

Научный рецензируемый журнал

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.

Периодичность 6 выпусков в год

Учредители

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

Сибирское отделение Российской академии наук

Главный редактор

В.К. Шумный – академик РАН, д-р биол. наук, профессор (Россия)

Заместители главного редактора

Н.А. Колчанов – академик РАН, д-р биол. наук, профессор (Россия)

Н.Б. Рубцов – д-р биол. наук, профессор (Россия)

Е.К. Хлесткина – д-р биол. наук, профессор (Россия)

Ответственный секретарь

Г.В. Орлова – канд. биол. наук (Россия)

Редакционный совет

В.С. Баранов – чл.-кор. РАН, д-р мед. наук (Россия)
Л.А. Беспалова – академик РАН, д-р с.-х. наук (Россия)
А. Бёрнер – д-р наук (Германия)
В.М. Говорун – академик РАН, д-р биол. наук (Россия)
И. Гроссе – д-р наук, проф. (Германия)
Г.Л. Дианов – д-р биол. наук, проф. (Великобритания)
Ю.Е. Дуброва – д-р биол. наук, проф. (Великобритания)
И.К. Захаров – д-р биол. наук, проф. (Россия)
И.А. Захаров-Гезехус – чл.-кор. РАН, д-р биол. наук (Россия)
С.Г. Инге-Вечтомов – академик РАН, д-р биол. наук (Россия)
И.Е. Керкис – д-р наук (Бразилия)
А.В. Кильчевский – чл.-кор. НАНБ, д-р биол. наук (Беларусь)
С.В. Костров – чл.-кор. РАН, д-р хим. наук (Россия)
Ж. Ле Гуи – д-р наук (Франция)
Б. Люгтенберг – д-р наук, проф. (Нидерланды)
В.И. Молодин – академик РАН, д-р ист. наук (Россия)
В.П. Пузырев – академик РАН, д-р мед. наук (Россия)
А.Ю. Ржецкий – канд. биол. наук, проф. (США)
И.Б. Рогозин – канд. биол. наук (США)
А.О. Рувинский – д-р биол. наук, проф. (Австралия)
К.Г. Скрябин – академик РАН, д-р биол. наук (Россия)
К.В. Славин – д-р наук, проф. (США)
И.А. Тихонович – академик РАН, д-р биол. наук (Россия)
Л.В. Хотылева – академик НАНБ, д-р биол. наук (Беларусь)
Э.К. Хуснутдинова – д-р биол. наук, проф. (Россия)
М.Ф. Чернов – д-р мед. наук (Япония)
С.В. Шестаков – академик РАН, д-р биол. наук (Россия)
Н.К. Янковский – академик РАН, д-р биол. наук (Россия)

Редакционная коллегия

Т.Г. Амстиславская – д-р биол. наук, доцент (Россия)
Е.Е. Андронов – канд. биол. наук (Россия)
Ю.С. Аульченко – д-р биол. наук (Россия)
Д.А. Афонников – канд. биол. наук, доцент (Россия)
Л.И. Афтanas – академик РАН, д-р мед. наук (Россия)
Е.В. Березиков – канд. биол. наук, проф. (Россия, Нидерланды)
С.А. Боринская – д-р биол. наук (Россия)
П.М. Бородин – д-р биол. наук, проф. (Россия)
М.И. Воевода – академик РАН, д-р мед. наук (Россия)
Т.А. Гавриленко – д-р биол. наук, доцент (Россия)
В.Н. Даниленко – д-р биол. наук, проф. (Россия)
С.А. Демаков – д-р биол. наук (Россия)
Е.А. Долгих – канд. биол. наук (Россия)
Н.Н. Дыгало – чл.-кор. РАН, д-р биол. наук (Россия)
С.Л. Киселев – д-р биол. наук, проф. (Россия)
В.А. Козлов – академик РАН, д-р мед. наук (Россия)
Ю.М. Константинов – д-р биол. наук, проф. (Россия)
А.В. Кочетов – чл.-кор. РАН, д-р биол. наук (Россия)
О. Кребс – д-р биол. наук, проф. (Германия)
И.Н. Лаврик – канд. хим. наук (Германия)
Л.А. Лутова – д-р биол. наук, проф. (Россия)
В.Ю. Макеев – чл.-кор. РАН, д-р физ.-мат. наук (Россия)
М.П. Мошкин – д-р биол. наук, проф. (Россия)
Н.А. Проворов – д-р биол. наук (Россия)
Д.В. Пышный – чл.-кор. РАН, д-р хим. наук (Россия)
А.В. Ратушный – канд. биол. наук (США)
Е.А. Салина – д-р биол. наук, проф. (Россия)
М.Г. Самсонова – д-р биол. наук (Россия)
В.А. Степанов – чл.-кор. РАН, д-р биол. наук (Россия)

Scientific Peer Reviewed Journal

VAVILOV JOURNAL OF GENETICS AND BREEDING

Vavilovskii Zhurnal Genetiki i Seleksii

Founded in 1997

Published 6 times annually

Founders

Federal State Budget Scientific Institution "The Federal Research Center Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences"

The Vavilov Society of Geneticists and Breeders

Siberian Branch of the Russian Academy of Sciences

Editor-in-Chief

V.K. Shumny, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

Deputy Editor-in-Chief

N.A. Kolchanov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

N.B. Rubtsov, Professor, Dr. Sci. (Biology), Russia

E.K. Khlestkina, Professor, Dr. Sci. (Biology), Russia

Executive Secretary

G.V. Orlova, Cand. Sci. (Biology), Russia

Editorial council

V.S. Baranov, Corr. Member of the RAS, Dr. Sci. (Medicine), Russia

L.A. Beshpalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia

A. Börner, Dr. Sci., Germany

M.F. Chernov, Dr. Sci. (Medicine), Japan

G.L. Dianov, Professor, Dr. Sci. (Biology), Great Britain

Yu.E. Dubrova, Professor, Dr. Sci. (Biology), Great Britain

V.M. Govorun, Full Member of the RAS, Dr. Sci. (Biology), Russia

J. Le Gouis, Dr. Sci., France

I. Grosse, Professor, Dr. Sci., Germany

S.G. Inge-Vechtomov, Full Member of the RAS, Dr. Sci. (Biology), Russia

I.E. Kerkis, Dr. Sci., Brazil

L.V. Khotyleva, Full Member of the NAS of Belarus, Dr. Sci. (Biology), Belarus

E.K. Khusnutdinova, Professor, Dr. Sci. (Biology), Russia

A.V. Kilchevsky, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology), Belarus

S.V. Kostrov, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia

B. Lugtenberg, Professor, Dr. Sci., Netherlands

V.I. Molodin, Full Member of the RAS, Dr. Sci. (History), Russia

V.P. Puzyrev, Full Member of the RAS, Dr. Sci. (Medicine), Russia

I.B. Rogozin, Cand. Sci. (Biology), United States

A.O. Ruvinsky, Professor, Dr. Sci. (Biology), Australia

A.Yu. Rzhetsky, Professor, Cand. Sci. (Biology), United States

S.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology), Russia

K.G. Skryabin, Full Member of the RAS, Dr. Sci. (Biology), Russia

K.V. Slavin, Professor, Dr. Sci., United States

I.A. Tikhonovich, Full Member of the RAS, Dr. Sci. (Biology), Russia

N.K. Yankovsky, Full Member of the RAS, Dr. Sci. (Biology), Russia

I.K. Zakharov, Professor, Dr. Sci. (Biology), Russia

I.A. Zakharov-Gezekhus, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

Editorial board

D.A. Afonnikov, Associate Professor, Cand. Sci. (Biology), Russia

L.I. Aftanas, Full Member of the RAS, Dr. Sci. (Medicine), Russia

T.G. Amstislavskaya, Associate Professor, Dr. Sci. (Biology), Russia

E.E. Andronov, Cand. Sci. (Biology), Russia

Yu.S. Aulchenko, Dr. Sci. (Biology), Russia

E.V. Berezikov, Professor, Cand. Sci. (Biology), Russia, Netherlands

S.A. Borinskaya, Dr. Sci. (Biology), Russia

P.M. Borodin, Professor, Dr. Sci. (Biology), Russia

V.N. Danilenko, Professor, Dr. Sci. (Biology), Russia

S.A. Demakov, Dr. Sci. (Biology), Russia

E.A. Dolgikh, Cand. Sci. (Biology), Russia

N.N. Dygalo, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

T.A. Gavrilenko, Associate Professor, Dr. Sci. (Biology), Russia

S.L. Kiselev, Professor, Dr. Sci. (Biology), Russia

A.V. Kochetov, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

Yu.M. Konstantinov, Professor, Dr. Sci. (Biology), Russia

V.A. Kozlov, Full Member of the RAS, Dr. Sci. (Medicine), Russia

O. Krebs, Professor, Dr. Sci. (Biology), Germany

I.N. Lavrik, Cand. Sci. (Chemistry), Germany

L.A. Lutova, Professor, Dr. Sci. (Biology), Russia

V.Yu. Makeev, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.), Russia

M.P. Moshkin, Professor, Dr. Sci. (Biology), Russia

N.A. Provorov, Dr. Sci. (Biology), Russia

D.V. Pyshnyi, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia

A.V. Ratushny, Cand. Sci. (Biology), United States

E.A. Salina, Professor, Dr. Sci. (Biology), Russia

M.G. Samsonova, Dr. Sci. (Biology), Russia

V.A. Stepanov, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

M.I. Voevoda, Full Member of the RAS, Dr. Sci. (Medicine), Russia

Геномика

746

ОБЗОР

Вычислительные проблемы анализа ошибок коротких прочтений ДНК при секвенировании следующего поколения

Р. те Боекхорст, Ф.М. Науменко, Н.Г. Орлова, Э.Р. Галиева, А.М. Спицина, И.В. Чадаева, Ю.Л. Орлов, И.И. Абнизова

756

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Определение числа копий эндогенных ретровирусов типа А у домашних свиней и диких кабанов

Р.Б. Айтназаров, Н.С. Юдин, Р.С. Кирильчук, Н.Н. Кочнев, С.П. Князев, М.И. Воевода

Анализ регуляторных последовательностей

762

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Анализ геномного распределения сайтов связывания транскрипционных факторов GAGA и CNC в развитии *Drosophila melanogaster*

И.И. Брусенцов, Д.А. Карагодин, Э.М. Баричева, Т.И. Меркулова

770

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Компьютерный анализ совместной локализации сайтов связывания транскрипционных факторов в геноме по данным ChIP-seq

А.И. Дергилев, А.М. Спицина, И.В. Чадаева, А.В. Свичкарев, Ф.М. Науменко, Е.В. Кулакова, Э.Р. Галиева, Е.Е. Витяев, М. Чен, Ю.Л. Орлов

779

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Оценка трансляционной значимости характеристик нуклеотидной последовательности мРНК млекопитающих на основе данных рибосомного профилирования

О.А. Волкова, Ю.В. Кондрахин, Р.Н. Шарипов

SNP-маркеры в биомедицине

787

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Кандидатные SNP-маркеры социального доминирования, способные влиять на родство ТАТА-связывающего белка к промоторам генов человека

И.В. Чадаева, Д.А. Рассказов, Е.Б. Шарыпова, Л.К. Савинкова, П.М. Пономаренко, М.П. Пономаренко

797

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Выявление новых регуляторных SNP, ассоциированных с предрасположенностью к развитию колоректального рака

Е.Ю. Леберфарб, Л.О. Брызгалов, И.И. Брусенцов, Т.И. Меркулова

Филогенетика и эволюция

804

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Эволюция СрG-островов путем тандемных дупликаций

В.Н. Бабенко, Ю.Л. Орлов, Ж.Т. Исакова, Д.А. Антонов, М.И. Воевода

816

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Комбинирование и оценка конгруэнтности филогенетических сигналов от нескольких генов с помощью геометрического подхода

В.М. Ефимов, В.Ю. Ковалева, Ю.Н. Литвинов

823

ОБЗОР

Структурно-функциональная дивергенция гомеологичных генов в аллополиплоидном геноме растений

А.Ю. Глаголева, О.Ю. Шоева, Е. К. Хлесткина

Системная биология и моделирование

832

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Анализ генной сети Neurogenesis:Prepattern, контролирующей первый этап формирования щетиночного узора у *Drosophila melanogaster*

Д.П. Фурман, Т.А. Бухарина

840

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Молекулярно-генетические механизмы взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открытоугольной глаукоме

О.В. Сайк, Н.А. Коновалова, П.С. Деменков, Н.В. Иванисенко, Т.В. Иванисенко, Д.Е. Иванощук, О.С. Коновалова, О.А. Подколодная, И.Н. Лаврик, Н.А. Колчанов, В.А. Иванисенко

- 848 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Математическая модель циркадного осциллятора млекопитающих: взаимодействие с системой NAD⁺/SIRT1 и возрастные изменения экспрессии генов циркадного осциллятора
Н.Л. Подкольный, Н.Н. Твердохлеб,
О.А. Подкольная

- 857 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Структурное моделирование мод связывания НАД⁺ с PARP-1
Н.В. Иванисенко, Д.А. Жечев, В.А. Иванисенко

Экспериментальная системная биология

- 863 **МЕТОДЫ И ПРОТОКОЛЫ**
Протокол анализа количественных характеристик опушения листа картофеля
А.В. Дорошков, М.А. Генаев, Д.А. Афонников

- 869 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Исследование влияния электромагнитного излучения терагерцового диапазона на протеом экстремофильной археи *Halorubrum saccharovorum*
Т.Н. Горячковская, С.Г. Константинова,
И.А. Мещерякова, С.В. Банникова, Е.А. Демидов,
А.В. Брянская, М.А. Щеглов, А.И. Семенов,
Д.Ю. Ощепков, В.М. Попик, С.Е. Пельтек

- 876 **ОБЗОР**
Стрессовые системы *Escherichia coli* и их роль в реакциях на воздействие терагерцового излучения
С.Е. Пельтек, Е.В. Демидова, В.М. Попик,
Т.Н. Горячковская

Генетика человека

- 887 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Полиморфизм регуляторной области гена *LCT* в некоторых тюркоязычных популяциях Алтае-Саянского региона
И.В. Пилипенко, М.С. Пристяжнюк, В.Ф. Кобзев,
М.И. Воевода, А.С. Пилипенко

- 894 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Динамика экспрессии транскрипционных факторов REL, RELA и IRF1 в макрофагоподобной линии U937 после воздействия диоксина
Е.В. Кашина, Д.Ю. Ощепков, Е.В. Антонцева,
М.Ю. Шаманина, Д.П. Фурман, В.А. Мордвинов

Экологическая генетика

- 899 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Симбиотическая бактерия *Wolbachia* в популяциях вредителя хвойных лесов *Dendrolimus superans sibiricus* Tschetverikov, 1908 (Lepidoptera: Lasiocampidae)
М.А. Юдина, В.В. Дубатовлов, Р.А. Быков,
Ю.Ю. Илинский

- 904 **ОБЗОР**
База данных генов, повышающих устойчивость пшеницы и родственных ей злаков к патогенным грибам
О.Г. Смирнова, А.В. Кочетов

Актуальные технологии генетики и клеточной биологии

- 909 **ОБЗОР**
Методы маркирования клеток для изучения судьбы клеточных поколений
А.М. Юнусова, Н.Р. Баттулин

- 918 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Морфофизиологические эффекты инсерционного мутагенеза гена контактин 5 (*Cntn5*) у трансгенных мышей
А.В. Смирнов, Н.А. Феофанова, Г.В. Концевая,
М.В. Анисимова, И.И. Ковригин, И.А. Серова,
М.П. Мошкин, Л.А. Герлинская, Н.Р. Баттулин

- 925 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Эффективное получение химерных мышей с использованием новой линии эмбриональных стволовых клеток
Г.В. Концевая, Н.А. Феофанова, А.Г. Мензоров,
И.Е. Пристяжнюк, А.В. Смирнов, Н.Р. Баттулин,
Л.А. Герлинская

930

МЕТОДЫ И ПРОТОКОЛЫ

Практическое руководство
по редактированию геномов
системой CRISPR/Cas9

*А.Г. Мензоров, В.А. Лукьянчикова, А.Н. Кораблев,
И.А. Серова, В.С. Фишман*

945

ОБЗОР

Зачем нужны коллекции
линий клеток

А.Г. Мензоров, О.Л. Серов

949

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Сравнительное изучение
репликативных свойств
противоопухолевых
рекомбинантных вирусов
осповакцины на культурах
клеток глиобластомы человека U87
и почки мартышки CV-1

*Р.А. Максюттов, И.В. Колосова, Т.В. Трезубчак,
И.А. Разумов, С.Н. Щелкунов*

Моделирование патологий

954

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Ренин-ангиотензин-альдостероновая
система у крыс линии НИСАГ
(ISIAN) со стресс-индуцированной
артериальной гипертензией

*А.Д. Дубинина, Е.В. Антонов, Л.А. Федосеева,
Е.Н. Пивоварова, А.Л. Маркель, Л.Н. Иванова*

959

ОБЗОР

Аутизм как проявление нарушения
молекулярных механизмов
регуляции развития и функций
синапсов

Е.А. Трифонова, Т.М. Хлебодарова, Н.Е. Грунтенко

968

Алфавитный указатель авторов
статей, опубликованных в журнале
в 2016 г.

Genomics

- 746 **REVIEW**
Computational problems of analysis of short next generation sequencing reads
R. te Boekhorst, F.M. Naumenko, N.G. Orlova, E.R. Galieva, A.M. Spitsina, I.V. Chadaeva, Y.L. Orlov, I.I. Abnizova

- 756 **ORIGINAL ARTICLE**
Determination of the copy numbers of type A porcine endogenous retroviruses in domestic pigs and wild boars
R.B. Aitnazarov, N.S. Yudin, R.S. Kiril'chuk, N.N. Kochnev, S.P. Knyazev, M.I. Voevoda

Analysis of regulatory sequences

- 762 **ORIGINAL ARTICLE**
Analysis of whole-genome binding patterns of GAGA and CNC transcription factors during *Drosophila melanogaster* development
I.I. Brusentsov, D.A. Karagodin, E.M. Baricheva, T.I. Merkulova

- 770 **ORIGINAL ARTICLE**
Computer analysis of co-localization of transcription factor binding sites in genome by ChIP-seq data
A.I. Dergilev, A.M. Spitsina, I.V. Chadaeva, A.V. Svichkarev, F.M. Naumenko, E.V. Kulakova, E.R. Galieva, E.E. Vityaev, M. Chen, Y.L. Orlov

- 779 **ORIGINAL ARTICLE**
Estimation of translational importance of mammalian mRNA nucleotide sequence characteristics based on ribosome profiling data
O.A. Volkova, Yu.V. Kondrakhin, R.N. Sharipov

SNP markers in biomedicine

- 787 **ORIGINAL ARTICLE**
Candidate SNP markers of social dominance, which may affect the affinity of the TATA-binding protein for human gene promoters
I.V. Chadaeva, D.A. Rasskazov, E.B. Sharypova, L.K. Savinkova, P.M. Ponomarenko, M.P. Ponomarenko

- 797 **ORIGINAL ARTICLE**
Detection of new regulatory SNPs associated with colorectal cancer predisposition
E.Yu. Leberfarb, L.O. Bryzgalov, I.I. Brusentsov, T.I. Merkulova

Phylogenetics and evolution

- 804 **ORIGINAL ARTICLE**
Evolution of CpG-islands by means of tandem duplications
V.N. Babenko, Y.L. Orlov, Zh.T. Isakova, D.A. Antonov, M.I. Voevoda

- 816 **ORIGINAL ARTICLE**
Combining and congruence evaluation of phylogenetic signals from different genes based on geometric approach
V.M. Efimov, V.Yu. Kovaleva, Yu.N. Litvinov

- 823 **REVIEW**
Structural and functional divergence of homoeologous genes in allopolyploid plant genomes
A.Y. Glagoleva, O.Y. Shoeva, E.K. Khlestkina

Systems biology and simulations

- 832 **ORIGINAL ARTICLE**
Analysis of the Neurogenesis:Prepattern gene network controlling the first stage in *Drosophila melanogaster* bristle pattern development
D.P. Furman, T.A. Bukharina

- 840 **ORIGINAL ARTICLE**
Molecular-genetic mechanisms of the interaction between processes of cell response to mechanical stress and neuronal apoptosis in primary open-angle glaucoma
O.V. Saik, N.A. Konovalova, P.S. Demenkov, N.V. Ivanisenko, T.V. Ivanisenko, D.E. Ivanoshchuk, O.S. Konovalova, O.A. Podkolodnaya, I.N. Lavrik, N.A. Kolchanov, V.A. Ivanisenko

- 848 **ORIGINAL ARTICLE**
Computational model for mammalian circadian oscillator: interacting with NAD⁺/SIRT1 pathway and age-related changes in gene expression of circadian oscillator
N.L. Podkolodnyy, N.N. Tverdokhleby, O.A. Podkolodnaya

- 857 **ORIGINAL ARTICLE**
Structural modeling of NAD⁺ binding modes to PARP-1
N.V. Ivanisenko, D.A. Zhechev, V.A. Ivanisenko

Experimental systems biology

- 863 **METHODS AND PROTOCOLS**
A protocol for analysis of the quantitative characteristics of leaf pubescence in potato
A.V. Doroshkov, M.A. Genaev, D.A. Afonnikov

- 869 **ORIGINAL ARTICLE**
The impact of terahertz radiation on an extremophilic archaean *Halorubrum saccharovororum* proteome
T.N. Goryachkovskaya, S.H. Konstantinova, I.A. Meshcheriakova, S.V. Bannikova, E.A. Demidov, A.V. Bryanskaya, M.A. Scheglov, A.I. Semenov, D.Yu. Oshchepkov, V.M. Popik, S.E. Peltek

- 876 **REVIEW**
Escherichia coli stress response systems and their reaction to terahertz radiation
S.E. Peltek, E.V. Demidova, V.M. Popik, T.N. Goryachkovskaya

Human genetics

- 887 **ORIGINAL ARTICLE**
Polymorphism of *LCT* gene regulatory region in Turkic-speaking populations of the Altay-Sayan region (southern Siberia)
I.V. Pilipenko, M.S. Pristyazhnyuk, V.F. Kobzev, M.I. Voevoda, A.S. Pilipenko

- 894 **ORIGINAL ARTICLE**
Dynamics of REL, RELA and IRF1 transcription factor expression in U937 macrophages after dioxin exposure
E.V. Kashina, D.Y. Oshchepkov, E.V. Antontseva, M.Y. Shamanina, D.P. Furman, V.A. Mordvinov

Ecological genetics

- 899 **ORIGINAL ARTICLE**
Wolbachia infection in populations of the coniferous forest pest *Dendrolimus superans sibiricus* Tschetverikov, 1908 (Lepidoptera: Lasiocampidae)
M.A. Yudina, V.V. Dubatolov, R.A. Bykov, Yu.Yu. Ilinsky

- 904 **REVIEW**
A database on genes increasing resistance of wheat and relative species against pathogenic fungi
O.G. Smirnova, A.V. Kochetov

Mainstream technologies in genetics and cell biology

- 909 **REVIEW**
Cell-marking techniques for cell lineage tracing
A.M. Yunusova, N.R. Battulin

- 918 **ORIGINAL ARTICLE**
Morphophysiological alterations caused by insertional mutagenesis of contactin 5 (*Cntn5*) gene in transgenic mice
A.V. Smirnov, N.A. Feofanova, G.V. Kontsevaya, M.V. Anisimova, I.I. Kovrigin, I.A. Serova, M.P. Moshkin, L.A. Gerlinskaya, N.R. Battulin

- 925 **ORIGINAL ARTICLE**
Efficient chimeric mouse production using a novel embryonic stem cell line
G.V. Kontsevaya, N.A. Feofanova, A.G. Menzorov, I.E. Pristyazhnyuk, A.V. Smirnov, N.R. Battulin, L.A. Gerlinskaya

- 930 **METHODS AND PROTOCOLS**
Genome editing using CRISPR/Cas9 system: a practical guide
A.G. Menzorov, V.A. Lukyanchikova, A.N. Korablev, I.A. Serova, V.S. Fishman
- 945 **REVIEW**
The necessity of cell banks
A.G. Menzorov, O.L. Serov
- 949 **ORIGINAL ARTICLE**
A comparative study of replicative properties of antitumor recombinant vaccinia viruses on human glioblastoma cell culture U87 and monkey kidney cell culture CV-1
R.A. Maksyutov, I.V. Kolosova, T.V. Tregubchak, I.A. Razumov, S.N. Shchelkunov
- Modeling of disorders**
- 954 **ORIGINAL ARTICLE**
Renin-angiotensin-aldosterone system in ISIAH rats with stress-induced arterial hypertension
A.D. Dubinina, E.V. Antonov, L.A. Fedoseeva, E.N. Pivovarova, A.L. Markel, L.N. Ivanova
- 959 **REVIEW**
Molecular mechanisms of autism as a form of synaptic dysfunction
E.A. Trifonova, T.M. Khlebodarova, N.E. Gruntenko
- 968 **Alphabetical author index for the list of papers published in the journal in 2016**

Настоящий выпуск журнала подготовлен при активном содействии ответственных редакторов – заведующего центром коллективного пользования «Биоинформатика» Н.Л. Подкольного и доктора биологических наук, профессора М.П. Мошкина – и содержит статьи по ряду актуальных направлений биоинформатики, системной компьютерной биологии и биотехнологии. Часть работ была представлена на Десятой международной мультikonференции по биоинформатике и системной биологии BGRS\SB-2016, прошедшей в августе–сентябре 2016 г. в Новосибирском академгородке. Широкий спектр проблем системной биологии нашел свое отражение в научных исследованиях, опубликованных в этом номере.

Рубрика «Геномика» представляет обзор по технологиям высокопроизводительного секвенирования ДНК и возникающим техническим и вычислительным задачам процессинга данных, а также оригинальную работу по геномному анализу эндогенных ретровирусов у домашних свиней и диких кабанов.

Раздел «Анализ регуляторных последовательностей» включает статьи по геномике, технологиям секвенирования, анализу регуляторных последовательностей ДНК. Рассмотрены проблемы геномного распределения сайтов связывания транскрипционных факторов, их совместной локализации в геноме, анализ данных ChIP-seq и рибосомного профилирования.

Проблема анализа маркеров однонуклеотидных полиморфизмов поднята в статьях по поиску новых регуляторных полиморфизмов в геноме человека, ассоциированных с предрасположенностью к заболеваниям (рубрика «SNP-маркеры в биомедицине»).

Эволюция геномных последовательностей рассматривается в работах по эволюции CpG-островов в геноме человека, функциональной дивергенции гомеологичных генов в геномах растений (рубрика «Филогенетика и эволюция»).

Представлена оригинальная работа по статистической оценке филогенетических сигналов с помощью геометрического подхода.

В рубрике «Системная биология и моделирование» приводится анализ генных сетей у дрозофилы, описан циркадный ритм у млекопитающих. Рассмотрены также молекулярно-генетические механизмы взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при глаукоме, компьютерное предсказание структур белковых комплексов.

Рубрика «Экспериментальная системная биология» включает исследование воздействия терагерцового излучения на прокариотические организмы. Здесь же дано описание протокола по количественному анализу фенотипических характеристик картофеля.

Статьи рубрики «Генетика человека» посвящены популяционному исследованию полиморфизмов локуса, контролирующего активность гена лактазы в двух тюркоязычных группах населения Южной Сибири – алтайских казахов и хакасов, а также исследованию функциональной роли некоторых элементов в промоторах генов *IRF1*, *REL*, *RELA* человека при действии ксенобиотика диоксина.

Рубрика «Экологическая генетика» включает статью по анализу инфицированности особей сибирского шелкопряда – крайне опасного вредителя хвойных деревьев, симбиотической бактерией *Wolbachia*. Компьютерные подходы к анализу взаимодействия патогенных грибов и злаковых культур нашли отражение в базе данных генов устойчивости пшеницы к патогенам.

В рубрике «Актуальные технологии генетики и клеточной биологии» представлены работы, посвященные созданию новых линий трансгенных животных и практическим аспектам использования системы редактирования генома CRISPR/Cas9 для внесения мутаций в геном млекопитающих. Помимо оригинальных статей раздел знакомит с двумя обзорами. В первом обзоре собраны примеры использования последних достижений в области клеточной биологии для маркирования потомков отдельных клеток *in vivo*, во втором – обоснована необходимость создания в Российской Федерации коллекции культур плюрипотентных клеток для развития биомедицинских исследований.

Статья из рубрики «Моделирование патологий» посвящена исследованию состояния ренин-ангиотензиновой системы у крыс линии НИСАГ – одной из моделей гипертонической болезни человека. Также в этот раздел входит обзор молекулярных механизмов синдрома аутизма, обусловленного моногенными мутациями, в котором предложена механизм-обоснованная терапия некоторых расстройств аутистического спектра.

Заместитель главного редактора
академик Н.А. Колчанов

Computational problems of analysis of short next generation sequencing reads

R. te Boekhorst¹, F.M. Naumenko², N.G. Orlova^{2, 3}, E.R. Galieva^{2, 4}, A.M. Spitsina², I.V. Chadaeva², Y.L. Orlov^{2, 4}, I.I. Abnizova⁵

¹ University of Hertfordshire, Hatfield, UK

² Novosibirsk State University, Novosibirsk, Russia

³ Novosibirsk State University of Architecture and Civil Engineering (Sibstrin), Novosibirsk, Russia

⁴ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

⁵ Wellcome Trust Sanger Institute, Cambridge, UK

Short read next generation sequencing (NGS) has significant impacts on modern genomics, genetics, cell biology and medicine, especially on meta-genomics, comparative genomics, polymorphism detection, mutation screening, transcriptome profiling, methylation profiling, chromatin remodelling and many more applications. However, NGS are prone for errors which complicate scientific conclusions. NGS technologies consist of shearing DNA molecules into collection of numerous small fragments, called a 'library', and their further extensive parallel sequencing. These sequenced overlapping fragments are called 'reads', they are assembled into contiguous strings. The contiguous sequences are in turn assembled into genomes for further analysis. Computational sequencing problems are those arising from numerical processing of sequenced samples. The numerical processing involves procedures such as: quality-scoring, mapping/assembling, and surprisingly, error-correction of a data. This paper is reviewing post-processing errors and computational methods to discern them. It also includes sequencing dictionary. We present here quality control of raw data, errors arising at the steps of alignment of sequencing reads to a reference genome and assembly. Finally this work presents identification of mutations ("Variant calling") in sequencing data and its quality control.

Keywords: next generation sequencing (NGS); DNA; sequencing technologies; statistical biases; genome polymorphisms; sequencing errors; review.

Вычислительные проблемы анализа ошибок коротких прочтений ДНК при секвенировании следующего поколения

Р. те Боекхорст¹, Ф.М. Науменко², Н.Г. Орлова^{2, 3}, Э.Р. Галиева^{2, 4}, А.М. Спицина², И.В. Чадаева², Ю.Л. Орлов^{2, 4}, И.И. Абнизова⁵

¹ Университет Хартфордшира, Хатфилд, Великобритания

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Новосибирский государственный архитектурно-строительный университет (Сибстрин)», Новосибирск, Россия

⁴ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

⁵ Институт Сэнгера, Велком Траст, Кембридж, Великобритания

Секвенирование следующего поколения (NGS) с помощью коротких прочтений ДНК вносит большой вклад в решение задач современной геномики, генетики, клеточной биологии и медицины, особенно в исследования метагеномики, сравнительной геномики, определение полиморфизмов, скрининг мутаций, транскриптомное профилирование, изучение ремоделирования хроматина и многие другие приложения. Секвенирование неустойчиво к техническим ошибкам, которые могут влиять на научные выводы. NGS технологии состоят из создания коллекции многочисленных коротких фрагментов ДНК, именуемой «библиотекой», получения молекулярных колоний и их дальнейшего массового параллельного секвенирования. Такие секвенированные фрагменты называются «прочтениями», они собираются (ассемблируются) в протяженные строки. Протяженные последовательности, в свою очередь, собираются в геномы для дальнейшего анализа. Вычислительные/процессинговые ошибки и сбои секвенирования – это ошибки, возникающие при последующей цифровой обработке секвенированных образцов. Последующая обработка (процессирование) включает процедуры оценки качества, картирования, ассемблирования и даже корректировки ошибочных данных. Данная статья рассматривает вычислительные ошибки процессирования, компьютерные и статистические подходы для их определения, а также представляет словарь терминологии секвенирования. Рассмотрены задачи идентификации мутаций («Определение вариаций») в данных секвенирования и контроль качества их определения. Определение вариаций включает локальные вариации, такие как одиночные нуклеотидные полиморфизмы, короткие вставки и делеции (инделлы), и масштабные вариации (инверсии, транслокации или большие

REVIEW

Received 18.11.2016

Accepted for publication 20.12.2016

© AUTHORS, 2016

инделы). Обсуждены проблемы контроля качества исходных (сырых) данных, ошибки, возникающие на этапах выравнивания прочтений последовательностей ДНК на референсный геном и последующего выравнивания/ассемблирования.

Ключевые слова: секвенирование следующего поколения; ДНК; технологии секвенирования; статистические неоднородности; геномные полиморфизмы; ошибки секвенирования; обзор.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

те Боекхорст Р., Науменко Ф.М., Орлова Н.Г., Галиева Э.Р., Спицина А.М., Чадаева И.В., Орлов Ю.Л., Абнизова И.И. Вычислительные проблемы анализа ошибок коротких прочтений ДНК при секвенировании следующего поколения. Вавиловский журнал генетики и селекции. 2016;20(6):746-755. DOI 10.18699/VJ16.191

HOW TO CITE THIS ARTICLE:

te Boekhorst R., Naumenko F.M., Orlova N.G., Galieva E.R., Spitsina A.M., Chadaeva I.V., Orlov Y.L., Abnizova I.I. Computational problems of analysis of short next generation sequencing reads. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):746-755. DOI 10.18699/VJ16.191

Here we will concentrate on second generation short read DNA sequencing (Liu et al., 2012), and will drop off the term ‘second generation short read’ while speaking about NGS further. The **terms in bold** are explained in the Glossary.

NGS technologies have as an essential feature of breaking DNA molecules into stack of numerous fragments, called a ‘library’. The end parts of these fragments are sequenced in parallel, and called ‘reads’. They are assembled into contiguous strings. These assembled sub-sequences are in turn assembled into genomes, and are subjects for further analysis.

We refer to (van Dijk et al., 2014; Anders et al., 2015) for detailed characteristics of NGS platforms.

The NGS data processing is arranged in a set of consecutive steps, called a pipeline. A common post-sequencing NGS pipeline (Mutarelli et al., 2014; Newell, 2014) consists of:

- (1) *Quality Control (QC) of initial data;*
- (2) *Mapping to a reference genome and/or assembly;*
- (3) *Post-mapping/assembly QC and re-calibration;*
- (4) *Variant calling and its QC;*
- (5) *Correcting of errors.*

The step 2 may be combined and/or substituted by a de-novo genome assembly (Baker, 2012) in case there is no reference for the sequenced genome.

We refer readers to reviews on NGS computational frameworks (Dolled-Filhart et al., 2013; Guo et al., 2014a; Mutarelli et al., 2014; Pabinger et al., 2014). The on-line sources (Li J. et al., 2012b; Hadfield, 2013; Ignatieva et al., 2015) can help researchers to build up their own pipelines. For ‘meta’ pipelines we refer to (Anders et al., 2015; Wolfinger et al., 2015) which describe tools to build up bespoke pipelines.

For each of the steps in a pipeline above we will review (a) what they are and what is their goal; (b) how it is done summarising the methodology, their advantages and pitfalls.

1. QC of initial data

For any platform, an initial unprocessed digital outputs of a sequencing are a **base calls** and their **qualities**.

Compared to a first generation Sanger sequencing (Sanger et al., 1992), NGS technologies are confronted by shorter read length, platform/instrument/sample specific biases (Harismendy et al., 2009), higher error rate, and irregular coverage. These factors lower the accuracy of NGS further analysis (e. g. variant calls and de-novo assembly) by introducing sequencing errors that may direct to mis-interpretation of data.

The main factors utilised in quality control of raw data to characterise sequencer’s performance and library preparation are:

Total read count; Proportion of high quality data; Nucleotide and quality distribution per cycle; Duplication rate (can be optical or amplification duplicates); Adapter’s counts; Proportion of bases per sample for pooled multiplexed data;
♦ *Total read count* shows general library effectiveness. It should be reasonably large to produce results of statistical significance.

♦ *Proportion of high quality ($Q > 30$) bases within Q value distribution* should be large: at least more than half. A base call is scored with low Q mostly because of sequencer’s preferences and faults (Abnizova et al., 2010; Ledergerber, Dessimoz, 2011). These low quality bases are typically trimmed or corrected (Kelley et al., 2010; Del Fabbro et al., 2013), so low Q and possibly wrong called data will not compromise downstream analysis. However, an error correction should be applied only to high-coverage and homogeneous data – an assumption that often fails for NGS data.

♦ *Quality-per-cycle distribution.* A random quality peaks/deeps per cycle point to some problems on machine during sequencing. Quality usually declines gradually with cycle as a result of increasing signal-to-noise ratio.

♦ *Duplicate reads*, appearing due to PCR (polymerase chain reaction) and optical problems, may lead to over-estimating of some variant contribution in the data. Duplicate removing is debated in (Pireddu et al., 2011; Davis et al., 2013). Thus, their proportion should be less than 10 %.

♦ *Proportion if adapters* should less than 10 % as well. Parts of adapter might be erroneously sequenced in the beginning of a read, and thus may bring artificial mutations (Martin, 2011; Li J. et al., 2012a). The popular tools for adapter removing are discussed in (Marroni et al., 2012; Jiang et al., 2014).

♦ *Di-multiplexing*, namely splitting up samples based on their tags, should be even across tags in theory. It is very important that the size of each pool is sufficient and equal (Mir et al., 2013) for pooled multiplexed samples. Fairly even di-multiplexing (Hadfield, 2013) provides less biased data.

♦ There is also a possibility to quality control a library before massive sequencing. The MiSeq QC (Illumina, 2014) enables performing a preliminary run on libraries before deep-sequencing on a bigger machine, such as HiSeq or HiSeqX.

Nevertheless, any individual QC metric should be regarded in context of its project (Guo et al., 2014a).

A lot of sequencers (Cox et al., 2010) generate a QC reports included into their processing pipeline, and these reports investigate mainly a general performance of the corresponding sequencer. They typically do not cover any effects of sample extraction and library preparations.

A special case is fastQC. It is built up to point to problems which developed either in the library preparation or on sequencer. It is a very fast and crude estimation of different metrics formed by stratified samples of the data.

Alternatively, FaQCs (Lo, Chain, 2014) records errors in the whole data. It also takes away low Q-value reads.

The NGS QC Toolkit (Patel, Jain, 2012), except of performing a quality check and generating descriptive statistics, trims low Q ends of reads and removes low Q bases. It also enables a conversion between various file formats of NGS data from Illumina and Roche 454 platforms.

One should be careful: what is removed might be a genuine biological signal. Nevertheless, any fluctuation from expected values for the QC metrics, might be a possible error.

2. Mapping/aligning to a reference genome and/or assembly

The further step is the matching of the reads to positions at the reference genome, so called mapping. This is done by aligning reads to sub-sequences of the reference genome to which they are most close in terms of nucleotide sequence. Computationally, mapping is the most time and memory consuming step (Day-Williams, Zeggini, 2011; Fonseca et al., 2012). It is also critical: any mistake in alignment will be subject to further processing and hence spread errors to the further stages of sequencing and analysis.

For the short reads of NGS, it is too inefficient in time and memory to use the well-known BLAST (Altschul et al., 1990) algorithm to map reads to genome. Therefore a particular memory and time optimised mapping algorithms are developed.

NGS mappers/aligners can be classified based on their methods: hash table indexing (Shang et al., 2014) or Burrows-Wheeler Transform (BWT) (Li, Durbin, 2010). They also differ by computer resource usage and sensitivity. Thus, they may lead to a different mapping results. Here we define sensitivity as a proportion of genome which is covered by at least one read after mapping. Mapping algorithms also vary in their ability to deal with particular sequencing platforms, quality of base, protocols and in the dealing with structural features of the DNA subject to sequencing, such as repetitive motives, gaps, deletions and insertions.

Both types of aligners typically pre-process and index both reference and reads before a search of matching read positions (in the reference genome) itself. A hash table is a kind of a look up table, only supplied with advanced structure of indexing. BWT usually compresses data in a particular way (modification of a suffix array) before matching. BWT aligners are less sensitive than hash table methods, but are faster and use less memory (Newell, 2014).

A sequence assembly refers to aligning and integrating short fragments from a sequenced DNA in order to recreate the original sequence. If the genome of an organism has not

been sequenced before, the assembly results in the first form of its reference genome. This procedure is called “de-novo assembly”. Sometimes a de-novo assembly is used together with alignment to reconstruct previously insufficiently covered and untrustworthy sequenced genome loci.

Present-day assembling algorithms for NGS comprise two main groups (Li Z. et al., 2012): (i) Overlap-layout-consensus (OLC) methods; and (ii) Eulerian/de Bruijn Graph (DBG) methods. Both groups apply a graph theory to deal with NGS data, but in OLC notation reads are nodes, while in DBG notation a k -mer is a node. A read’s overlapping sequences stand for graph edges in both groups of assemblers.

We refer to (Li, Homer, 2010; Nagarajan, Pop, 2013; Otto et al., 2014; Pightling et al., 2014) or comparisons and benchmark aligners tests; for exhaustive literature on assemblers read (Nagarajan, Pop, 2013; Chin et al., 2014; Shang et al., 2014). An online list of aligners is regularly updated at (<http://www.ebi.ac.uk>).

What can go wrong?

- *Reference mistakes.* One should understand that an alignment step is apparently dependent on a reference’s accuracy. In the case of incorrect reference many reference errors could be mistreated as high quality genetic variants.
- A bias *shared by* most technologies is that their accuracy decreases with the number of sequencing cycles, thus an error of mapping the end of a read grows (Balint, 2016; Sameith et al., 2016).
- Of the more specific defects, we refer to: *platform-dependent issues; the type of protocol used; complications due to the functional and structural complexity of the sample DNA.*
- *Read Length and Error Rates*

Read lengths span from 70–1500 bp (Newell, 2014) depending on the sequencing platform. If reads are short it is harder to match them precisely to a unique genomic location.

Some sequencing platforms allow for longer read lengths than others (for example 200 bp by Ion Torrent and 700 bp by Roche’s 454, while Illumina’s reads of 100–250 bp) which makes mapping more precise. However, this advantage is defeated by their higher mismatch-error rate; aligners throw away reads with too many mismatches on the basis of a pre-set mismatch error rate.

- *Platform-dependent issues*

The technology on which a platform is founded may be prone to a certain sequencing mistakes, resulting in platform-specific error characteristic.

A ‘light-based’ sequencing platforms, Illumina, SOLiD, and Complete Genomics, employ fluorescent dye’s labelling to measure a signal strength for a successive sequencing cycle. The light-based platforms are known to be impaired by GC-bias, i.e. a low coverage of either **GC-rich** or **GC-poor** DNA regions (Chen et al., 2013; Rieber et al., 2013). Its origin is likely to be a fragmentation or/and cloning procedures during library preparation (Benjamini, Speed, 2012; Ross et al., 2013).

The light-based platforms typically are disadvantaged by single nucleotide miss-identifications. The SOLiD platform is known to have difficulties with sequencing palindromic sequences (Huang et al., 2012).

Ion Torrent’s Personal Genome Machine (PGM) (Niu et al., 2010) utilises semiconductor sequencing technology that oper-

ates on acidity (pH) rather than on light. Roche's 454 (Niu et al., 2010) employs a pyro-sequencing technology. An accuracy of both technologies depends on the length of sub-sequences of identical nucleotides ("homo-polymers") because of similar computational approaches to evaluate a homo-polymer length. Defective flow-calls result in insertion/deletion (indel) errors: they are largely homo-polymer-associated errors in case when short homo-polymers are frequent while long are rare (Bragg et al., 2013; Li et al., 2013).

Recognizing indels from NGS is known to be very daring (Li et al., 2013), because 'indel by itself obstructs with precise mapping'. To map indels precisely, pair-end (PE) information is employed (Albers et al., 2011). It is valid for indels half a size of reads. Longer deletions are detected by a split-read method.

To distinguish long insertions a de-novo assembly of weakly covered regions is required (Li et al., 2013).

- *Sequence-specific errors*: For pyrosequencing platforms, a 'homopolymer-associated errors' result into throwing away repetitive DNA after mapping. Indel errors are known to be context-dependent. Moreover, for Ion Torrent, GC-poor organisms have higher error rate and poorer coverage than GC-balanced. The nucleotide context of Illumina errors is reported in (Minoche et al., 2011).

- *DNA complexity: DNA functionality causes aligning biases*
The study of NGS artefacts in (Schwartz et al., 2011) showed that less linguistically complex sequences of introns are less covered with reads than more complex sequences of exons. The authors discovered that peaks of mapped reads were associated with biological features, such as intron-exon junction, expression level, splice sites and transcription length.

Similarly, the authors of (Auerbach et al., 2009) found that regions proximal to promoters are prone for sonication breakage, and hence are the subjects of regional bias. These regions are the primary cause of an uneven read coverage, retaining a large peaks of aligned reads.

- *DNA complexity: Repetitive DNA causes assembly problem*

A particular troublesome feature of the sequential structure of many genomes is the occurrence of long chunks of repetitive DNA (so-called "repeats"): repetitive DNA is frequently overlooked, miss-mapped and miss-assembled by all platforms (McCoy et al., 2014).

Around half of human genome is comprised of repetitive DNA (de Koning et al., 2011), the fraction of repeats is even larger for some plant genomes (Feschotte et al., 2002). Even though repetitive DNA is functionally important, NGS sequencing often fails to sequence it flawlessly (Alkan et al., 2011b; Ye et al., 2011). Most current technologies are error-prone while handling repeats.

But granting a repetitive DNA stretch is sequenced correctly, it might be compromised by similar DNA in other genome location, and lead to mis-alignment. And finally, repetitive DNA is often a hot-spot of real biological mutations and structural variations (Orlov et al., 2006; Medvedev et al., 2009; Safronova et al., 2015, 2016).

In addition to various repetitive DNA, a short indels and segmental duplications are also difficult to align and assemble (McCoy et al., 2014) because of ambiguity at which location to map an identical DNA subsequence.

The main assumption of assembly (similar reads belong to the same location) is breached by various repeats and polymorphisms. An assembly is computationally not tractable for genomes where the ratio of repeat length to read length is large (Nagarajan, Pop, 2013).

When whole long repetitive stretch were sequenced together with their flanking regions, it would be easier to detect it within genome. Therefore, longer reads could solve this problem (Huddleston et al., 2014).

- *Diversity of protocols: PE and MP methods*

The types of sequencing protocols depend on a researcher's question: e.g. reads sequenced in pairs (pair end, PE) (Medvedev et al., 2009) or singles (SE). PE reads are designed to detect direction and distance between reads, therefore reads containing complex DNA can be mapped uniquely (Miller et al., 2010; Alkan et al., 2011a).

A sub-type of PE reads, the long inserts reads (up to 5 KB), frequently named as **mate-pair libraries (MP)** (Park, 2013) are valuable to connect long repeats (including **repetitive transposable elements**) and structural variations.

Longer reads can solve the assembly and mapping problems. With longer reads it is easier to establish a correct genomic location for a sequenced DNA. Therefore, a new synthetic long reads (McCoy et al., 2014) from the Illumina TruSeq are developed. They are as long as 3d generation PacBio (Sharon et al., 2013), but much more accurate, having as low error rate as 0.03 % per base.

These synthetic long reads are assembled from Illumina short reads, by combination of laboratory and computational efforts (Voskoboinik et al., 2013). Nonetheless, there are still some imperfections left: gaps in assembly and a low coverage for repetitive AT-rich regions.

Regrettably, when some problems are reduced, a new ones arise. The essential problems of MP (Park, 2013) are: (i) extremely elaborated construction of their libraries, and (ii) common mistakes of mapping: 'inward facing' reads as a substitute of 'outward facing'. This mistake results into **chimeric reads** (Illumina). Another problems are: unexpectedly small **insert sizes** (Nextera), underrepresentation of the AT-rich sequences (SOLiD) and unplanned spontaneous secondary fragmentation (Roche).

- *Sequencing errors* (Abnizova et al., 2012; Ross et al., 2013) is another threat for aligners. Clearly, if a read encloses more mismatches than allowed by aligner settings, than it will be discarded, even if it accommodates biological signal.

Another objection significant discordance of assemblers (Magoc et al., 2013): different assemblers yield very unequal amount of assembled reads for the same data sets, specifically for homologous genome regions.

3. Post-mapping/assembly QC and re-calibration

Mapping is known (Li H. et al., 2009) to be the a primary cause of sequencing biases. Therefore it is recommended that one reviews the quality of mapped reads before in-depth scientific analysis.

3.1. Mapping metrics

To safeguard an adequate aligners' performance, there are several QC metrics:

- ◇ Number and proportion of high quality reads mapped;
- ◇ Coverage uniformity and average/median depth of coverage;
- ◇ Quality-associated metrics: mapping quality, base quality score distribution, etc.;
- ◇ Insert size distribution; most frequent *k*-mers;
- ◇ Protocol-linked metrics: fraction of discordant pairs; capture efficiency;
- ◇ Forward and reverse read strand symmetry;
- ◇ Error rate per cycle;
- ◇ Possible contaminations metrics, see Chapter 4; GC-bias metrics;
- ◇ For a confident variant call two metrics are of great importance: sequencing coverage depth and uniformity.

Ideally, an even read coverage is expected along genome, to escape local biases. On the other hand, coverage is known (Minoche et al., 2011) to be non-uniform along genome, depending on the regional function, composition (Rieber et al., 2013) and many other features.

- ◇ The Q-value/score is a commonly used measure of base call quality (Bonfield, Staden, 1995; Ewing et al., 1998). The quality Q-scores compress different types of information about the quality of base calls into a confidence (of error) value. Quality score is commonly accepted input for majority of analysis tools, assemblers and aligners in order to produce accurate results.

However, in a raw fastq/bam files these Qs are inferred or predicted. The predictions are based on a set of measurements of a base call, and on previous observations of the values of the measurements. The inferred Q-values are assigned by the means of pre-computed look up table, so called calibration table (Brockman et al., 2008; Abnizova et al., 2010).

A sequencer's errors are typically of low Q, and come from technological and hardware shortcomings.

The infamous sources of errors for Illumina sequencers are: phasing and pre-phasing, dye label X-talk, molecule degradation with time and G-quenching (IDT, 2011). The phase inaccuracy results from base-incorporation errors on a sequencer machine. A G-quenching is an effect of previous nucleotide G; a base quality is typically low for this G-preceded base call (Abnizova et al., 2010, 2012). It was strongly pronounced for the v3 version HiSeq, and dramatically reduced for HiSeqX10 and X5.

- ◇ Contaminated sequences (due to different reasons) may bring up artefacts during variant calling (Schmieder, Edwards, 2011).
- ◇ A capture efficiency for exome sequencing is a proportion of useful data (Garcia-Garcia et al., 2016). It is normally 40–75 % (Guo et al., 2014a), and should not be too small for statistically sound results.

And likewise to the section 1, any inconsistency with expected values for a sample investigated should be cautioning.

3.2. Assembly metrics

In the non-existence of reference genome, the assembly metrics are:

- ◇ Total number of contigs or scaffolds: the less the better;
- ◇ Contig or scaffolds sizes: max, mean and N50. N50 is defined as the length of the scaffold/contig, which overlaps the midpoint of length-ordered concatenation of scaffolds/contigs;

- ◇ Total size of scaffolds. It should be close to an expected size of a genome sequenced;
- ◇ Number of Ns should be limited. (The created gaps in assembly are filled with the uninformative base-pair character 'N'.)

An assembly accuracy and several normalised metrics are possible to assess in case when a reference genome exists. Note that normalization accounts only on those parts of assembly, which can be mapped to a reference genome by standard local alignment tools.

- ◇ Sensitivity of assembly is defines as a percent of genome assembled.
- ◇ Normalised N50 for contigs and for scaffolds is more complicated than for contigs because of N-filler of gaps (Mäkinen et al., 2012).

3.3. Q re-calibration

A predicted Qs often do not correspond to an actual Qs for a certain run/lane/library. In this incident (and in case heterogeneous data are combined) it is suggested to re-calibrate the data (Ewing et al., 1998; Massingham, Goldman, 2012). In the WTSI we implemented the in-house recalibration and error analysis tools (Abnizova et al., 2010). Instead of trimming an ambiguous base calls, we warn (low Q) about possible sequencing errors. Trustworthy Q-value is known to increase SNP call accuracy (Li, Stoneking, 2012) more than hard filtering.

4. Variant calling and its QC

Variant calling from NGS data is defined as a computational methods for establishing an event of genetic variant resulting from NGS experiments (Lawrence, 2014; Zhang et al., 2015).

Variant calling involves small-range variants (Kojima et al., 2013), such as single nucleotide polymorphisms (SNPs), short insertions and deletions (indels), and large-range structural variants, copy number variants (CNV) and structural variants (SV). A SVs are inversions, translocations, or large indels. All types of variants are identified by comparison to a reference genome.

Fraction of variation in genomes is significant: e.g. for human genome, SNPs comprise around 0.1 %, although SV's contribution is estimated as 1.2 % (Tattini et al., 2015) and CNV's contribution is as large as 15 % (Wong et al., 2010).

A variant calling is crucial for comparative genomics and genetics of human diseases. A valuable variant calling application is clinical testing: identifying disease-associated mutations (Chin et al., 2013).

Variant calls are implemented in two ways: (i) after aligning reads, or (ii) after assembling reads. Sometimes these steps are combined. SNPs and small indels can be identified by alignment of short sequencing reads to a reference genome. However, larger structural variants and repetitive regions in the genome are harder to find.

Structural variation can disturb genes or regulatory elements, therefore whole-genome sequencing is not complete without assembly and detection of structural variation (Li H. et al., 2009). In the (i) case, a position of each read relative to the reference genome is identified first. After reads are aligned,

a set of QC steps, involving recalibration, duplicate removing, and indel-realignment, are done before variant calling.

In the (ii) case, an assembly of un-processed reads is performed first, and only after this the assembly is set against a reference genome (if the later exists). Variant detection after assembling is beneficial to individual genes (Olson et al., 2015), but it loses power when applied to a whole genome: in the absence of a reference genome it is not possible to identify other genome's contaminations; spurious variants can not be verified by raw reads after assembling.

Somatic versus germline mutation

Variant calling from NGS is well utilised in genetics of human diseases. There are three typical ways how NGS data is applied in the area: (a) detection of causal germline mutations in Mendelian disorders (Lettice et al., 2008; Stitzel et al., 2011); (b) detection of putative genes for complex diseases with GWAS (Day-Williams, Zeggini, 2011; Lander, 2011; Marian, 2012); (c) detection of somatic and constitutional mutations in cancer (Walther et al., 2009).

It is more complicated to identify a somatic mutation than a germline mutation (Pabinger et al., 2014).

To identify somatic mutations in cancer, they typically compare tumor vs/and normal samples for the same individual (Vissers et al., 2011; Yan et al., 2011).

An annotation step is commonly performed after a variant call (Li R. et al., 2009; Wang et al., 2010; Yang, Wang, 2015). Annotation is done by utilising a public databases, e. g. dbSNP, dbVar (Lin et al., 2015). The next step is visualization (Nielsen et al., 2010).

A set of metrics to assess a quality of variant call is listed below (Guo et al., 2014b; Jun et al., 2015):

- ♦ **Ti/Tv ratio**, individually for whole genome sequencing (WGS) and whole exome sequencing (WES) (should be 2 and 3);
- ♦ **Heterozygosity ratio**;
- ♦ Number of known and of new SNPs per person: should be not more than 200;
- ♦ Cross species and within species contamination; **genotype** consistency;
- ♦ SNP spatial density; QC per SNP;
- ♦ Strand, cycle, allele balance, reference allele biases; haplotypes scores;
- ♦ Performance metrics: Sensitivity and specificity of single nucleotide variant call.

One can combine these metrics by a machine learning methods (DePristo et al., 2011; Jun et al., 2015). In order to minimise false positive (FP), some variant callers do a lot of filtering and trimming using metrics above: by applying a minimum depth of coverage threshold, by masking of homo-polymers and repeats, by trimming poor quality bases from a read etc. Unfortunately, while reducing FP, one can increase false negative (FN) by applying these filters (Olson et al., 2015).

♦ To assess a goodness of a variant caller, one should use a performance metrics: accuracy, sensitivity and specificity, (Olson et al., 2015) given a reliable benchmarking test sets and reference.

A comprehensive review of post-map QC is performed in (Wyllie, 2013; Guo et al., 2014b). The GATK (DePristo et al., 2011) utilises variant QC metrics for their variant calls, applying genotyping and known SNP information for a variant

QC and annotation. However, there seems to be no a standard evaluation of a variant caller (Olson et al., 2015) so far.

5. Correction of errors

A definite amount of errors is the result of sequencing and post-processing imperfections. One way to tackle them is to Q-score possible known artefacts low, so they would be not used by further analysis. Another way is to correct errors using a knowledge about error sources for various platforms' errors (Edgar, Flyvbjerg, 2015; Olson et al., 2015) and computational biases.

An error correction *after mapping* is correction of a mismatch between sequenced read and a reference. *After/during assembling* error correction is a general agreement of base calls across all reads belonging to the same assembled location.

There are multiple attempts to correct sequencing errors. However, an error correction might introduce new type of errors: mis-correction errors (Yang et al., 2013; Fujimoto et al., 2014). And these errors are more difficult to correct back than technological errors.

A sound comparison of NGS platforms is done in (Yang et al., 2013) together with very good explanation of modern error-correction methods. Surprisingly, the paper is very convincing that one should NOT introduce new mis-correcting errors. Additionally, it also does not look promising to correct reads without understanding causes of sequencing/library errors. The work (Fujimoto et al., 2014) confirms that error correction methods can not handle heterozygosity, and they introduce new mis-correction errors.

There are approaches to correct for known context biases, such as GGGGT error patterns for Illumina (Minoche et al., 2011; Nakamura et al., 2011). However, new Illumina releases (e. g. HiSeqX10) are almost free from old type motif-dependency, and new artefacts (such as larger context dependence on a next base) appear.

Error models are used in (Janin et al., 2014) to realistically simulate individual sequencing runs and/or technologies. These models are mostly empirically derived and context-based. A comparison of genomes without assembling them is introduced by (Patro, Kingsford, 2015).

It might be beneficial to do so for de novo sequenced genomes. However, possible PCR biases in coverage are not included in the model. Some studies, such as (Orton et al., 2015) developed a computational error model of Illumina's sample processing, which involves experimental steps. This model infers possible genomic genome locations of PCR errors.

As a conclusion, one should be informed of possible biases, and make decisions depending on their study's aim. Overall conclusion is in necessity to use short sequencing reads error correction for the mapping and processing NGS data, depending on sequencing platforms. Details of error corrections publications will be presented in next paper.

Acknowledgement

Authors are grateful to M.P. Ponomarenko, G.V. Orlova, G.V. Vasiliev and M. Chen for science discussion. Development and testing of genomic computer pipelines was supported by Russian Science Foundation (project 14-24-00123). Application of software for plant transcriptomics was supported by RFBR (16-54-53064).

Glossary

The well-known **Sanger sequencing method** (Sanger et al., 1992) is called a first-generation DNA sequencing technology. The next generation sequencing technologies (Liu et al., 2012) include: (i) 2nd generation sequencing, the massive parallel sequencing of relatively short DNA fragments (Dolled-Filhart et al., 2013); and (ii) 3d generation sequencing, in which single DNA molecules hence much longer fragments (Schadt et al., 2010) are sequenced. In this paper we will focus on 2nd generation DNA sequencing, and will omit the term '2nd generation' while mentioning NGS further.

With NGS technologies, bases are inferred from light/chemistry intensity signals, a process commonly referred to as **base-calling**. The sequenced bases are assigned A, C, G or T letters depending on the intensity.

The **Q-value/score** is the most well accepted measure of base call quality (Bonfield, Staden, 1995).

The quality Q-scores compress a variety of types of information about the quality of base calls into a probability-of-error value.

Mapping or aligning is the matching of the reads to locations at the reference genome. This is done by aligning reads to stretches of the reference genome to which they are most similar in terms of nucleotide sequence.

A **sequence assembly** refers to aligning and merging short fragments from a DNA sequence in order to reconstruct the original sequence.

If the genome of a species has not been sequenced before, the assembly of the reads results in the first version of its reference genome. This is called "**de-novo assembly**".

Multiplex is a library containing various samples labelled with bar codes.

Sample **multiplexing** is a useful technique when targeting specific genomic regions or working with smaller genomes. To accomplish this, individual "**barcode**" sequences are added to each sample so they can be distinguished and sorted during data analysis. **Pooling** samples exponentially increases the number of samples analyzed in a single run, without drastically increasing cost or time.

Di-multiplexing is separating samples based on their tags, ideally should be even across tags.

Adapter. The vast majority of next-generation sequencing experiments will attach adapter sequence to the sequencing construct. In many cases these are standard sequences that can be obtained from the vendor and/or sequencing centre. Unfortunately sometimes adapter information is not properly tracked and attached as metadata to the raw sequencing data and may not be known for a given sample.

PF (purity-filtered) data: PF-filtering is known as throwing away data with low maximum intensity signal (purity, Illumina terminology).

GC-content is a measure of the relative frequency of the cytosine (C) and guanine (G) bases, in comparison with the adenine (A) and thymine (T) bases. A genome is called **GC-rich** if significantly more than 50 % of its bases are G or C.

Mate-pair libraries. *Mate-pair* is different from "*paired-end*" in the sense of how the sequence library is made. In "*Mate-pair*" sequencing, 2–5 kb fragments are selected and sequenced from both end, thus giving information how nucleotides far apart are linked together. *Mate-pairs* are more ideal for studying genomic structural rearrangement and help de novo genome assembly. They also facilitate sensitive structural variant (SV) detection across a widened SV size-spectrum and in repetitive areas of the genome.

Insert size = DNA fragment size.

Ti/Tv (sometimes called Ts/Tv): the ratio of transitions vs. transversions in SNPs. Transitions are mutations within the same type of nucleotide: pyrimidine-pyrimidine mutations (C ↔ T) and purine-purine mutations (A ↔ G). Transversions are mutations from a pyrimidine to a purine or vice versa.

The **heterozygosity ratio** is the number of heterozygous sites in an individual divided by the number of non-reference homozygous sites.

Error-correction is an attempt to correct a mismatch between sequenced reads and/or reference (if it is available).

Genomic variant or mutation is a permanent alteration of the nucleotide sequence of the genome of an organism.

A **single nucleotide polymorphism** or simple nucleotide polymorphism, (SNP), is a variation in a single nucleotide which may occur at some specific position in the genome, where each variation is present to some appreciable degree within a population (e.g. >1 %).

Structural variation (also genomic structural variation) is the variation in structure of an organism's chromosome.

Structural variation consists of many kinds of variation in the genome of one species, and usually includes microscopic and sub-microscopic types, such as deletions, duplications, copy-number variants, insertions, inversions and translocations.

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule.

WGS – whole genome sequencing.

WES – whole exome sequencing.

In Illumina, **PCR and size selection** steps have been implicated in GC-bias. PCR is known to preferentially amplify GC-moderate sequences, while size selection involves DNA heating which leads to a GC-poor fragment's underrepresentation. Avoiding these steps helps to limit the GC-bias.

BAM File – binary version of SAM file, a typical output of the secondary phase of data analysis.

Coverage – this value indicates the coverage of an analysed sequence with respect to its length, usually expressed as a percentage; sometimes the term is also used for the depth of reading.

Long-Reads – strategy for sequencing samples prepared by *Mate-Pair-End* method.

Mate Pair-End-Read – strategy for sample preparation where the longer fragment (thousands of bases) is circularized using labelled adapters, the molecule is subsequently fragmented, but only the fragments containing the labelled adapters are sequenced.

Paired-End-Read – a method of reading a fragment where the fragment is first read from one end and then from the other.

Read Depth – DNA = number of times a nucleotide is read; RNA = total number of reads per sample.

Read Length – the number of read bases per fragment, respectively the maximum length of the fragment, which can be sequenced at a time (indicated in bases).

Single-Read – a method of reading a fragment where the fragment is read from one end only during sequencing.

SNP – Single-Nucleotide Polymorphism = sequence divergence in the range of a single base.

SNP Calling – process of detecting SNPs in the sequences obtained.

Variant Calling is a process of detection of sequence variants in the sequences obtained.

Heterozygosity occurs when an individual has two different alleles of a gene/loci.

Chimeric reads are reads with DNA sequences originating from two different samples.

Conflict of interest

The authors declare no conflict of interest.

References

- Abnizova I., Leonard S., Skelly T., Brown A., Jackson D., Gourtovaia M., Qi G., Te Boekhorst R., Faruque N., Lewis K., Cox T. Analysis of context-dependent errors for illumina sequencing. *J. Bioinform. Comput. Biol.* 2012;10(2):1241005.
- Abnizova I., Skelly T., Naumenko F., Whiteford N., Brown C., Cox T. Statistical comparison of methods to estimate the error probability in short-read Illumina sequencing. *J. Bioinform. Comput. Biol.* 2010;8(3):579-591.
- Albers C.A., Lunter G., MacArthur D.G., McVean G., Ouwehand W.H., Durbin R. Dindel: accurate indel calls from short-read data. *Genome Res.* 2011;21(6):961-973.
- Alkan C., Cardone M.F., Catacchio C.R., Antonacci F., O'Brien S.J., Ryder O.A., Purgato S., Zoli M., Della Valle G., Eichler E.E., Ventura M. Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Res.* 2011a;21(1):137-145.
- Alkan C., Sajjadian S., Eichler E.E. Limitations of next-generation genome sequence assembly. *Nat. Methods.* 2011b;8(1):61-65.
- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J. Mol. Biol.* 1990;215(3):403-410.
- Anders S., Pyl P.T., Huber W. HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166-169.
- Auerbach R.K., Euskirchen G., Rozowsky J., Lamarre-Vincent N., Moqtaderi Z., Lefrancois P., Struhl K., Gerstein M., Snyder M. Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. USA.* 2009;106(35):14926-14931.
- Baker M. De novo genome assembly: what every biologist should know. *Nature Methods.* 2012;9.
- Balint B. Decreased sequencing accuracy at the 3' end of SBS Illumina Reads. 2016.
- Benjamini Y., Speed T.P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 2012;40(10):e72.
- Bonfield J.K., Staden R. The application of numerical estimates of base calling accuracy to DNA sequencing projects. *Nucleic Acids Res.* 1995;23(8):1406-1410.
- Bragg L.M., Stone G., Butler M.K., Hugenholtz P., Tyson G.W. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 2013;9(4):e1003031.
- Brockman W., Alvarez P., Young S., Garber M., Giannoukos G., Lee W.L., Russ C., Lander E.S., Nusbaum C., Jaffe D.B. Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res.* 2008;18(5):763-770.
- Chen Y.C., Liu T., Yu C.H., Chiang T.Y., Hwang C.C. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One.* 2013; 8(4):e62856.
- Chin E.L.H., da Silva C., Hegde M. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genet.* 2013;14:6.
- Chin F.Y.L., Leung H.C.M., Yiu S.M. Sequence assembly using next generation sequencing data – challenges and solutions. *Sci. China Life Sciences.* 2014;57(11):1140-1148.
- Cox M.P., Peterson D.A., Biggs P.J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics.* 2010;11:485.
- Davis M.P., van Dongen S., Abreu-Goodger C., Bartonicek N., Enright A.J. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods.* 2013;63(1):41-49.
- Day-Williams A.G., Zeggini E. The effect of next-generation sequencing technology on complex trait research. *Eur. J. Clin. Invest.* 2011; 41(5):561-567.
- de Koning A.P., Gu W., Castoe T.A., Batzer M.A., Pollock D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 2011;7(12):e1002384.
- Del Fabbro C., Scalabrin S., Morgante M., Giorgi F.M. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One.* 2013;8(12):e85024.
- DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., del Angel G., Rivas M.A., Hanna M., McKenna A., Fennel T.J., Kernysky A.M., Sivachenko A.Y., Cibulskis K., Gabriel S.B., Altshuler D., Daly M.J. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 2011;43(5):491-498.
- Dolled-Filhart M.P., Lee M., Jr., Ou-Yang C.W., Haraksingh R.R., Lin J.C. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *Sci. World J.* 2013;730210.
- Edgar R.C., Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics.* 2015; 31(21):3476-3482.
- Ewing B., Hillier L., Wendl M.C., Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 1998;8(3):175-185.
- Feschotte C., Jiang N., Wessler S.R. Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 2002;3(5):329-341.
- Fonseca N.A., Rung J., Brazma A., Marioni J.C. Tools for mapping high-throughput sequencing data. *Bioinformatics.* 2012;28(24):3169-3177.
- Fujimoto M., Bodily P.M., Okuda N., Clement M.J., Snell Q. Effects of error-correction of heterozygous next-generation sequencing data. *BMC Bioinformatics.* 2014;15(Suppl.7):S3.
- Garcia-Garcia G., Baux D., Faugere V., Moclyn M., Koenig M., Claustres M., Roux A.F. Assessment of the latest NGS enrichment capture methods in clinical context. *Sci. Rep.* 2016;6:20948.
- Guo Y., Ye F., Sheng Q., Clark T., Samuels D.C. Three-stage quality control strategies for DNA re-sequencing data. *Brief Bioinform.* 2014a;15(6):879-889.
- Guo Y., Zhao S., Sheng Q., Ye F., Li J., Lehmann B., Pietenpol J., Samuels D.C., Shyr Y. Multi-perspective quality control of Illumina exome sequencing data using QC3. *Genomics.* 2014b;103(5-6):323-328.
- Hadfield J. Quality control for your NGS data. 2013.
- Harismendy O., Ng P.C., Strausberg R.L., Wang X., Stockwell T.B., Beeson K.Y., Schork N.J., Murray S.S., Topol E.J., Levy S., Frazer K.A. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* 2009;10(3):R32.
- Huang Y.F., Chen S.C., Chiang Y.S., Chen T.H., Chiu K.P. Palindromic sequence impedes sequencing-by-ligation mechanism. *BMC Syst. Biol.* 2012;6(Suppl.2):S10.
- Huddleston J., Ranade S., Malig M., Antonacci F., Chaisson M., Hon L., Sudmant P.H., Graves T.A., Alkan C., Dennis M.Y., Wilson R.K., Turner S.W., Korlach J., Eichler E.E. Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res.* 2014;24(4):688-696.
- IDT. 2011. G-sequencing.
- Ignatieva E.V., Podkolodnaya O.A., Orlov Y.L., Vasiliev G.V., Kolchanov N.A. Regulatory genomics: Integrated experimental and computer approaches. *Genetika.* 2015;51(4):409-429.
- Illumina. 2014. Sequencing Library QC on the MiSeq® System.
- Janin L., Schulz-Trieglaff O., Cox A.J. BEETL-fastq: a searchable compressed archive for DNA reads. *Bioinformatics.* 2014;30(19):2796-2801.
- Jiang H., Lei R., Ding S.W., Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics.* 2014;15:182.
- Jun G., Wing M.K., Abecasis G.R., Kang H.M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* 2015;25(6):918-925.
- Kelley D.R., Schatz M.C., Salzberg S.L. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.* 2010;11(11):R116.
- Kojima K., Nariai N., Mimori T., Takahashi M., Yamaguchi-Kabata Y., Sato Y., Nagasaki M. A statistical variant calling approach from pedigree information and local haplotyping with phase informative reads. *Bioinformatics.* 2013;29(22):2835-2843.

- Lander E.S. Initial impact of the sequencing of the human genome. *Nature*. 2011;470(7333):187-197.
- Lawrence M. Introduction to variant calling. *Bioconductor*. 2014.
- Ledergerber C., Dessimoz C. Base-calling for next-generation sequencing platforms. *Brief Bioinform.* 2011;12(5):489-497.
- Lettice L.A., Hill A.E., Devenney P.S., Hill R.E. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* 2008;17(7):978-985.
- Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
- Li H., Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform.* 2010;11(5):473-483.
- Li J.W., Robison K., Martin M., Sjodin A., Usadel B., Young M., Olivares E.C., Bolser D.M. The SEQanswers wiki: a wiki database of tools for high-throughput sequencing analysis. *Nucleic Acids Res.* 2012a;40(Database iss.):D1313-1317.
- Li J.W., Schmieder R., Ward R.M., Delenick J., Olivares E.C., Mittelman D. SEQanswers: an open access community for collaboratively decoding genomes. *Bioinformatics*. 2012b;28(9):1272-1273.
- Li M.K., Stoneking M. A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biology*. 2012;13(5).
- Li R., Yu C., Li Y., Lam T.W., Yiu S.M., Kristiansen K., Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009;25(15):1966-1967.
- Li S., Li R., Li H., Lu J., Li Y., Bolund L., Schierup M.H., Wang J. SOAPindel: efficient identification of indels from short paired reads. *Genome Res.* 2013;23(1):195-200.
- Li Z., Chen Y., Mu D., Yuan J., Shi Y., Zhang H., Gan J., Li N., Hu X., Liu B., Yang B., Fan W. Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Brief Funct. Genomics*. 2012;11(1):25-37.
- Lin K., Smit S., Bonnema G., Sanchez-Perez G., de Ridder D. Making the difference: integrating structural variation detection tools. *Brief Bioinform.* 2015;16(5):852-864.
- Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L., Law M. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012;251364.
- Lo C.C., Chain P.S. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*. 2014;15:366.
- Magoc T., Pabinger S., Canzar S., Liu X., Su Q., Puiu D., Tallon L.J., Salzberg S.L. GAGE-B: an evaluation of genome assemblers for bacterial organisms. *Bioinformatics*. 2013;29(14):1718-1725.
- Makinen V., Salmela L., Ylinen J. Normalized N50 assembly metric using gap-restricted co-linear chaining. *BMC Bioinformatics*. 2012;13:255.
- Marian A.J. Molecular genetic studies of complex phenotypes. *Transl. Res.* 2012;159(2):64-79.
- Marroni F., Pinosio S., Morgante M. The quest for rare variants: pooled multiplexed next generation sequencing in plants. *Front Plant. Sci.* 2012;3:33.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*. 2011;17(1).
- Massingham T., Goldman N. All your Base: a fast and accurate probabilistic approach to base calling. *Genome Biology*. 2012;13(2).
- McCoy R.C., Taylor R.W., Blauwkamp T.A., Kelley J.L., Kertesz M., Pushkarev D., Petrov D.A., Fiston-Lavier A.S. Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLoS One*. 2014;9(9):e106689.
- Medvedev P., Stanciu M., Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods*. 2009;6(Suppl.11):S13-20.
- Miller J.R., Koren S., Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics*. 2010;95(6):315-327.
- Minoche A.E., Dohm J.C., Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* 2011;12(11):R112.
- Mir K., Neuhaus K., Bossert M., Schober S. Short barcodes for next generation sequencing. *PLoS One*. 2013;8(12):e82933.
- Mutarelli M., Marwah V., Rispoli R., Carrella D., Dharmalingam G., Oliva G., di Bernardo D. A community-based resource for automatic exome variant-calling and annotation in Mendelian disorders. *BMC Genomics*. 2014;15(Suppl.3):S5.
- Nagarajan N., Pop M. Sequence assembly demystified. *Nat. Rev. Genet.* 2013;14(3):157-167.
- Nakamura K., Oshima T., Morimoto T., Ikeda S., Yoshikawa H., Shima Y., Ishikawa S., Linak M.C., Hirai A., Takahashi H., Altaf-Ul-Amin Md., Ogasawara N., Kanaya S. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* 2011;39(13):e90.
- Newell F. NGS mapping, errors and quality control. Australia: Univ. of Queensland, 2014.
- Nielsen C.B., Cantor M., Dubchak I., Gordon D., Wang T. Visualizing genomes: techniques and challenges. *Nat. Methods*. 2010;7(Suppl.3):S5-S15.
- Niu B., Fu L., Sun S., Li W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics*. 2010;11:187.
- Olson N.D., Lund S.P., Colman R.E., Foster J.T., Sahl J.W., Schupp J.M., Keim P., Morrow J.B., Salit M.L., Zook J.M. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Front. Genet.* 2015;6:235.
- Orlov Y.L., Te Boekhorst R., Abnizova I.I. Statistical measures of the structure of genomic sequences: entropy, complexity, and position information. *J. Bioinform. Comput. Biol.* 2006;4:523-36.
- Orton R.J., Wright C.F., Morelli M.J., King D.J., Paton D.J., King D.P., Haydon D.T. Distinguishing low frequency mutations from RT-PCR and sequence errors in viral deep sequencing data. *BMC Genomics*. 2015;16:229.
- Otto C., Stadler P.F., Hoffmann S. Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics*. 2014;30(13):1837-1843.
- Pabinger S., Dander A., Fischer M., Snajder R., Sperk M., Efremova M., Krabichler B., Speicher M.R., Zschocke J., Trajanoski Z. A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform.* 2014;15(2):256-278.
- Park N., Shirley L., Gu Y., Keane T.M., Swerdlow H., Quail M.A. An improved approach to mate-paired library preparation for Illumina sequencing. *Methods Next-Generation Sequencing*. 2013;1(1):10-20.
- Patel R.K., Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012;7(2):e30619.
- Patro R., Kingsford C. Data-dependent bucketing improves reference-free compression of sequencing reads. *Bioinformatics*. 2015;31(17):2770-2777.
- Pightling A.W., Petronella N., Pagotto F. Choice of reference sequence and assembler for alignment of *Listeria monocytogenes* short-read sequence data greatly influences rates of error in SNP analyses. *PLoS One*. 2014;9(8):e104579.
- Pireddu L., Leo S., Zanetti G. SEAL: a distributed short read mapping and duplicate removal tool. *Bioinformatics*. 2011;27(15):2159-2160.
- Rieber N., Zaparka M., Lasitschka B., Jones D., Northcott P., Hutter B., Jäger N., Kool M., Taylor M., Lichter P., Pfister S., Wolf S., Brors B., Eils R. Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. *PLoS One*. 2013;8(6):e66621.
- Ross M.G., Russ C., Costello M., Hollinger A., Lennon N.J., Hegarty R., Nusbaum C., Jaffe D.B. Characterizing and measuring bias in sequence data. *Genome Biol.* 2013;14(5):R51.
- Safronova N.S., Babenko V.N., Orlov Y.L. 117 analysis of SNP containing sites in human genome using text complexity estimates. *J. Biomol. Structure Dynamics*. 2015;33(Suppl.1):73-74. DOI 10.1080/07391102.2015.1032750.

- Safronova N.S., Ponomarenko M.P., Abnizova I.I., Orlova G.V., Chadaeva I.V., Orlov Y.L. Flanking monomer repeats determine decreased context complexity of single nucleotide polymorphism sites in the human genome. *Rus. J. Genet. Appl. Res.* 2016;6(8):809-815. DOI 10.1134/S2079059716070121.
- Sameith K., Roscito J.G., Hiller M. Iterative error correction of long sequencing reads maximizes accuracy and improves contig assembly. *Brief Bioinform.* 2016;1-8.
- Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. (1977). *Biotechnology.* 1992;24:104-108.
- Schadt E.E., Turner S., Kasarskis A. A window into third-generation sequencing. *Hum. Mol. Genet.* 2010;19(R2):R227-240.
- Schmieder R., Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One.* 2011;6(3):e17288.
- Schwartz S., Oren R., Ast G. Detection and removal of biases in the analysis of next-generation sequencing reads. *PLoS One.* 2011;6(1):e16685.
- Shang J., Zhu F., Vongsangnak W., Tang Y., Zhang W., Shen B. Evaluation and comparison of multiple aligners for next-generation sequencing data analysis. *Biomed. Res. Int.* 2014;309650.
- Sharon D., Tilgner H., Grubert F., Snyder M. A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* 2013; 31(11):1009-1014.
- Stitzel N.O., Kiezun A., Sunyaev S. Computational and statistical approaches to analyzing variants identified by exome sequencing. *Genome Biol.* 2011;12(9):227.
- Tattini L., D'Aurizio R., Magi A. Detection of genomic structural variants from next-generation sequencing data. *Front Bioeng. Biotechnol.* 2015;3:92.
- van Dijk E.L., Auger H., Jaszczyszyn Y., Thermes C. Ten years of next-generation sequencing technology. *Trends Genet.* 2014;30(9):418-426.
- Vissers L.E., Fano V., Martinelli D., Campos-Xavier B., Barbuti D., Cho T.J., Dursun A., Kim O.H., Lee S.H., Timpani G., Nishimura G., Unger S., Sass J.O., Veltman J.A., Brunner H.G., Bonafe L., Dionisi-Vici S., Superti-Furga A. Whole-exome sequencing detects somatic mutations of IDH1 in metaphyseal chondromatosis with D-2-hydroxyglutaric aciduria (MC-HGA). *Am. J. Med. Genet. A.* 2011;155A(11):2609-2616.
- Voskoboinik A., Neff N.F., Sahoo D., Newman A.M., Pushkarev D., Koh W., Passarelli B., Fan H.C., Mantalas G.L., Palmeri K.J., Ishizuka K.J., Gissi C., Griggio F., Ben-Shlomo R., Corey D.M., Penland L., White R.A.III, Weissman I.L., Quake S.R. The genome sequence of the colonial chordate, *Botryllus schlosseri*. *ELife.* 2013;2:e00569.
- Walther A., Johnstone E., Swanton C., Midgley R., Tomlinson I., Kerr D. Genetic prognostic and predictive markers in colorectal cancer. *Nat. Rev. Cancer.* 2009;9(7):489-499.
- Wang K., Li M., Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
- Wolfinger M.T., Fallmann J., Eggenhofer F., Amman F. ViennaNGS: A toolbox for building efficient next-generation sequencing analysis pipelines. *F1000Res.* 2015;4:50.
- Wong K., Keane T.M., Stalker J., Adams D.J. Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. *Genome Biology.* 2010;11(12).
- Wyllie M. Comprehensive analysis of clinical trials data shows unequivocally that Phosphodiesterase Inhibitors (PDEi) improve orgasm. The power of meta-analysis? *BJU Int.* 2013;111(2):190-191.
- Yan X.J., Xu J., Gu Z.H., Pan C.M., Lu G., Shen Y., Shi J.Y., Zhu Y.M., Tang L., Zhang X.W., Liang W.-X., Mi J.-Q., Song H.-D., Li K.-Q., Chen Z., Chen S.-J. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet.* 2011;43(4):309-315.
- Yang H., Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* 2015;10(10):1556-1566.
- Yang X., Chockalingam S.P., Aluru S. A survey of error-correction methods for next-generation sequencing. *Brief Bioinform.* 2013; 14(1):56-66.
- Ye L., Hillier L.W., Minx P., Thane N., Locke D.P., Martin J.C., Chen L., Mitreva M., Miller J.R., Haub K.V., Dooling D.J., Mardis E.R., Wilson R.K., Weinstock G.M., Warren W.C. A vertebrate case study of the quality of assemblies derived from next-generation sequences. *Genome Biol.* 2011;12(3):R31.
- Zhang W., Ng H.W., Shu M., Luo H., Su Z., Ge W., Perkins R., Tong W., Hong H. Comparing genetic variants detected in the 1000 genomes project with SNPs determined by the International HapMap Consortium. *J. Genetics.* 2015;94(4):731-740.

Определение числа копий эндогенных ретровирусов типа А у домашних свиней и диких кабанов

Р.Б. Айтназаров^{1, 2}, Н.С. Юдин^{1, 2, 3} , Р.С. Кирилчук², Н.Н. Кочнев⁴, С.П. Князев⁴, М.И. Воевода^{1, 2, 3}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт терапии и профилактической медицины» Сибирского отделения Российской академии медицинских наук, Новосибирск, Россия

⁴ Федеральное государственное бюджетное образовательное учреждение высшего образования «Новосибирский государственный аграрный университет», Новосибирск, Россия

Современная трансплантология испытывает острейший недостаток в донорских органах. Для решения проблемы был предложен подход, основанный на использовании органов и тканей животных для пересадки человеку (ксенотрансплантация). Однако широкому развитию этого направления препятствует риск передачи человеку зоонозных инфекционных заболеваний. По экономическим и этическим критериям, а также благодаря сходству по анатомическим и физиологическим параметрам, свинья является наиболее оптимальным источником органов для ксенотрансплантации. В геноме свиньи содержатся эндогенные ретровирусы (PERV) типа А которые могут инфицировать клеточные линии человека *in vitro*. Специально для биомедицинских целей, в том числе ксенотрансплантации, в Институте цитологии и генетики СО РАН была выведена селекционная группа сибирских миниатюрных свиней. Цель работы заключалась в анализе числа копий PERV А у сибирских миниатюрных свиней, пород-основателей ландрас и крупная белая, а также диких кабанов. Число копий PERV определяли методом абсолютного количественного анализа с использованием красителя SYBR Green. В качестве стандарта использовали образец с известным числом копий, который получали методом конечных разведений. Число копий PERV А в калибровочных образцах ДНК сибирских миниатюрных свиней составило 2.4, 3.6 и 4.9 копии PERV А на клетку, что хорошо совпадает с данными других авторов. Медиана числа копий PERV А на клетку составила 4.5 у сибирских миниатюрных свиней, 1.3 у ландрасов, 1.0 у представителей крупной белой породы и 0.8 у диких кабанов. Достоверные различия в числе копий ретровируса обнаружены между миниатюрными свиньями и дикими кабанями. Таким образом, геном сибирских миниатюрных свиней содержит значительное число копий потенциально патогенных для человека ретровирусов PERV типа А. Для ксенотрансплантации необходимо отбирать животных с наименьшим числом ретровирусов в геноме. Метод количественного определения числа копий PERV А с использованием красителя SYBR Green позволяет выявить таких животных и проводить селекцию сибирских миниатюрных свиней на уменьшение этого показателя.

Ключевые слова: свинья; сибирская миниатюрная свинья; крупная белая порода; порода ландрас; дикий кабан; свиной эндогенный ретровирус; ДНК; ген *envA*; число копий; ксенотрансплантация.

Determination of the copy numbers of type A porcine endogenous retroviruses in domestic pigs and wild boars

R.B. Aitnazarov^{1, 2}, N.S. Yudin^{1, 2, 3} , R.S. Kiril'chuk², N.N. Kochnev⁴, S.P. Knyazev⁴, M.I. Voevoda^{1, 2, 3}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Research Institute of Internal and Preventive Medicine SB RAMS, Novosibirsk, Russia

⁴ Novosibirsk State Agrarian University, Novosibirsk, Russia

Modern transplantology is in need of transplants. To solve this problem, the use of animal organs and tissues for grafting to humans (xenografts) has been proposed. However, the progress in this direction is hampered by the risk of zoonotic infection of recipients. With regard to economic and ethical criteria and to the anatomical and physiological similarity to humans, the pig is the best source of xenografts. The pig genome harbors type A porcine endogenous retroviruses (PERV), which can infect human cell lines *in vitro*. A population of Siberian minipigs was raised at the Institute of Cytology and Genetics just for xenografting. The goal of the present study is to analyze the copy numbers of PERV A in Siberian minipigs, their founder breeds Landrace and Large White, and wild boars. The copy numbers of PERVs have been determined by absolute measurement with SYBR Green dye. End-point dilutions of a sample with a known copy number have been used for reference. The PERV A copy numbers in standard samples of Siberian minipig DNA are 2.4, 3.6, and 4.9 per cell, which is consistent with data obtained by other scientists. Minipigs and wild boars show a significant difference in retrovirus copy numbers. Thus, the Siberian minipig genome has a considerable number of type A PERVs, conceivably pathogenic to humans. It is necessary to select animals with minimum PERV numbers in the genome for xenografting. The method

of PERV A quantitation with SYBR Green allows detection of such animals and selection of Siberian minipigs for reduction of this index.

Key words: pig; Siberian mini pig; Large White breed; Landras breed; wild boar; porcine endogenous retrovirus; DNA; *envA* gene; copy number; xenotransplantation.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Айтназаров Р.Б., Юдин Н.С., Кирилчук Р.С., Кочнев Н.Н., Князев С.П., Воевода М.И. Определение числа копий эндогенных ретровирусов типа А у домашних свиней и диких кабанов. Вавиловский журнал генетики и селекции. 2016;20(6):756-761. DOI 10.18699/VJ16.192

HOW TO CITE THIS ARTICLE:

Aitnazarov R.B., Yudin N.S., Kiril'chuk R.S., Kochnev N.N., Knyazev S.P., Voevoda M.I. Determination of the copy numbers of type A porcine endogenous retroviruses in domestic pigs and wild boars. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):756-761. DOI 10.18699/VJ16.192

Одна из важнейших проблем современной трансплантологии – недостаток донорских органов. Решить эту проблему можно за счет использования для пересадки человеку органов животных (ксенотрансплантация). Свинья считается наиболее вероятным донором органов для ксенотрансплантации в силу анатомического и физиологического сходства с человеком, а также по этическим и экономическим причинам (Ekser et al., 2015). Основным препятствием для широкого внедрения ксенотрансплантации остается риск зоонозных инфекционных заболеваний, передающихся человеку от животных. Существенно снизить риск передачи инфекции при ксенотрансплантации можно путем выращивания свиней в специфических, свободных от патогенов условиях (Shimatsu et al., 2013). Однако при этом сохраняется опасность передачи человеку свинных эндогенных ретровирусов (Porcine Endogenous RetroVirus – PERV). PERV является РНК-содержащим ретровирусом, но в его цикле репликации также используется ДНК. Поэтому PERV способны встраиваться в геном клеток в качестве провирусов и не могут быть удалены путем выращивания свиней в свободных от патогенов условиях (Yudin et al., 2011). По предварительным оценкам, геном свиньи может содержать от 6 до 10 способных к репликации провирусов, от 30 до 50 полноразмерных копий PERV и от 100 до 200 локусов, содержащих их частичную последовательность (Niebert, Tonjes, 2005). Существует риск развития инфекции PERV у реципиентов ксенотрансплантатов, распространения PERV от инфицированных пациентов к тесно контактирующему с ними медперсоналу и далее на общую популяцию (Denner, 2016).

Выделяют три типа ретровирусов PERV – А, В и С, которые сходны друг с другом по последовательностям ДНК генов *gag* (group-specific-antigens) и *pol* (polymerase), но существенно различаются по последовательности рецептор-связывающего домена гена *env* (envelope), кодирующего белок оболочки вируса (Le Tissier et al., 1997; Akiyoshi et al., 1998). PERV А и В, помимо клеточных линий свиньи, могут инфицировать некоторые линии клеток человека *in vitro*, а PERV С может реплицироваться только в клетках свиньи (Kimsa et al., 2014). Исследования больных, которых лечили с использованием живых тканей или органов свиньи, до сих пор не выявили развития инфекции PERV у человека *in vivo* (Godehardt et al., 2015). Однако не ясно, являются эти данные результатом реального отсут-

ствия продукции вируса клетками свиньи или результатом эффективной инактивации высвобождаемых вирусных частиц иммунной системой. Поэтому потенциальная опасность переноса PERV при ксенотрансплантации по-прежнему сохраняется.

Несколько пород лабораторных мини-свиней были выведены за рубежом специально для нужд ксенотрансплантации (Mogozov et al., 2015). Селекционная группа сибирских миниатюрных свиней была создана в Институте цитологии и генетики (ИЦиГ) СО РАН путем скрещивания свиней вьетнамской черной и пятнистой пород, свиней пород ландрас и крупной белой, а также диких кабанов европейского и среднеазиатского подвидов (Тихонов, 2010). В геноме большинства сибирских миниатюрных свиней присутствуют ретровирусы PERV всех трех типов (Айтназаров и др., 2014б). Обогащенность геномов миниатюрных свиней ретровирусами, вероятно, обусловлена их внесением от пород-основателей. Известно, что дикий кабан имеет наименьшее число копий PERV всех типов в геноме по сравнению со свиньями домашних пород, в том числе крупной белой (Mang et al., 2001). Повышение этого показателя у домашних свиней, несомненно, связано с увеличением риска зоонозной инфекции. Однако число копий PERV в геноме сибирских миниатюрных свиней неизвестно.

Целью работы было определение числа копий тропного к клеткам человека ретровируса PERV типа А у сибирских миниатюрных свиней, используемых для ксенотрансплантации, а также у представителей некоторых пород-основателей.

Материалы и методы

Животные. Образцы крови сибирских миниатюрных свиней ($n = 10$) были взяты в Центре коллективного пользования «Генофонды пушных и сельскохозяйственных животных» ИЦиГ СО РАН. Образцы крови свиней пород крупная белая ($n = 6$) и ландрас ($n = 3$) собраны в хозяйствах Новосибирской области. Образцы крови центральноевропейского кабана *Sus scrofa scrofa* ($n = 3$) получены из Воронежского биосферного заповедника, а среднеазиатского кабана ($n = 1$) *Sus scrofa nigripes* – в Чуйской долине (Киргизия). При анализе образцы дикого кабана объединяли в одну группу. ДНК выделяли методом протеолитической обработки с последующей экстракцией фенолом.

Определение числа копий PERV A. Стандартные образцы для количественного анализа числа копий PERV были получены методом конечных разведений (Sedlak, Jerome, 2013). Для этого проводили амплификацию фрагмента гена *envA* в серии последовательных десятикратных или двукратных разведений исходного образца ДНК. Условия амплификации фрагмента гена *envA* ретровирусов PERV размером 356 п. н. были описаны нами ранее (Никитин и др., 2008). Число копий PERV A в стандартных образцах определяли на основании распределения Пуассона по формуле $C = -\ln(1 - p)$, где C – среднее число копий, p – доля реплицированных образцов, в которых прошла амплификация. Результат подтверждали в двух независимых экспериментах.

Далее определение числа копий PERV осуществляли методом абсолютного количественного анализа по графику стандартов с различным числом копий PERV A на приборе Rotor-Gene Q (Qiagen, Нидерланды). Амплификацию выполняли с использованием набора реагентов для проведения ПЦР в реальном времени в присутствии красителя SYBR Green I (Синтол, Россия) по протоколу фирмы-изготовителя. Для каждого образца ДНК ПЦР проводили не менее трех раз. Данные обрабатывали с помощью программы Rotor-Gene 6000 Series software версия 1.8.17.5. При вычислениях принимали, что клетка свиньи содержит то же количество геномной ДНК, что и клетка человека (6 пг ДНК на клетку).

Статистический анализ. Для определения нормальности распределения использовали критерий Колмогорова–Смирнова. В сравниваемых группах вычисляли медиану, 25- и 75%-й квартили количества копий PERV на клетку. Поскольку исследованный признак не подчинялся нормальному распределению, для сравнения показателей между группами применяли непараметрический критерий Краскела–Уоллиса, а попарные сравнения проводили путем апостериорного анализа средних рангов для всех групп. Критическое значение уровня статистической значимости при проверке нулевых гипотез принимали равным 0.05. Обработку результатов проводили с помощью пакета STATISTICA версии 8.0.

Результаты

Методом конечных разведений было определено число копий PERV A в трех образцах ДНК сибирских миниатюрных свиней, которые в дальнейшем использовали для построения калибровочного графика. Исходная концентрация ДНК в образцах варьировала от 1376 до 1972 нг/мкл. При разведении около $1:40 \times 10^{-6}$ наблюдали стохастическую амплификацию реплицированных образцов (рис. 1). Расчеты показали, что образцы ДНК трех мини-свиней содержали 2.4, 3.6 и 4.9 копии PERV A на клетку.

Был разработан метод количественной ПЦР в реальном времени для определения числа копий гена *envA* с использованием в качестве калибровки образцов, в которых число копий определялось методом конечных разведений. Исследования показали, что детекция целевой последовательности ДНК была линейной (средний коэффициент регрессии $R^2 = 0.99$) в диапазоне от 20 копий до 20 тыс. копий на пробу, что показывает высокую чувствительность метода (рис. 2).

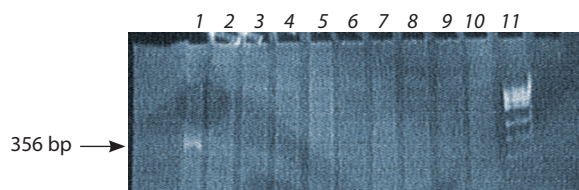


Fig. 1. Determination of the copy number of the *envA* gene by the end-point dilution method in a miniature pig DNA sample.

Lanes 1–9 are amplified products from nine replicate dilutions $1:40 \times 10^{-6}$ of the same DNA sample; lane 10, negative control (water); lane 11, DNA ladder from 100 to 1000 bp, at 100-bp increments.

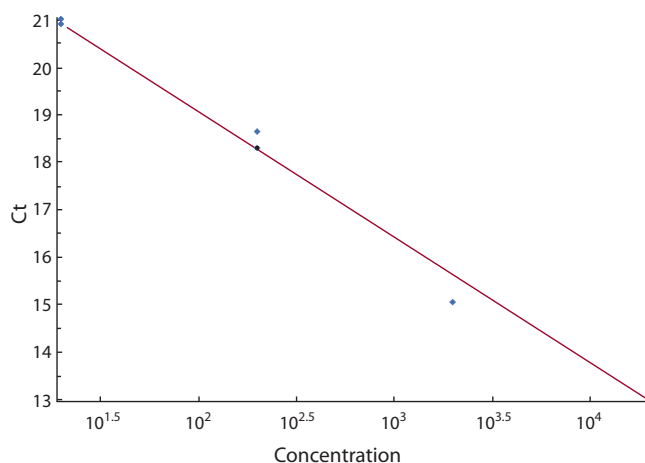


Fig. 2. Typical calibration curve for the *envA* gene.

A standard miniature pig DNA sample was diluted to four concentrations: 20000, 2000, 200, and 20 copies. Real-time PCR was run individually, and the corresponding Ct values were measured. The calibration curve for *envA* was constructed by plotting the Ct values (Y-axis) on the scale of logarithmic copy numbers in standard samples (X-axis).

С помощью метода количественной ПЦР в реальном времени было проведено сравнение числа копий гена *envA* ретровируса PERV в образцах ДНК крови у домашних свиней и диких кабанов (см. таблицу). Оказалось, что число копий на клетку у представителей этих групп достоверно различается ($H = 7.88$; $p < 0.05$). При попарном сравнении выявлены достоверные различия в числе копий ретровируса между миниатюрными свиньями и дикими кабанками ($p < 0.05$).

Обсуждение

Острый дефицит тканей и органов в трансплантологии привел к внедрению ксенотрансплантации органов от свиньи к человеку. Однако широкое развитие этого метода сдерживается возможностью передачи человеку патогенов свиньи, в особенности PERV. Передача PERV человеку может быть отправной точкой пандемии, угрожающей здоровью не только реципиентов, но и всего общества в целом. Многие клетки человека экспрессируют рецепторы, специфичные для PERV типов A и B, в то время как рецепторы, специфичные для PERV C, на них в боль-

Copy numbers of PERV A per cell in the genomic DNA of domestic pigs and wild boars

<i>Sus scrofa</i> varieties	Number of animals	Median	25 th –75 th percentiles	min–max
Siberian miniature pig	10	4.5	2.2–9.2	0.6–58.9
Large White breed	6	1.0	0.2–12.0	0.2–28.1
Landrace breed	3	1.3	1.2–1.6	1.2–1.6
Wild boar	4	0.8	0.4–1.1	0.1–1.2*

* $p < 0.05$ as compared to miniature pigs.

шинстве случаев не обнаружены (Marcucci et al., 2009; Nakaya et al., 2011). Соответственно, политропные вирусы PERV A и PERV B способны инфицировать клетки человека, а экотропный вирус PERV C инфицирует только клетки свиньи. Показано, что рекомбинанты PERV A/C также способны инфицировать клетки человека, причем их провирусы могут интегрироваться *de novo* в геномы соматических (но не герминативных) клеток (Denner, 2008; Karlas et al., 2010).

Эксперименты доказали, что PERV могут инфицировать некоторые линии клеток человека *in vitro* (Specke et al., 2001; Lee et al., 2008; Yu et al., 2009; Kimsa et al., 2014). Перенос эндогенных ретровирусов свиньи и, в некоторых случаях, даже их репликация были продемонстрированы на культурах первичных эндотелиальных клеток, гепатоцитов и мононуклеарных клеток периферической крови человека (Specke et al., 2001; Li et al., 2006; Frühauf et al., 2009). Эндотелиальные клетки представляют особый интерес, поскольку этот тип клеток является основным тканевым барьером для PERV.

В настоящее время риск распространения вызванной PERV инфекции среди населения носит гипотетический характер, поскольку все больные находятся под пристальным эпидемиологическим контролем. Но если ксенотрансплантация станет более рутинной процедурой, то ситуация может быстро измениться. Поэтому для того, чтобы исключить какой-либо риск инфекции PERV, желательно получить свиней, утративших все тропные к человеку PERV. Экспрессия РНК PERV найдена практически у всех изученных пород домашних свиней (Yudin et al., 2011). Однако полногеномное секвенирование ДНК может помочь характеризовать и отбирать животных-доноров с пониженной инфекционной способностью PERV. Поскольку большинство копий PERV, вероятно, являются дефектными и вклад в репликацию компетентных вирусов вносят только несколько полноразмерных копий, экспрессия РНК в целом не должна коррелировать с выработкой инфекционных вирионов. Как только эти критические локусы будут идентифицированы, может быть достигнут (путем селекции или нокаутных технологий) контроль за способными к репликации PERV. Так, недавно появилось сообщение об успешной инактивации всех 62 копий PERV в геноме линии эпителиальных клеток почки свиньи PK15 с помощью технологии CRISPR/Cas9 (Yang et al., 2015). Определение числа копий PERV в геноме – первый шаг на пути к созданию безвирусных свиней с помощью биотехнологических манипуляций.

Существует два подхода для количественного определения уровня ДНК- или РНК-мишеней в образце ткани – абсолютный и относительный (Ребриков и др., 2009). Абсолютный количественный анализ определяет точное число копий мишени, как правило, путем сравнения значения порогового цикла ПЦР-реакции *Ct* с калибровочной кривой. При относительном количественном анализе изменение в экспрессии мишени нормализуется относительно экспрессии контрольного референсного гена. В настоящее время для абсолютного количественного анализа ДНК или РНК используются два метода ПЦР в реальном времени: с интеркалирующим красителем SYBR Green или с TaqMan-зондами. Использование красителя SYBR Green по сравнению с зондами TaqMan позволяет более эффективно и с меньшими затратами проанализировать большое число генов-мишеней. Поэтому на основе именно этого подхода был разработан метод количественного определения числа копий PERV (Yu et al., 2007; Zhang et al., 2010), который использован в нашей работе.

Предварительно, в соответствии с рекомендациями авторов метода, нами были подготовлены три калибровочных образца ДНК, в которых методом конечных разведений определялось число копий PERV A (Sedlak, Jerome, 2013). Исходя из предположения, что клетка свиньи содержит такое же количество ДНК, что и клетка человека (Yu et al., 2007), для образцов получены значения 2.4, 3.6 и 4.9 копии PERV A на клетку. Эти значения хорошо согласуются с числом копий гена *envA* в мононуклеарных клетках периферической крови (Yu et al., 2007) и других тканях (Zhang et al., 2010) у миниатюрных свиней китайских пород.

Селекционная группа сибирских миниатюрных свиней была выведена специально для медико-биологических целей (Тихонов, 2010). С использованием пластинок роста тел позвонков новорожденных мини-свиней в Новосибирском НИИ травматологии и ортопедии им. Я.Л. Цивьяна Минздрава России разработан метод получения хондротрансплантатов для коррекции дистрофических и травматических изменений хрящевой ткани, а также процессов роста при идиопатическом сколиозе (Тихонов, 2010). Проводятся исследования по апробации и производству биопротезов клапанов сердца, кровеносных сосудов и перикардального лоскута для интракардиальной хирургии и ангиопластики на основе материалов от мини-свиней (Тихонов, 2010).

С помощью метода количественного определения с использованием красителя SYBR Green нами определено

число копий PERV A в клетках крови у сибирских миниатюрных свиней, двух пород-основателей и диких кабанов. Оказалось, что у сибирских миниатюрных свиней число копий PERV A существенно выше, чем у дикого кабана. Аналогичная закономерность была найдена при анализе числа копий генов *pol* для других пород свиней (Mang et al., 2001). Авторы статьи выдвинули предположение, что инбридинг приводит к повышению числа копий PERV в геноме. Однако последующее изучение параметров генетического разнообразия по микросателлитным маркерам не подтвердило эту гипотезу (Quereda et al., 2012). Ранее нами было показано, что, наряду с «вертикальной» передачей вирусов PERV в ряду поколений, у свиней также возможен «горизонтальный» перенос вирусных частиц при тесном контакте между животными (Айтназаров и др., 2014а). Поскольку частота контактов между миниатюрными свиньями существенно выше, чем между дикими кабанами, обнаруженные нами различия могут быть связаны с «горизонтальным» путем передачи вирусной инфекции.

Медианы числа копий PERV A у свиней пород ландрас и крупная белая в нашем исследовании оказались существенно ниже, чем показатели, опубликованные другими авторами (Mang et al., 2001; Mazurek et al., 2013). Однако это несовпадение не вызывает удивления. Исследования свиней крупной белой породы показали, что хромосомное распределение и число копий PERV существенно различаются у разных особей (Bosch et al., 2000; Herring et al., 2001). Значительные внутривидовые различия были получены и для свиней других пород (Ma et al., 2010; Lee et al., 2011; Liu et al., 2011).

Полученный нами большой разброс числа копий вируса в пересчете на одну клетку может иметь два объяснения. С одной стороны, он может отражать низкую чувствительность использованного нами метода определения числа копий PERV A в области низких концентраций ДНК-мишени. В таком случае селекция свиней на снижение этого признака в популяции будет эффективной. С другой стороны, полученные значения могут быть связаны с индивидуальным разбросом среди животных по количеству копий PERV A, возникшим вследствие вариации процессов встройки/вырезания копий провируса в ходе онтогенеза. В этом случае селекция будет неэффективной и отбор целесообразно проводить исключительно на индивидуальном уровне, непосредственно перед трансплантацией, путем исследования различных органов на степень их пораженности ретровирусом. Поскольку другие авторы также наблюдали значительный разброс между животными по этому показателю (Mazurek et al., 2013), этот вопрос требует дальнейшего изучения.

Таким образом, геном сибирских миниатюрных свиней содержит существенное число копий тропных к клеткам человека ретровирусов PERV типа A. Этот факт должен учитываться при использовании органов животных для ксенотрансплантации. Метод количественного определения числа копий PERV A с применением красителя SYBR Green позволяет выбрать для ксенотрансплантации животных с наименьшим количеством ретровирусов в геноме и может быть использован для селекции мини-свиней на снижение этого показателя.

Acknowledgments

This work was supported by the Siberian Branch of the Russian Academy of Sciences, Integrated Program II.2, project 0324-2015-0019.

Conflict of interest

The authors declare no conflict of interest.

References

- Aitnazarov R.B., Nikitin S.V., Knyazev S.P., Yudin N.S. The saturation of genome of the pig by porcine endogenous retroviruses: the influence of heredity and environment. *Innovatsii i prodovolstvennaya bezopasnost* = Innovations and food security. 2014;2:41-49. (in Russian)
- Aitnazarov R.B., Yudin N.S., Nikitin S.V., Ermolayev V.I., Voevoda M.I. Identification of whole genomes of endogenous retroviruses in Siberian miniature pigs. *Rus. J. Genetics: Applied Research*. 2014;4(6):523-525. DOI 10.1134/S2079059714060021.
- Akiyoshi D.E., Denaro M., Zhu H., Greenstein J.L., Banerjee P., Fishman J.A. Identification of a full-length cDNA for an endogenous retrovirus of miniature swine. *J. Virology*. 1998;72(5):4503-4507.
- Bosch S., Arnauld C., Jestin A. Study of full-length porcine endogenous retrovirus genomes with envelope gene polymorphism in a specific-pathogen-free Large White swine herd. *J. Virology*. 2000;74(18): 8575-8581.
- Denner J. Recombinant porcine endogenous retroviruses (PERV-A/C): a new risk for xenotransplantation? *Archives Virology*. 2008;153(8): 1421-1426. DOI 10.1007/s00705-008-0141-7.
- Denner J. How active are porcine endogenous retroviruses (PERVs)? *Viruses*. 2016;8(8):E215. DOI 10.3390/v8080215.
- Ekser B., Cooper D.K., Tector A.J. The need for xenotransplantation as a source of organs and cells for clinical transplantation. *Intern. J. Surgery*. 2015;23:199-204. DOI 10.1016/j.ijssu.2015.06.066.
- Frühauf J.H., Mertsching H., Giri S., Frühauf N.R., Bader A. Porcine endogenous retrovirus released by a bioartificial liver infects primary human cells. *Liver International*. 2009;29(10):1553-1561. DOI 10.1111/j.1478-3231.2009.02087.x.
- Godehardt A.W., Rodrigues Costa M., Tönjes R.R. Review on porcine endogenous retrovirus detection assays—impact on quality and safety of xenotransplants. *Xenotransplantation*. 2015;22(2):95-101. DOI 10.1111/xen.12154.
- Herring C., Quinn G., Bower R., Parsons N., Logan N.A., Brawley A., Elsome K., Whittam A., Fernandez-Suarez X.M., Cunningham D., Onions D., Langford G., Scobie L. Mapping full-length porcine endogenous retroviruses in a large white pig. *J. Virology*. 2001;75(24): 12252-12265.
- Karlas A., Irgang M., Votteler J., Specke V., Ozel M., Kurth R., Denner J. Characterisation of a human cell-adapted porcine endogenous retrovirus PERV-A/C. *Ann. Transplantation*. 2010;15(2):45-54.
- Kimsa M.C., Strzalka-Mrozik B., Kimsa M.W., Gola J., Nicholson P., Lopata K., Mazurek U. Porcine endogenous retroviruses in xenotransplantation – molecular aspects. *Viruses*. 2014;6(5):2062-2083. DOI 10.3390/v6052062.
- Le Tissier P., Stoye J.P., Takeuchi Y., Patience C., Weiss R.A. Two sets of human-tropic pig retrovirus. *Nature*. 1997;389(6652):681-682.
- Lee D., Kim N.Y., Bae G.E., Lee H.J., Kwon M., Kim S.S., Lee H.T., Yang J.M., Kim Y.B. Transmissible infection of human 293T cells with porcine endogenous retroviruses subgroup A from NIH-miniature pig. *Transplantation Proceed.* 2008;40(10):3742-3745. DOI 10.1016/j.transproceed.2008.09.035.
- Lee D., Lee J., Yoon J.K., Kim N.Y., Kim G.W., Park C., Oh Y.K., Kim Y.B. Rapid determination of PERV copy number from porcine genomic DNA by real-time polymerase chain reaction. *Animal Biotechnol.* 2011;22(4):175-180. DOI 10.1080/10495398.2011.595294.
- Li Z., Ping Y., Shengfu L., Yangzhi Z., Jingqiu C., Youping L., Hong B. Variation of host cell tropism of porcine endogenous retroviruses expressed in chinese Banna minipig inbred. *Intervirology*. 2006;49(4): 185-191.

- Liu G., Li Z., Pan M., Ge M., Wang Y., Gao Y. Genetic prevalence of porcine endogenous retrovirus in chinese experimental miniature pigs. *Transplantation Proceed.* 2011;43(7):2762-2769. DOI 10.1016/j.transproceed.2011.06.061.
- Ma Y., Yang Y., Lv M., Yan Q., Zheng L., Ding F., Wu J., Tian K., Zhang J. Real-time quantitative polymerase chain reaction with SYBR green I detection for estimating copy numbers of porcine endogenous retrovirus from Chinese miniature pigs. *Transplantation Proceed.* 2010;42(5):1949-1952. DOI 10.1016/j.transproceed.2010.01.054.
- Mang R., Maas J., Chen X., Goudsmit J., van der Kuyl A.C. Identification of a novel type C porcine endogenous retrovirus: evidence that copy number of endogenous retroviruses increases during host inbreeding. *J. General Virology.* 2001;82(Pt. 8):1829-1834.
- Marcucci K.T., Argaw T., Wilson C.A., Salomon D.R. Identification of two distinct structural regions in a human porcine endogenous retrovirus receptor, HuPAR2, contributing to function for viral entry. *Retrovirology.* 2009;6(3):1-15. DOI 10.1186/1742-4690-6-3.
- Mazurek U., Kimsa M.C., Strzalka-Mrozik B., Kimsa M.W., Adam-ska J., Lipinski D., Zeyland J., Szalata M., Slomski R., Jura J., Smo-rag Z., Nowak R., Gola J. Quantitative analysis of porcine endogenous retroviruses in different organs of transgenic pigs generated for xenotransplantation. *Current Microbiology.* 2013;67(4):505-514. DOI 10.1007/s00284-013-0397-3.
- Morozov V.A., Morozov A.V., Rotem A., Barkai U., Bornstein S., Denner J. Extended microbiological characterization of Göttingen minipigs in the context of xenotransplantation: detection and vertical transmission of hepatitis E virus. *PLoS One.* 2015;10(10):e0139893. DOI 10.1371/journal.pone.0139893.
- Nakaya Y., Shojima T., Yasuda J., Imakawa K., Miyazawa T. Epigenetic regulation on the 5'-proximal CpG island of human porcine endogenous retrovirus subgroup A receptor 2/GPR172B. *Microbes Infect.* 2011;13(1):49-57. DOI 10.1016/j.micinf.2010.09.014.
- Niebert M., Tonjes R.R. Evolutionary spread and recombination of porcine endogenous retroviruses in suiformes. *J. Virology.* 2005;79(1):649-654.
- Nikitin S.V., Yudin N.S., Knyazev S.P., Aitnazarov R.B., Kobzev V.F., Bekenev V.A., Savvina M.A., Ermolaev V.I. Frequency of chromosomes carrying endogenous retroviruses in the populations of domestic pig and wild boar. *Genetika = Genetics (Moscow).* 2008;44(6):789-797. (in Russian)
- Quereda J.J., Herrero-Medrano J.M., Abellana J.M., García-Nicolás O., Martínez-Alarcón L., Pallarés F.J., Ramírez P., Muñoz A., Ramis G. Porcine endogenous retrovirus copy number in different pig breeds is not related to genetic diversity. *Zoonoses Public Health.* 2012; 59(6):401-407. DOI 10.1111/j.1863-2378.2012.01467.x.
- Rebrikov D.V., Samatov G.A., Trofimov D.Y., Semenov P.A., Savilova A.M., Kofiadi I.A., Abramov D.D. PTsR «v real'nom vremeni» [Real-time PCR]. Moscow, BINOM. *Laboratoriya znaniy Publ.*, 2009. (in Russian)
- Sedlak R.H., Jerome K.R. Viral diagnostics in the era of digital polymerase chain reaction. *Diagnostic Microbiology Infectious Disease.* 2013;75(1):1-4. DOI 10.1016/j.diagmicrobio.2012.10.009.
- Shimatsu Y., Yamada K., Horii W., Hirakata A., Sakamoto Y., Waki S., Sano J., Saitoh T., Sahara H., Shimizu A., Yazawa H., Sachs D.H., Nunoya T. Production of cloned NIBS (Nippon Institute for Biological Science) and α -1, 3-galactosyltransferase knockout MGH miniature pigs by somatic cell nuclear transfer using the NIBS breed as surrogates. *Xenotransplantation.* 2013;20(3):157-164. DOI 10.1111/xen.12031.
- Specke V., Rubant S., Denner J. Productive infection of human primary cells and cell lines with porcine endogenous retroviruses. *Virology.* 2001;285(2):177-180.
- Tikhonov V.N. *Laboratornye mini-svin'i, genetika i mediko-biologicheskoe ispolzovanie* [Laboratory Mini-pigs: Genetics and Biomedical Use]. Novosibirsk, SB RAS Publ., 2010. (in Russian)
- Yang L., Güell M., Niu D., George H., Lesha E., Grishin D., Aach J., Shrock E., Xu W., Poci J., Cortazio R., Wilkinson R.A., Fishman J.A., Church G. Genome-wide inactivation of porcine endogenous retroviruses (PERVs). *Science.* 2015;350(6264):1101-1104. DOI 10.1126/science.aad1191.
- Yu P., Zhang L., Li S.F., Cheng J.Q., Lu Y.R., Zeng Y.Z., Li Y.P., Bu H. A rapid method for detection of the copy number of porcine endogenous in swine. *J. Rapid Methods Automation Microbiology.* 2007; 15:199-205.
- Yu P., Zhang P., Zhang L., Li S.F., Cheng J.Q., Lu Y.R., Li Y.P., Bu H. Studies on long-term infection of human cells with porcine endogenous retrovirus. *Acta Virologica.* 2009;53(3):169-174.
- Yudin N.S., Aitnazarov R.B., Ermolaev V.I. Porcine endogenous retroviruses: what are the risks of infection transmission in xenotransplantation? *Rus. J. Genet. Appl. Res.* 2011;1(6):532-539. DOI 10.1134/S207905971106013X.
- Zhang P., Yu P., Wang W., Zhang L., Li S., Bu H. An effective method for the quantitative detection of porcine endogenous retrovirus in pig tissues. *In Vitro Cellular & Develop. Biology – Animal.* 2010; 46(5):408-410. DOI 10.1007/s11626-009-9264-8.

Анализ геномного распределения сайтов связывания транскрипционных факторов GAGA и CNC в развитии *Drosophila melanogaster*

И.И. Брусенцов , Д.А. Карагодин, Э.М. Баричева, Т.И. Меркулова

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

На основании доступных данных ChIP-seq и ChIP-chip экспериментов, выполненных с использованием антител к транскрипционным факторам GAGA и CNC, проведен анализ распределения по геному участков связывания этих факторов в эмбриогенезе (возраст эмбрионов 0–12 и 16–24 ч), а также на стадии белой предкуколки дрозофилы. Показано, что основная часть мест связывания GAGA и CNC попадает в промоторные районы и интроны генов, при этом максимальная плотность пиков связывания обоих факторов приходится на область старта транскрипции. Показано также, что у эмбрионов в возрасте как 0–12, так и 16–24 ч развития наблюдается неслучайная колокализация GAGA и CNC, в то время как на стадии белой предкуколки колокализации этих транскрипционных факторов в масштабе генома выявить не удается. Для того чтобы очертить круг генов, для которых возможна совместная регуляция GAGA и CNC, было осуществлено исследование их совместного распределения в аннотированных регуляторных районах (промоторная область и участки, соответствующие 5'-UTR и 3'-UTR мРНК). Оказалось, что совокупности генов, в регуляторных районах которых обнаружено связывание обоих факторов, сильно различаются на разных стадиях. Если у эмбрионов в возрасте 0–12 ч 353 гена характеризуются пересечением пиков ChIP-seq GAGA и CNC, а в возрасте 16–24 ч число таких генов составляет 611, то всего лишь 61 ген является «общим» для обеих стадий. Предполагается, что различные подгруппы генов-мишеней этих факторов регулируются разными сочетаниями изоформ GAGA и CNC, паттерны экспрессии которых изменяются в ходе эмбриогенеза дрозофилы. Функциональный анализ генов, в регуляторных районах которых найдена колокализация GAGA и CNC на всех исследованных стадиях развития, показывает обогащение генами, контролирующими эмбриогенез, развитие нервной системы и крыла дрозофилы.

Ключевые слова: транскрипционные факторы; GAGA; CNC; развитие; *D. melanogaster*.

Analysis of whole-genome binding patterns of GAGA and CNC transcription factors during *Drosophila melanogaster* development

I.I. Brusentsov , D.A. Karagodin, E.M. Baricheva, T.I. Merkulova

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

On the basis of available data of ChIP-seq and ChIP-chip experiments performed using antibodies against GAGA and CNC transcription factors, genome-wide binding mapping of these factors at hours 0–12 and 16–24 of *Drosophila* embryogenesis, as well as on white pre-pupae stage, was conducted. It was shown that the bulk of GAGA and CNC binding falls into promoter regions and introns, with the maximal density of peaks in the vicinity of the transcription start site. Moreover, in both 0–12 and 16–24 hour old embryos GAGA and CNC are frequently co-localized, while on white pre-pupae stage there is no co-localization of these factors on a genome-wide scale. In order to select a set of genes potentially co-regulated by GAGA and CNC, the study of their co-binding in annotated regulatory regions (promoter areas and segments corresponding to the 5'-UTR and 3'-UTR of mRNA) was performed. The results obtained clearly demonstrated that the sets of genes characterized by co-binding of both factors vary greatly at different stages. Thus from 353 genes with overlapped GAGA and CNC binding loci on the 0–12 hour old embryos and 611 genes on the 16–24 hour old embryos only 61 genes “belong” to both stages. For an explanation it is proposed that different sets of target genes are regulated by combinations of various GAGA and CNC isoforms, which are characterized by distinct expression patterns during *drosophila* embryogenesis. Functional annotation analysis of genes, in whose regulatory regions both GAGA and CNC were found at all investigated stages, demonstrates enrichment by genes controlling embryogenesis, neurogenesis and wing development. The data obtained suggest the interaction of GAGA and CNC during *D. melanogaster* embryogenesis.

Key words: transcription factors; GAGA; CNC; development; *D. melanogaster*.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Брусенцов И.И., Карагодин Д.А., Баричева Э.М., Меркулова Т.И. Анализ геномного распределения сайтов связывания транскрипционных факторов GAGA и CNC в развитии *Drosophila melanogaster*. Вавиловский журнал генетики и селекции. 2016;20(6):762-769. DOI 10.18699/VJ16.193

HOW TO CITE THIS ARTICLE:

Brusentsov I.I., Karagodin D.A., Baricheva E.M., Merkulova T.I. Analysis of whole-genome binding patterns of GAGA and CNC transcription factors during *Drosophila melanogaster* development. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):762-769. DOI 10.18699/VJ16.193

ORIGINAL ARTICLE

Received 11.03.2016

Accepted for publication 26.04.2016

© AUTHORS, 2016

Развитие и жизнедеятельность многоклеточного организма осуществляются за счет дифференциальной экспрессии многих тысяч генов, определенные наборы которых транскрибируются в различных типах клеток на разных этапах развития и отвечают на разнообразные внешние сигналы (Tomancak, Ohler, 2010; Меркулова и др., 2013). Известно, что регуляция экспрессии большинства генов происходит на транскрипционном уровне, на котором центральную роль играют транскрипционные факторы (ТФ) – регуляторные белки, опознающие определенные последовательности ДНК в регуляторных районах (Frith et al., 2008; Charoensawan et al., 2010). В зависимости от состава используемых сайтов связывания ТФ в регуляторных районах генов формируются различные ансамбли этих белков, и такая комбинаторика сайтов и взаимодействующих с ними ТФ имеет ключевое значение для определения уровня экспрессии различных генов, обусловленного стадией развития организма, типом клетки и внешними воздействиями (Harbison et al., 2004; Ochoa-Espinosa, Small, 2006; Bulger, Groudine, 2011; Меркулова и др., 2013).

D. melanogaster на протяжении многих лет является одной из наиболее популярных моделей для изучения сети генных взаимодействий и роли различных ТФ на разных стадиях онтогенеза, включая эмбриональное развитие (Adryan, Teichmann, 2010; Tomancak, Ohler, 2010). В ходе изучения эмбриогенеза *D. melanogaster* было выявлено более 10 ТФ (Bicoid, Nanos, Kruppel, Knirps, Giant, Hunchback, Caudal и др.) координированное действие которых определяет ранние стадии формирования эмбриона (Ingham, 1988; Rivera-Pomar, Jäckle, 1996; Segal et al., 2008). Проведенные позднее полногеномные исследования показали, что в контроле онтогенеза дрозофилы участвуют сотни ТФ, хотя конкретная роль большинства из них еще не определена (Adryan, Teichmann, 2010; Hammonds et al., 2013). До сих пор очень мало известно также о взаимодействии различных ТФ друг с другом в ходе развития дрозофилы. Поэтому представляется очевидным, что наряду с общим анализом полногеномных данных необходимы более конкретизированные исследования накопленного экспериментального материала, в частности изучение колокализации отдельных ТФ в масштабе генома и выявление круга генов, совместно регулируемых этими факторами.

Транскрипционный фактор GAGA, кодируемый геном *Trithorax-like* (*Trl*), необходим для нормального развития *D. melanogaster* (Farkas et al., 1994). Он экспрессируется на всех стадиях развития, но особенно активно во время эмбриогенеза (<http://flybase.org/reports/FBgn0013263.html>). В ходе анализа полногеномных профилей связывания ТФ было установлено, что GAGA специфично связывается с так называемыми HOT-регионами (или Hotspot), местами с повышенной плотностью сайтов связывания разных ТФ по сравнению с остальным геномом (Kvon et al., 2012; Slattery et al., 2014). Поскольку GAGA присутствует в большинстве тканей и на всех стадиях онтогенеза, должен существовать механизм, обеспечивающий изменение наборов генов, регулируемых им в разных условиях. ТФ GAGA представлен несколькими изоформами, и одним из возможных механизмов специфичности его действия

является связывание с разными сайтами различных его гомо- и гетеродимеров (Espinosa et al., 1999). Другим возможным механизмом может быть взаимодействие GAGA с другими ТФ и белками хроматина для обеспечения изменения специфичности связывания с разными сайтами (Pagans et al., 2002). В этом случае изменение перекрывания профилей связывания ТФ на разных стадиях развития будут обеспечивать изменение набора регулируемых генов. В рамках проекта modENCODE (www.modencode.org; Roy et al., 2010) были получены профили связывания для более чем 80 ТФ на разных стадиях развития дрозофилы. Однако в настоящее время только для очень небольшого количества этих факторов получены профили связывания для нескольких следующих друг за другом стадий развития, что дает возможность проследить динамику колокализации районов связывания разных ТФ и сделать предположения об их возможном взаимодействии в течение развития. Для GAGA профили были получены у эмбрионов нескольких возрастов, а также на стадии белой предкуколки (БП). Только для трех ТФ (SuHW, STAT92E, CNC) также получены профили для этих же периодов развития. SuHW является важным компонентом инсуляторных комплексов, и взаимное расположение его сайтов и сайтов связывания GAGA исследовалось в том числе и на геномном уровне (Nègre et al., 2010). Для STAT92E и CNC имеются литературные данные о перекрывании их сайтов связывания с сайтами GAGA, полученные при анализе профилей только для 0–12-часовых эмбрионов (Nègre et al., 2011). На наш взгляд, наибольший интерес для изучения возможного взаимодействия с GAGA представляет CNC, кодируемый геном *cap-n-collar* (*cnc*). CNC является важным регулятором эмбриогенеза дрозофилы, в частности он необходим для формирования головного сегмента эмбриона (Mohler et al., 1995).

Целью настоящей работы было изучение распределения сайтов связывания для двух важных регуляторов эмбрионального развития *D. melanogaster* – ТФ GAGA и CNC – в течение эмбрионального развития и на стадии БП. Мы определили также круг генов, которые, возможно, совместно регулируются этими факторами.

Материалы и методы

В анализ были взяты данные полногеномных экспериментов по иммунопреципитации хроматина из проекта modENCODE (www.modencode.org; Roy et al., 2010). Эксперименты были выполнены с использованием антител к константным областям транскрипционных факторов GAGA и CNC на линии *D. melanogaster* iso1 (*y, bw cn sp*) для эмбрионов в возрасте 0–12 и 16–24 ч и на стадии БП. Экспериментальные данные для эмбрионов 0–12 ч были получены с применением технологии ChIP-chip, остальные – с помощью ChIP-seq. Исходные данные были извлечены из архивов NCBI (<http://www.ncbi.nlm.nih.gov/geo/> для ChIP-chip и <http://www.ncbi.nlm.nih.gov/sra/> для ChIP-seq) (таблица). В работе использовали версию 6 аннотации генома дрозофилы (Hoskins et al., 2015; ftp://ftp.ensembl.org/pub/release-83/gtf/drosophila_melanogaster/Drosophila_melanogaster.BDGP6.83.gtf.gz).

Поскольку ChIP-chip для 0–12-часовых эмбрионов проводили в рамках одного эксперимента, данные о про-

Raw ChIP-seq and ChIP-on-chip data and peak detection results

Transcription factor	Stage	ID GEO	Number of peaks	Mean peak size (bp)
GAGA	Embryo 0–12 h	GSM409071	3 902	1 826
	Embryo 16–24 h	SRR063878, SRR063879, SRR063880*	6 682	658
	White prepupa	SRR099124, SRR099125, SRR099126*	18 668	346
CNC	Embryo 0–12 h	GSM399627	697	1 037
	Embryo 16–24 h	SRR390242, SRR390243, SRR390244*	823	618
	White prepupa	SRR350870, SRR350871, SRR350872*	415	568

* input control.

филе распределения пиков взяты без предварительной обработки.

Предпроцессинг данных ChIP-seq проводили с использованием программы Trimmomatic-3.2.2 (Bolger et al., 2014). Средняя длина прочтения составила 36 п. н. Выравнивание проводили на последовательность генома дрозофилы версии 6 с помощью пакета программ Bowtie2 (Langmead, Salzberg, 2012). Для удаления эффекта дубликации прочтений использовали пакет программ SAMtools 1.2 (Li et al., 2009). Поиск пиков связывания проводили с помощью программы MACs2 2.0.9 (qvalue = 0.01) (Zhang et al., 2008). В качестве контроля для бутстреп-анализа были сформированы искусственные данные по каждому транскрипционному фактору, содержащие то же количество пиков для каждой хромосомы и размеры пиков, соответствующие средней протяженности пиков в экспериментальных данных, но со случайным распределением по геному. Относительно этой же выборки оценивали обогащение различных районов генома пиками связывания CNC и GAGA, полученными в эксперименте. Плотность распределения пиков (ρ) в различных районах генома (промоторы, 5'- и 3'-UTRs, интроны, кодирующие районы и межгенные промежутки) рассчитывали по формуле:

$$\rho = \frac{Nr}{N+Sr},$$

где Nr – количество пиков данного ТФ, попадающих в определенный район генома, N – общее количество пиков данного ТФ, Sr – суммарная площадь данного района генома в мегабазах.

Координаты HOT-районов и STARR-Seq энхансеров взяты из ранее опубликованных работ (Roy et al., 2010; Arnold et al., 2013). Преобразование ординат к версии 6 аннотации генома дрозофилы производили с использованием Drosophila Sequence Coordinates Converter (http://flybase.org/static_pages/downloads/COORD.html).

Анализ обогащения генов-мишеней выполняли с использованием пакета программ WebGestalt (<http://www.webgestalt.org/>).

Результаты

При сравнении результатов экспериментов по иммунопреципитации хроматина было выявлено значительно большее количество мест связывания GAGA фактора с ДНК по сравнению с CNC (см. таблицу). Причем если в случае с CNC это количество остается относительно ста-

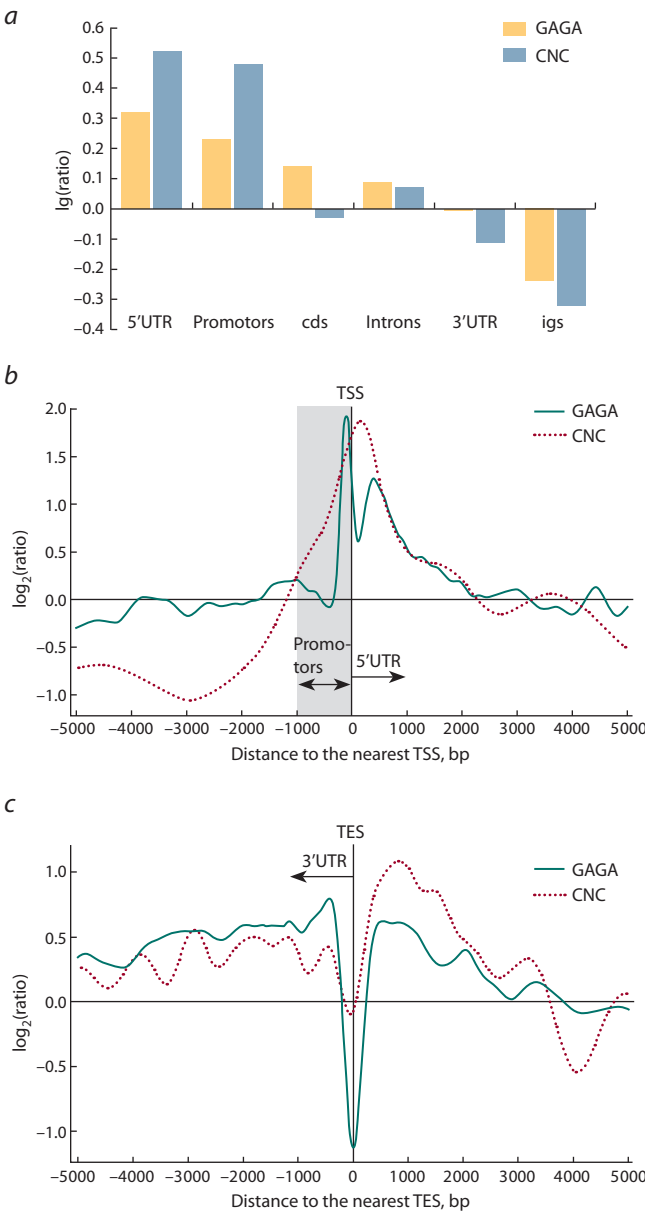


Fig. 1. (a) Distribution of CNC and GAGA-binding peaks in the *Drosophila* genome. Designations: cds, coding sequences; igs, intergenic spacers. (b) Regions enriched and depleted in TFBSs around TSS. (c) Same around TES. Y-axis: logarithmic ratios between the densities of GAGA and CNC peaks around (b) TSS and (c) TES and the densities of the peaks in an artificial sample in the same region (see Materials and methods).

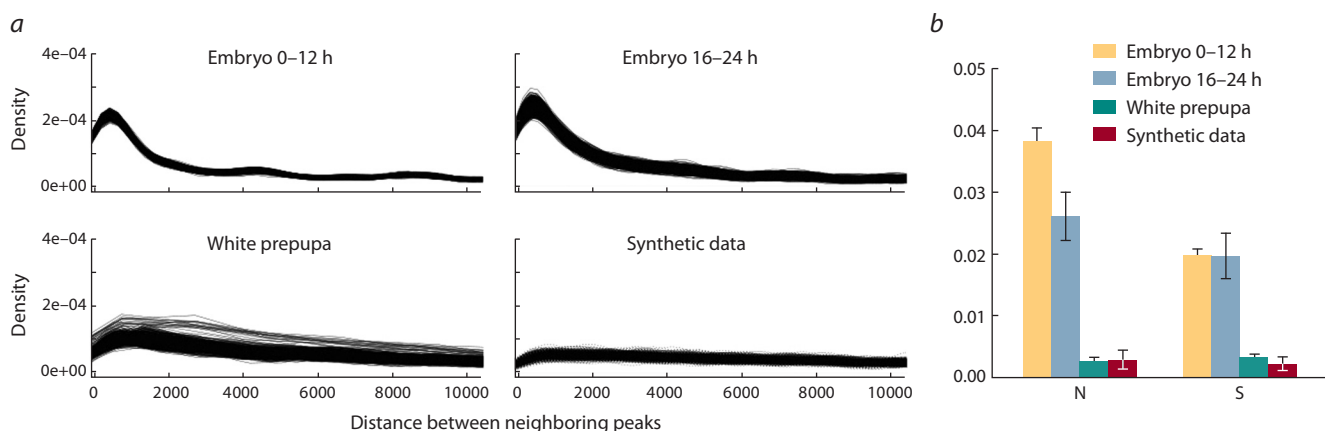


Fig. 2. Arrangement of GAGA- and CNC-binding peaks in the *D. melanogaster* genome.

(a) Distance between the tops of neighboring CNC and GAGA peaks, bp; (b) shares of the numbers (N) and coverage areas (S) of overlapping peaks in the total number of peaks $N_{\text{over}} / (N_{\text{CNC}} + N_{\text{GAGA}})$ and the total coverage area $S_{\text{over}} / (S_{\text{CNC}} + S_{\text{GAGA}} - S_{\text{over}})$. Error bars indicate 95% Gaussian confidence ranges ($\pm 1.96 \sigma$).

бильным во всех стадиях, то количество пиков связывания GAGA растет по мере развития мух.

При оценке распределения пиков связывания GAGA и CNC по геному для всех исследованных стадий было выявлено, что районы начала генов, включающие промоторную область (интервал от старта транскрипции до -1000 п. н.) и участок, соответствующий 5'-UTR мРНК, существенно обогащены пиками обоих факторов (рис. 1, а). Следует отметить достаточно высокую плотность пиков GAGA в кодирующих областях генов (см. рис. 1, а). Возможно, это связано с участием GAGA в процессе элонгации транскрипции (O'Brien et al., 1995). Максимальная плотность пиков и GAGA и CNC приходится на область старта транскрипции (TSS) (см. рис. 1, б). Такой характер распределения мест связывания в геноме зафиксирован и для других транскрипционных факторов (Farnham, 2009) и подтверждает принятую точку зрения о важной роли промоторных районов генов в регуляции транскрипции (Меркулова и др., 2013). Минимальная плотность пиков обоих факторов наблюдается в районе сайта терминации транскрипции (TES), что ранее было зафиксировано для GAGA (Nègre et al., 2010), а для CNC наблюдается впервые.

Для исследования возможного взаимодействия транскрипционных факторов была проведена оценка минимального расстояния между пиками GAGA и CNC на разных стадиях развития мух. Чтобы нивелировать влияние разницы в количестве пиков, полученных в разных экспериментах, на результат вычисления, для каждой стадии были случайным образом сформированы подвыборки, содержащие 3902 пиков GAGA и 415 CNC, и на этих подвыборках оценивали расстояние между вершинами соседних пиков ТФ. Достоверность определяли с использованием бутстреп-анализа (999 реплик). Результаты представлены на рис. 2. Видно, что на эмбриональной стадии развития пик исследуемых ТФ тяготеют друг к другу. На стадии БП колокализации сайтов GAGA и CNC на геномном уровне практически не наблюдается: из рис. 2, б видно, что

расстояние между пиками на этой стадии достоверно не отличается от расстояний между пиками в искусственно созданной нами выборке. При оценке количества и площади перекрытия пиков мы также видим колокализацию CNC и GAGA на эмбриональных стадиях и ее отсутствие на стадии БП в масштабе генома (см. рис. 2, б).

Для исследования функциональной роли участков возможного совместного связывания CNC и GAGA мы проанализировали перекрытие всех таких участков с HOT-регионами и STARR-seq энхансерами. HOT-регионы – это участки ДНК, в которых плотность сайтов связывания ТФ повышена (Moogman et al., 2006). Было продемонстрировано, что HOT-регионы часто обладают энхансерной активностью (Kvon et al., 2012). Также ранее был предложен экспериментальный метод для полногеномного выявления энхансеров (STARR-seq), с помощью которого авторы выявили более 10000 участков, обладающих энхансерной активностью в S2 и OSC клетках (Arnold et al., 2013). Было установлено, что районы колокации GAGA и CNC с высокой частотой перекрываются и с HOT-регионами, и с энхансерами (рис. 3). С HOT-регионами перекрывается 90 %, а с энхансерами – 42 % районов колокации CNC и GAGA, что достоверно больше, чем перекрытие каждого из ТФ по отдельности. Таким образом, значительная часть районов колокации может обладать энхансерной активностью и участвовать в активации экспрессии рядом расположенных генов.

Особый интерес представляет возможная колокализация изучаемых ТФ в регуляторных районах генов. В качестве таких районов были выбраны промоторная область и прилегающий к ней участок, соответствующий 5'-UTR мРНК, а также участок, соответствующий 3'-UTR мРНК. Оказалось, что у эмбрионов в возрасте 0–12 ч 353 гена характеризуются пересечением пиков связывания GAGA и CNC в этих районах, в 12–24 ч – число таких генов составляет 611, а на стадии БП – 180 (рис. 4). Из приведенных на рис. 4 данных видно также, что число ге-

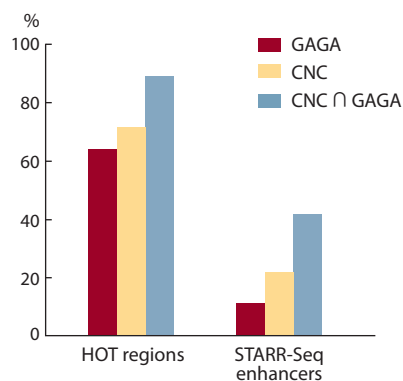


Fig. 3. Shares of GAGA- and CNC-binding regions and their putative linkage (CNC ∩ GAGA) that overlap HOT regions and S2 enhancers.

Data are integrated over all *Drosophila* developmental stages investigated. The share of regions with CNC and GAGA colocalization is significantly ($p < 0.01$ according to Fisher's exact test) greater than the share of each factor in both cases.

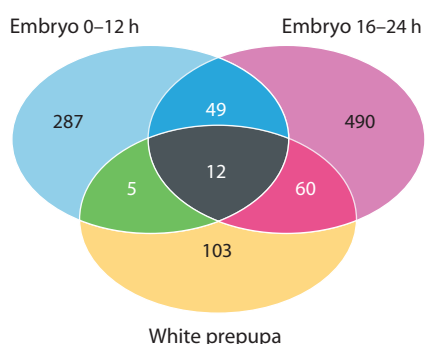


Fig. 4. Numbers of genes with peaks of GAGA and CNC-binding sites at the 0–12 h and 16–24 h embryonic stages and at the white prepupa stage.

нов, в регуляторных районах которых обнаружено совместное связывание GAGA и CNC на более одной стадии, весьма невелико. Так, только в 12 из них колокализация ТФ наблюдалась на всех стадиях. Для эмбрионов в возрасте 0–12 и 16–24 ч количество таких «общих» районов равно 61, в 0–12 ч и БП – 17, а в 16–24 ч и БП – 72. Наиболее вероятным объяснением этому явлению служит изменение паттернов экспрессии изоформ GAGA и CNC в ходе эмбриогенеза дрозофилы. Это позволяет предполагать, что разные подгруппы генов-мишеней GAGA и CNC регулируются различными сочетаниями изоформ GAGA и CNC.

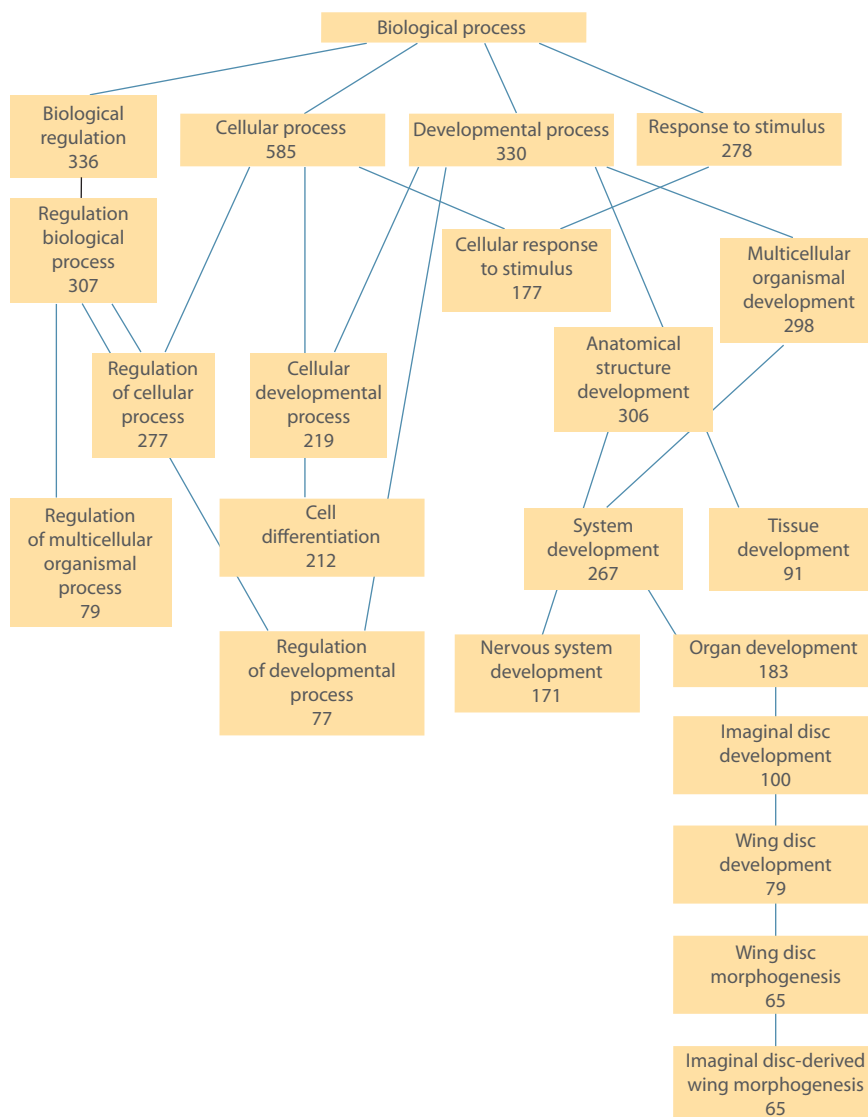


Fig. 5. Involvement of genes targeted by CNC and GAGA TFs in various biologic processes according to analysis with WebGestalt software ($p < 0.02$).

Мы проанализировали функции всех выявленных нами предполагаемых генов-мишеней обоих ТФ (1006 генов). В результате выяснилось, что совокупность этих генов обогащена определенными группами. Так, по термину «молекулярные функции» выявлено повышенное содержание генов, кодирующих транскрипционные факторы (60 генов). По термину «биологические функции» среди генов-мишеней обоих ТФ также наблюдалось существенное обогащение генами-регуляторами различных процессов (336 генов). Это, вероятно, связано с тем, что *cnc* и *Trl* находятся на вершине регуляторных каскадов, а их мишени-регуляторы осуществляют регуляцию второго порядка. Более детально процессы, в которые вовлечены предполагаемые гены-мишени, представлены на рис. 5. Среди них интересны процессы развития нервной системы и морфогенеза крыла. Представленность генов, связанных с этими процессами, не одинакова на разных стадиях развития. Так, если обогащение по генам развития нервной системы наблюдается на всех изучаемых стадиях (эмбрионы 0–12 ч, эмбрионы 16–24 ч, БП), то обогащение по генам, вовлеченным в морфогенез крыльев, – только на стадии эмбриогенеза 16–24 ч, что коррелирует со временем инвагинации имажинальных дисков (Campos-Ortega, Hartenstein, 1997).

Обсуждение

Разработка полногеномных технологий изучения экспрессии генов, функциональной организации хроматина и локализации различных ТФ в масштабе генома (RNA-seq, ChIP-chip, ChIP-seq, DamID) привела к существенному расширению знаний о характерных для разных типов клеток профилях транскрипции генов, модификаций хроматина и распределения ТФ, а также об изменениях этих профилей в ходе развития многоклеточного организма (Filion et al., 2010; Roy et al., 2010; Lelli et al., 2012; Diehl, Boyle, 2016). Однако в настоящее время темпы накопления экспериментальных данных, получаемых с помощью указанных технологий, значительно превышают темпы извлечения из них биологической информации. Так, в базах проекта modENCODE содержится уже довольно много данных ChIP-seq экспериментов по определению полногеномного распределения ряда ТФ на разных стадиях развития *D. melanogaster*, однако их анализ далек от завершения. В частности, для весьма немногих ТФ выполнены работы по выявлению совместной локализации этих факторов (Slattery et al., 2014), которая может служить весомым указанием на их взаимодействие в процессе регуляции тех или иных наборов генов. Среди факторов, представляющих интерес для подобного изучения, мы выделили два важных регулятора развития дрозофилы, ТФ GAGA и CNC, для которых в modENCODE содержатся данные ChIP-chip и ChIP-seq экспериментов, выполненных во время 0–12 ч и 16–24 ч развития эмбриона, а также на стадии белой предкуколки.

Известно, что гены *Trl* и *cnc* экспрессируются на всех стадиях онтогенеза дрозофилы (Перельгина и др., 1993; Soeller et al., 1993; <http://www.flyexpress.net/>). Экспрессия *Trl* достигает максимума у эмбрионов 4–12 ч развития, а затем постепенно снижается (Soeller et al., 1993; Bhat et al., 1996; Karagodin et al., 2013). В раннем эмбриогенезе транскрипты *Trl* равномерно распределены по всему эмбриону, а в дальнейшем транскрипция наблюдается преимущественно в формирующейся нервной системе (Перельгина и др., 1993). Основными транскриптами гена *Trl* являются два: длиной 2.5 и 3.0 т.н. На стадии 0–4 ч развития эмбрионов преобладает 2.5 т.н. транскрипт, а начиная с 4 ч увеличивается количество транскрипта длиной 3.0 т.н., и с 8–12 ч он становится преобладающим и остается таковым до конца эмбриогенеза (Soeller et al., 1993; Karagodin et al., 2013). Известно, что транскрипт длиной 2,5 т.н. соответствует изоформе белка GAGA-519, а транскрипт длиной 3.0 т.н. — изоформе GAGA-581 (Karagodin et al., 2013). Несмотря на то что соотношение основных транскриптов гена *Trl* меняется в ходе онтогенеза, установлено, что только в первые 6 ч развития эмбрионов в них доминирует белок GAGA-519, а в дальнейшем обе изоформы выявляются примерно в равных количествах (Benyajati et al., 1997).

Транскрипты гена *cnc* на первых стадиях эмбриогенеза так же, как и транскрипты гена *Trl*, равномерно распределены во всем эмбрионе, а в дальнейшем транскрипция *cnc* в основном ограничена областью формирующейся головы (<http://www.flyexpress.net/>). Выявлены три главных транскрипта гена *cnc* длиной 6.6; 5.4; 3.3 т.н. (McGinnis et al., 1998). У 2–8-часовых эмбрионов преобладающими

являются транскрипты 5.4 и 4.0 т.н., у 8–12-часовых эмбрионов основной транскрипт — 5.4 т.н., а у 12–24-часовых присутствуют все три транскрипта. Выявлены соответствующие этим транскриптам изоформы белка CNC длиной 533 а.о. (изоформа А, 3.3 т.н. транскрипт), 805 а.о. (изоформа В, 5.4 т.н. транскрипт) и 1383 а.о. (изоформа С, 6.6 т.н. транскрипт). Транскрипты, дающие начало изоформам А и С, и сами белки нарабатываются еще во время оогенеза, поэтому они выявляются в раннем эмбриогенезе еще до начала зиготической транскрипции. В дальнейшем эти изоформы появляются вновь уже после 11 ч эмбриогенеза и присутствуют до конца этого процесса. Значительное количество изоформы В выявляется в головном отделе эмбриона на стадиях 6–14 (3–12 ч) эмбриогенеза, тогда как в другое время этот белок обнаруживается в эмбрионе в очень малом количестве (McGinnis et al., 1998). Таким образом, если разные изоформы белка GAGA выявляются в течение большей части эмбриогенеза приблизительно в равных количествах, то разные изоформы белка CNC распределены по разным периодам эмбриогенеза достаточно специфично.

Анализ взаимного расположения пиков связывания CNC и GAGA позволил выявить районы, в которых пики перекрываются на определенной стадии развития. В этих районах сайты связывания CNC и GAGA располагаются достаточно близко друг к другу, для того чтобы было возможно влияние связывания одного из ТФ на связывание второго и даже физическое взаимодействие между ними для совместного связывания. Нужно отметить, что данные ChIP-seq и ChIP-chip относятся к целым эмбрионам и БП, поэтому нельзя исключить наложение пиков ТФ от разных типов клеток. Однако следует учитывать высокую степень перекрывания районов экспрессии генов *cnc* и *Trl*. Так, в раннем эмбриогенезе GAGA и CNC присутствуют во всех клетках эмбриона, а на более поздних стадиях их экспрессия перекрывается в головном сегменте эмбриона. Поэтому кажется маловероятным, что большая часть наблюдаемого перекрытия возникает за счет наложения пиков от разных типов клеток. Как следствие, поиск мест совместной локализации CNC и GAGA служит первым этапом для выявления возможной совместной регуляции генов этими ТФ в течение развития дрозофилы.

Проведенный нами анализ высветил новые аспекты возможного совместного действия этих факторов в ходе эмбрионального развития *D. melanogaster*. Мы показали, что в периоды 0–12 и 16–24 ч развития эмбриона наблюдается неслучайная колокализация GAGA и CNC, в то время как на стадии белой предкуколки колокализации этих факторов в масштабе генома выявить не удастся. Данные указывают на возможность совместной регуляции GAGA и CNC определенного круга генов во время эмбриогенеза. Для того чтобы выявить эти гены, было осуществлено исследование колокализации изучаемых ТФ в аннотированных регуляторных районах (промоторная область и участки, соответствующие 5'-UTR и 3'-UTR мРНК). Оказалось, что совокупности генов, в регуляторных районах которых обнаружено связывание обоих ТФ, сильно различаются на разных стадиях. Так, если в период 0–12 ч эмбриогенеза 353 гена характеризуются пересечением пиков ChIP-seq GAGA и CNC, а в

период 12–24 ч число таких генов составляет 611, то всего лишь 61 ген является «общим» для обоих периодов. Мы предполагаем, что такое варьирование числа генов, характеризующихся совместным связыванием обоих белков, определяется тем, что на разных стадиях развития преобладают различные изоформы CNC и GAGA, которые, по-видимому, обладают определенной избирательностью по отношению к сайтам связывания этих белков на ДНК и/или в составе хроматина.

Функциональный анализ генов, в регуляторных районах которых найдена колокализация ТФ GAGA и CNC, выделяет группы генов, контролирующих эмбриогенез, развитие нервной системы и развитие крыла дрозофилы. Следует отметить, что ТФ GAGA так же, как и ТФ CNC, влияют на ход этих процессов (Iyer et al., 2013; Schertel et al., 2015). Так, хорошо известно, что GAGA регулирует активность многих генов, контролирующих эмбриогенез дрозофилы (Granok et al., 1995). Ген *cnc* также принимает участие в этом процессе, влияя на работу генов развития головы (Veraksa et al., 2000). Однако до настоящего времени не было данных о совместном участии GAGA и CNC в регуляции транскрипции конкретных генов. Полученные нами результаты указывают на существование обширной и динамично меняющейся сети генов, совместно регулируемых этими ТФ в ходе эмбрионального развития дрозофилы, и открывают новые перспективы в изучении этого процесса.

Acknowledgments

This work was supported by State Budgeted Project 0324-2015-0003 and the Russian Foundation for Basic Research, project 16-04-01018-a.

Conflict of interest

The authors declare no conflict of interest.

References

- Adryan B., Teichmann S.A. The developmental expression dynamics of *Drosophila melanogaster* transcription factors. *Genome Biol.* 2010; 11(4):R40.
- Arnold C.D., Gerlach D., Stelzer C., Boryn L.M., Rath M., Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science.* 2013;339(6123):1074-1077.
- Benyajati C., Mueller L., Xu N., Pappano M., Gao J., Mosammaparast M., Conklin D., Granok H., Craig C., Elgin S. Multiple isoforms of GAGA factor, a critical component of chromatin structure. *Nucl. Acids Res.* 1997;25(16):3345-3353.
- Bhat K.M., Farkas G., Karch F., Gyurkovics H., Gausz J., Schedl P. The GAGA factor is required in the early *Drosophila* embryo not only for transcriptional regulation but also for nuclear division. *Development* (Cambridge, England). 1996;122(4):1113-1124.
- Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (Oxford, England). 2014; 30(15):2114-2120.
- Bulger M., Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell.* 2011;144(3):327-339.
- Campos-Ortega J.A., Hartenstein V. *The Embryonic Development of Drosophila Melanogaster*. Berlin; Heidelberg: Springer, 1997.
- Charoensawan V., Wilson D., Teichmann S.A. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucl. Acids Res.* 2010;38(21):7364-7377.
- Diehl A.G., Boyle A.P. Deciphering ENCODE. *Trends Genet.* 2016; 32(4):238-249.
- Espinas M.L., Jimenez-Garcia E., Vaquero A., Canudas S., Bernues J., Azorin F. The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. *J. Biol. Chemistry.* 1999;274:16461-16469.
- Farkas G., Gausz J., Galloni M., Reuter G., Gyurkovics H., Karch F. The Trithorax-like gene encodes the *Drosophila* GAGA factor. *Nature.* 1994;371(6500):806-808.
- Farnham P.J. Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* 2009;10(9):605-616.
- Filion G.J., van Bommel J.G., Braunschweig U., Talhout W., Kind J., Ward L.D., Brugman W., de Castro I.J., Kerkhoven R.M., Bussemaker H.J., van Steensel B. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell.* 2010; 143(2):212-224.
- Frith M.C., Valen E., Krogh A., Hayashizaki Y., Carninci P., Sandelin A. A code for transcription initiation in mammalian genomes. *Genome Res.* 2008;18(1):1-12.
- Granok H., Leibovitch B.A., Shaffer C.D., Elgin S.C. Chromatin. GAGA over GAGA factor. *Curr. Biol.* 1995;5(3):238-241.
- Hammonds A.S., Bristow C.A., Fisher W.W., Weiszmarm R., Wu S., Hartenstein V., Kellis M., Yu B., Frise E., Celniker S.E. Spatial expression of transcription factors in *Drosophila* embryonic organ development. *Genome Biol.* 2013;14(12):R140.
- Harbison C.T., Gordon D.B., Lee T.I., Rinaldi N.J., Macisaac K.D., Danford T.W., Hannett N.M., Tagne J.-B., Reynolds D.B., Yoo J., Jennings E.G., Zeitlinger J., Pokholok D.K., Kellis M., Rolfé P.A., Takusagawa K.T., Lander E.S., Gifford D.K., Fraenkel E., Young R.A. Transcriptional regulatory code of a eukaryotic genome. *Nature.* 2004;431(7004):99-104.
- Hoskins R.A., Carlson J.W., Wan K.H., Park S., Mendez I., Galle S.E., Booth B.W., Pfeiffer B.D., George R.A., Svirskas R., Krzywinski M., Schein J., Accardo M.C., Damia E., Messina G., Méndez-Lago M., de Pablos B., Demakova O.V., Andreyeva E.N., Boldyreva L.V., Marra M., Carvalho A.B., Dimitri P., Villasante A., Zhimulev I.F., Rubin G.M., Karpen G.H., Celniker S.E. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* 2015;25(3):445-458.
- Ingham P.W. The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature.* 1988;335(6185):25-34.
- Iyer E.P.R., Iyer S.C., Sullivan L., Wang D., Meduri R., Graybeal L.L., Cox D.N. Functional genomic analyses of two morphologically distinct classes of *Drosophila* sensory neurons: post-mitotic roles of transcription factors in dendritic patterning. *PloS ONE.* 2013;8(8):e72434.
- Karagodin D.A., Omelina E.S., Fedorova E.V., Baricheva E.M. Identification of functionally significant elements in the second intron of the *Drosophila melanogaster* Trithorax-like gene. *Gene.* 2013;520(2): 178-184.
- Kvon E.Z., Stampfel G., Yanez-Cuna J.O., Dickson B.J., Stark A. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Develop.* 2012;26(9): 908-913.
- Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 2012;9(4):357-359.
- Lelli K.M., Slattery M., Mann R.S. Disentangling the many layers of eukaryotic transcriptional regulation. *Ann. Rev. Genet.* 2012;46(1):43-68.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. & 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England). 2009;25(16):2078-2079.
- McGinnis N., Ragnhildstveit E., Veraksa A., McGinnis W. A cap “n” collar protein isoform contains a selective Hox repressor function. *Development* (Cambridge, England). 1998;125(22):4553-4564.
- Merkulova T.I., Ananko E.A., Ignat'eva E.V., Kolchanov N.A. Regulatory transcription codes in eukaryotic genomes. *Genetika = Genetics* (Moscow). 2013;49(1):37-54. (in Russian)
- Mohler J., Mahaffey J.W., Deutsch E., Vani K. Control of *Drosophila* head segment identity by the bZIP homeotic gene *cnc*. *Development* (Cambridge, England). 1995;121(1):237-247.

- Moorman C., Sun L.V., Wang J., de Wit E., Talhout W., Ward L.D., Greil F., Lu X.-J., White K.P., Bussemaker H.J., van Steensel B. Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 2006;103:12027-12032.
- Nègre N., Brown C.D., Ma L., Bristow C.A., Miller S.W., Wagner U., ... White K.P. A cis-regulatory map of the *Drosophila* genome. *Nature*. 2011;471(7339):527-531. <http://doi.org/10.1038/nature09990>.
- Nègre N., Brown C.D., Shah P.K., Kheradpour P., Morrison C.A., Henikoff J.G., ... White K.P. A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genetics*. 2010;6(1):e1000814. <http://doi.org/10.1371/journal.pgen.1000814>.
- O'Brien T., Wilkins R.C., Giardina C., Lis J.T. Distribution of GAGA protein on *Drosophila* genes in vivo. *Genes Development*. 1995;9(9):1098-1110.
- Ochoa-Espinosa A., Small S. Developmental mechanisms and cis-regulatory codes. *Curr. Opin. Gen. Dev.* 2006;16(2):165-170.
- Pagans S., Ortiz-Lombardia M., Espinas M.L., Bernues J., Azorin F. The *Drosophila* transcription factor tramtrack (TTK) interacts with Trithorax-like (GAGA) and represses GAGA-mediated activation. *Nucl. Acids Res.* 2002;30(20):4406-4413.
- Perelygina L.M., Baricheva E.M., Sebeleva T.E., Kokoza V.A. The evolutionarily conserved gene Nc70F expressed in nerve tissue of *Drosophila melanogaster* encodes a protein homologous to the mouse delta transcription factor. *Genetika = Genetics (Moscow)*. 1993;29(10):1597-1607.
- Rivera-Pomar R., Jäckle H. From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet.* 1996;12(11):478-483.
- Roy S., Ernst J., Kharchenko P.V., Kheradpour P., Nègre N., Eaton M.L., Landolin J.M., Bristow C.A., Ma L.J., Lin M.F., ... Kellis M. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*. 2010;330(6012):1787-1797.
- Schertel C., Albarca M., Rockel-Bauer C., Kelley N.W., Bischof J., Hens K., van Nimwegen E., Basler K., Deplancke B. A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating *Drosophila* wing development. *Genome Res.* 2015;25(4):514-523.
- Segal E., Raveh-Sadka T., Schroeder M., Unnerstall U., Gaul U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature*. 2008;451(7178):535-540.
- Slattery M., Ma L., Spokony R.F., Arthur R.K., Kheradpour P., Kundaje A., Nègre N., Crofts A., Ptashkin R., Zieba J., Ostapenko A., Suchy S., Victorsen A., Jameel N., Grundstad A.J., Gao W., Moran J.R., Rehm E.J., Grossman R.L., Kellis M., White K.P. Diverse patterns of genomic targeting by transcriptional regulators in *Drosophila melanogaster*. *Genome Res.* 2014;24(7):1224-1235.
- Soeller W.C., Oh C.E., Kornberg T.B. Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol. Cell. Biol.* 1993;13(12):7961-7970.
- Tomancak P., Ohler U. Mapping the complexity of transcription control in higher eukaryotes. *Genome Biol.* 2010;11(4):115.
- Veraksa A., McGinnis N., Li X., Mohler J., McGinnis W. Cap "n" collaborates with a small Maf subunit to specify pharyngeal development and suppress deformed homeotic function in the *Drosophila* head. *Development (Cambridge, England)*. 2000;127(18):4023-4037.
- Zhang Y., Liu T., Meyer C.A., Eeckhoutte J., Johnson D.S., Bernstein B.E., Nusbaum C., Myers R.M., Brown M., Li W., Liu X.S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137.

Компьютерный анализ совместной локализации сайтов связывания транскрипционных факторов в геноме по данным ChIP-seq

А.И. Дергилев^{1, 2}✉, А.М. Спицина¹, И.В. Чадаева^{1, 2}, А.В. Свичкарев^{1, 3}, Ф.М. Науменко¹, Е.В. Кулакова¹,
Э.Р. Галиева^{1, 2}, Е.Е. Витяев^{2, 4}, М. Чен⁵, Ю.Л. Орлов^{1, 2}✉

¹ Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

² Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

³ Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого», Санкт-Петербург, Россия

⁴ Федеральное государственное бюджетное учреждение науки Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, Новосибирск, Россия

⁵ Университет Чжэцзян, Ханчжоу, Китай

Разработана компьютерная программа расчета кластеров сайтов связывания различных транскрипционных факторов (ТФ) по данным геномных координат пиков профиля ChIP-seq (Chromatin Immunoprecipitation-sequencing). Рассмотрены статистические особенности распределения сайтов связывания транскрипционных факторов в геноме мыши, полученных с помощью ChIP-seq экспериментов в эмбриональных стволовых клетках. Определены кластеры сайтов, содержащие четыре (и более) сайта связывания различных транскрипционных факторов в геноме мыши, описано их расположение относительно регуляторных районов генов. Подтверждено наличие двух типов совместной локализации сайтов: кластеры, содержащие сайты связывания факторов Oct4, Nanog, Sox2, расположенные в дистальных районах, и кластеры с сайтами связывания n-Мус, c-Мус, находящиеся в основном в промоторных районах генов мыши. Анализ новых данных ChIP-seq по связыванию транскрипционных факторов Nr5a2, Tbx3, Сер, SRF, USF1 в том же типе клеток подтвердил разделение кластеров сайтов связывания транскрипционных факторов на два типа: содержащие сайты связывания регуляторов плюрипотентности (Oct4, Nanog и Sox2) и не включающие их. Разработана компьютерная программа статистической обработки данных о расположении сайтов в генах, использующая экспериментальные данные локализации сайтов, которые получены методами ChIP-seq в геномах мыши и человека. С помощью этой программы выявлены закономерности локализации сайтов связывания транскрипционных факторов различных типов. Рассчитаны расстояния между ближайшими сайтами связывания ТФ группы Oct4, Nanog, Sox2 и сайтами связывания других факторов в кластерах сайтов, которые служат основой для анализа совместного связывания белковых комплексов с ДНК. Рассчитана доля присутствия известных нуклеотидных мотивов сайтов связывания транскрипционных факторов в геномных участках ChIP-seq. Пересчитаны весовые матрицы для таких нуклеотидных мотивов. Показана корреляция присутствия мотивов с интенсивностью связывания ChIP-seq. Программы, реализующие разработанные компьютерные методы оценки кластеризации сайтов связывания различных транскрипционных факторов для новых данных ChIP-seq, доступны по запросу к авторам.

Ключевые слова: сайты связывания; нуклеотидные мотивы; эмбриональные стволовые клетки; поиск закономерностей; ChIP-seq; энхансеры.

Computer analysis of co-localization of transcription factor binding sites in genome by ChIP-seq data

A.I. Dergilev^{1, 2}✉, A.M. Spitsina¹,
I.V. Chadaeva^{1, 2}, A.V. Svichkarev^{1, 3},
F.M. Naumenko¹, E.V. Kulakova¹, E.R. Galieva^{1, 2},
E.E. Vityaev^{2, 4}, M. Chen⁵, Y.L. Orlov^{1, 2}✉

¹ Novosibirsk State University, Novosibirsk, Russia

² Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

³ Peter the Great St. Petersburg Polytechnic University, Saint-Petersburg, Russia

⁴ Institute of Mathematics SB RAS, Novosibirsk, Russia

⁵ Zhejiang University, Hangzhou, China

Statistical features of the distribution of transcription factor binding sites in the mouse genome that are obtained by ChIP-seq experiments in embryonic stem cells have been considered. Clusters of sites that contain four or more different transcription factor binding sites in the mouse genome have been defined, also their location relatively to the regulatory regions of genes has been described. The presence of two types of site co-localization has been shown: clusters containing binding sites for factors Oct4, Nanog, Sox2, located in the distal regions, and clusters containing binding sites n-Myc, c-Myc, mainly located in the promoter regions of mouse genes. Analysis of new ChIP-seq data about binding of transcription factors Nr5a2, Tbx3 in the same cell type has confirmed the division of clusters of transcription factors binding sites into two types: those containing the binding sites of regulators of pluripotency (Oct4, Nanog, and others) and those not. The computer program of the statistical data processing of gene location and chromatin domains that analyzes experimental data of site localization obtained by ChIP-seq in the mouse genome and the human genome has been developed. The presence of preferences at position of transcription factor binding sites of various types has been revealed, the distances between the nearest groups of TF binding sites Oct4, Nanog, Sox2 and TF binding sites n-Myc and c-Myc have been calculated using this program. The presence

of nucleotide motifs of transcription factor binding sites in the selected areas of ChIP-seq has been estimated, nucleotide motifs have been refined. A correlation between the presence of motifs and the intensity of ChIP-seq binding has been shown. Computer methods for estimating the clustering of different transcription factors binding sites for new data ChIP-seq have been developed. Programs are available upon the request to the authors.

Key words: transcription factor binding sites; embryonic stem cells; data mining; regularity discovery; ChIP-seq; enhancers.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Дергилев А.И., Спицина А.М., Чадаева И.В., Свичкарев А.В., Науменко Ф.М., Кулакова Е.В., Галиева Э.Р., Витяев Е.Е., Чен М., Орлов Ю.Л. Компьютерный анализ совместной локализации сайтов связывания транскрипционных факторов в геноме по данным ChIP-seq. Вавиловский журнал генетики и селекции. 2016;20(6):770-778. DOI 10.18699/VJ16.194

HOW TO CITE THIS ARTICLE:

Dergilev A.I., Spitsina A.M., Chadaeva I.V., Svichkarev A.V., Naumenko F.M., Kulakova E.V., Galieva E.R., Vityaev E.E., Chen M., Orlov Y.L. Computer analysis of co-localization of transcription factor binding sites in genome by ChIP-seq data. Vavilovskii Zhurnal Genetiki i Selektzii = Vavilov Journal of Genetics and Breeding. 2016;20(6):770-778. DOI 10.18699/VJ16.194

Исследование регуляции экспрессии генов эукариот в масштабе генома требует изучения сайтов связывания транскрипционных факторов (ССТФ), контролирующих транскрипцию генов, их геномную локализацию, определение генов-мишеней ТФ. В последние годы благодаря методам высокопроизводительного секвенирования ChIP-seq, ChIP-on-chip, ChIP-PET и другим технологиям, сопряженным с иммунопреципитацией хроматина (ChIP – Chromatin ImmunoPrecipitation), появился огромный массив качественно новых данных, позволяющих оценить регуляторный потенциал клетки, в том числе исследовать все сайты связывания заданного транскрипционного фактора в геноме (Орлов, 2014; Игнатьева и др., 2015).

Развитие высокоэффективных экспериментальных методик измерения экспрессии генов, построения карт ДНК-белковых взаимодействий позволяет исследовать совместное расположение сайтов связывания транскрипционных факторов в геноме (Kuznetsov et al., 2007; Chen et al., 2008). Возникают задачи анализа огромных объемов данных и поиска закономерностей организации регуляторных районов генов с помощью статистических, логических и биоинформационных методов (Кулакова и др., 2015; Спицина и др., 2015).

Классическими задачами анализа экспериментальных данных ChIP-seq являются картирование последовательностей прочтений ДНК (ридов), определение пиков геномного профиля связывания, последующая реконструкция нуклеотидных мотивов сайтов связывания по геномным последовательностям, представленным в таких пиках (Орлов, 2014; He et al., 2015). В данной работе мы рассматриваем задачу определения групп (кластеров) сайтов связывания в геноме и анализ расположения мотивов в них.

С появлением первых данных о сайтах связывания ТФ был разработан набор компьютерных методов предсказания сайтов только по нуклеотидной последовательности, поиска закономерностей взаимного расположения сайтов в промоторных районах (Babenko et al., 1999; Витяев и др., 2001; Боева, 2016). Первые работы основывались на выявлении олигонуклеотидных мотивов и были ограничены статистической базой (десятки, в лучшем случае сотни последовательностей в базах данных) (Heinemeyer et al.,

1998). Появление полногеномных методов секвенирования дало толчок к поиску сайтов и регуляторных районов с помощью геномных профилей ChIP-seq, определению предпочтений геномной локализации, нуклеосомной упаковки (Goh et al., 2010; Орлов, 2014).

В этой связи уникальный ресурс по одновременному анализу связывания большого числа различных транскрипционных факторов методами ChIP-seq для анализа как мотивов, так и регуляторных генных сетей представляет работа X. Chen с коллегами (Chen et al., 2008). Эти данные о сайтах связывания послужили основой целого ряда исследований, уточнявших протоколы обработки ChIP-seq, продолжающихся и в настоящее время (Orlov et al., 2009; Kuznetsov et al., 2010; He et al., 2015; Zhang, Wang, 2015). В нашем исследовании рассматривается аспект анализа регуляторных районов на основе кластеризации сайтов связывания различных транскрипционных факторов.

Выбор транскрипционных факторов в упоминаемой выше работе (Chen et al., 2008) был связан с анализом поддержания плюрипотентного состояния клеток. Так, например, транскрипционный фактор Oct4, кодируемый геном *Pou5f1*, – это POU домен-содержащий ТФ, существенный для эмбриональных стволовых клеток (ЭСК) и раннего эмбрионального развития. В частности, Oct4 взаимодействует с Sox2 (HMG-содержащий ТФ), совместно они воздействуют на множество генов в ЭСК человека (Boyer et al., 2005; Loh et al., 2006). В серии работ, включая известную публикацию S. Yamanaka (Takahashi, Yamanaka, 2006), было показано, что Oct4 и Sox2 вместе с c-Myc и Klf4 достаточны для репрограммирования фибробластов в индуцированные плюрипотентные стволовые клетки, которые функционально похожи на ЭСК. Другой транскрипционный фактор – Nanog – представляет собой гомеодомен-содержащий фактор, который может поддерживать состояние плюрипотентности в ЭСК. Однако для этого требуются и другие транскрипционные регуляторы. К ним относятся Esrrb и Zfx, которые регулируют самобновление ЭСК (Ivanova et al., 2006; Loh et al., 2006). Известны также ключевые компоненты сигнальных путей, опосредованных BMP и LIF – Smad1 и STAT3 (Chen et al., 2008). Полногеномный эксперимент ChIP-seq для

фактора Tbx3 в ЭСК мыши показал, что Tbx3 регулирует гены, ассоциированные с состоянием плюрипотентности, и факторы репрограммирования, и, таким образом, относится к группе транскрипционных факторов поддержания плюрипотентности (Han et al., 2010). Еще один важный транскрипционный фактор, E2f1, задействован в регуляции клеточного цикла, где показана ассоциация участков его связывания с промоторными районами генов (Bieda et al., 2006).

Для расширения анализа к данным, полученным в работе X. Chen с коллегами (2008), были добавлены координаты сайтов связывания нескольких транскрипционных факторов, определенные с помощью ChIP-seq в том же типе клеток – Cep, SRF, USF1 (Sirito et al., 1998; Xu et al., 2014; Kuzniewska et al., 2015).

Cep – активатор транскрипции, который связывает ту же последовательность ДНК, что и p53. Он играет важную роль в нормальном развитии, обеспечивает мейозное деление клеток. Следующий транскрипционный фактор, SRF, принадлежит к семейству факторов MADS (MCM1, Agamous, Deficiens и SRF). Этот белок связывается с элементом ответа в сыворотке (SRE) в промоторном участке генов-мишеней. Кроме того, SRF регулирует активность многих ранних генов, например *C-FOS*, и тем самым участвует в регуляции клеточного цикла, апоптоза, клеточного роста и дифференцировки клеток (Kuzniewska et al., 2015). Последний ТФ в нашем добавленном списке – USF1 – кодирует основную спираль лейциновой застёжки и может функционировать в качестве фактора клеточной транскрипции (Sirito et al., 1998). Следует отметить, что пониженный уровень USF1 у мышей увеличивает скорость метаболизма.

С помощью компьютерного анализа данных ChIP-seq, доступных в GEONCBI, построены полногеномные карты сайтов связывания транскрипционных факторов в эмбриональных стволовых клетках в геноме мыши для факторов c-Myc, Oct4, Nanog, Sox2, E2f1, n-Myc, Tbx3, Eset, Nr5a2, Smad2 (Chen et al., 2008; Han et al., 2010; Heng et al., 2010; Lee et al., 2011), а также Cep, SRF, USF1 (Sirito et al., 1998; Xu et al., 2014; Kuzniewska et al., 2015).

В нашей работе были использованы координаты сайтов этих факторов связывания в одном и том же типе клеток, представленные в указанных выше публикациях. Найдена корреляция между присутствием мотивов (процент мотивов в пиках ChIP-seq) и интенсивностью связывания ChIP-seq (высота пика) в геноме мыши. Разработанное программное обеспечение может быть использовано для анализа кластеров сайтов связывания новых наборов транскрипционных факторов в геномах эукариот, включая данные из проектов ENCODE (<https://genome.ucsc.edu/ENCODE/>), modENCODE, FactorBook (<http://www.factorbook.org>). Исследование закономерностей расположения нуклеотидных мотивов и контекстных сигналов в геномной ДНК найденных кластеров сайтов с помощью методов интеллектуального анализа данных и знаний (Data Mining) позволит описать точные закономерности структурной организации регуляторных районов генов, в том числе энхансерных районов в геноме для их аннотации и распознавания на основе нуклеотидных последовательностей.

Материалы и методы

В работе были использованы полногеномные карты сайтов связывания транскрипционных факторов в эмбриональных стволовых клетках, построенных по данным ChIP-seq для c-Myc, Oct4, Nanog, Sox2, E2f1, n-Myc, Tbx3, Eset, Nr5a2, Smad2 в геноме мыши (Chen et al., 2008; Han et al., 2010; Heng et al., 2010; Lee et al., 2011). Были изучены закономерности расположения нуклеотидов в кластерах сайтов с помощью программы Match TM и других программ, разработанных на языке Python 2.7 (IDE PyCharm 4.0.4) с использованием библиотек, реализованных на языке C++. Разработан набор утилит командной строки, при этом каждая утилита (на языке Python) реализует одну из требуемых функций. Был задействован ряд скриптов для симуляции числа кластеров в зависимости от общего числа сайтов и размера генома, в том числе только размеров промоторных областей или заданных районов хромосом. С помощью скриптов в среде R были построены тепловые карты и проведены расчеты корреляций встречаемости пар сайтов.

Поиск мотива выполняли с использованием весовой матрицы из базы данных TRANSFAC (Heinemyer et al., 1998), уточнение – по выравниванию найденных нуклеотидных последовательностей, содержащих мотив в различных наборах данных. Общая схема итеративного компьютерного поиска и уточнения мотивов по данным ChIP-seq представлена на рис. 1.

Результаты и обсуждение

Использованы геномные карты сайтов связывания большого набора транскрипционных факторов в геноме мыши на одном типе клеток (Chen et al., 2008; Han et al., 2010; Heng et al., 2010). Необходимо отметить, что эксперименты иммунопреципитации хроматина могут получать сигнал непрямыми ДНК-белковыми взаимодействиями, поэтому присутствие нуклеотидных мотивов не является необходимым условием и требует дополнительного уточнения. Особенность использованного подхода состоит в изучении одновременного расположения сайтов связывания нескольких различных транскрипционных факторов, предполагается возможность их взаимодействия и согласованного функционирования в исследуемом типе ткани. Показано наличие геномных участков, совместно занятых несколькими сайтами различных факторов, определенных с помощью ChIP-seq. Такие участки можно назвать «горячими точками» совместной локализации ТФ. Они с большой вероятностью представляют собой функционально важные регуляторные участки транскрипции генов. Необходима дальнейшая интеграция данных о связывании сайтов с анализом экспрессии генов (Orlov et al., 2012; Полуниин и др., 2014; Спицина и др., 2015) с учетом дальнедействующих взаимодействий хроматина (Li et al., 2014; Кулакова и др., 2015).

Перерасчет карт локализации сайтов связывания Tbx3 (Han et al., 2010) и исследованных ранее кластеров из работы X. Chen с коллегами (2008) показал совместную локализацию связывания Tbx3 с группой Oct4-Sox2-Nanog (Доп. материалы 1)¹.

¹ Дополнительные материалы 1–4 см. в Приложении 1 по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2016-20/appx5.pdf>

Статистика расположения сайтов связывания транскрипционных факторов

Анализ расположения сайтов показал, что отдельные участки генома заняты одновременно несколькими различными транскрипционными факторами, связанными с геномной ДНК на очень близком расстоянии (десятки нуклеотидов), или даже с перекрытием участков связывания. Некоторые обогащенные связыванием ТФ районы могут появиться по случайным причинам – близкое расположение сайтов еще не означает их функциональной общности или кооперативного связывания. В то же время некоторые геномные районы, обогащенные сайтами связывания (Chen et al., 2008), могут функционировать как дистальные энхансеры и действительно привлекают кооперативно связывающиеся белковые факторы, физически контактирующие друг с другом при связывании с ДНК. Примеры экспериментального определения таких контактирующих белков, образующих комплекс энхансесомы, были показаны ранее (Panne et al., 2007) (рис. 2).

Рассмотрим расположение сайтов, определенных с помощью ChIP-seq, в пределах 500 нт. Для статистического разделения неслучайных комбинаций сайтов от «шума» – ожидаемого по случайным причинам числа кластеров сайтов (групп близко расположенных позиций пиков связывания ChIP-seq) – был разработан алгоритм, принимающий во внимание число связанных районов, интенсивность сигнала ChIP-seq в связывании для каждого ТФ. Первый шаг состоял в формальном определении кластера сайтов связывания. Два участка связывания (пики ChIP-seq) включались в кластер, если центральные позиции пиков были удалены не более чем на заданное расстояние (200 нт) друг от друга.

Для оценки вероятности получения таких комбинаций было построено распределение кластеров, которые могут образоваться по случайным причинам, по размерам, в том числе размерам сайтов и хромосом, с использованием подходов, представленных в работах (Chen et al., 2008; Orlov et al., 2009). В целом три сайта связывания различных ТФ в одном и том же геномном локусе могут рассматриваться как неслучайная комбинация. Для более строгого сравнения использовали размер, занимаемый только промоторами генов, без пересечения (определяли 2.5 Кб перед стартом транскрипции и 500 нт – после него), что значительно меньше размера всего генома, доступного для картирования. Связывание четырех и более ТФ одновременно достаточно для принятия гипотезы на уровне 1 % FDR (с вероятностью ошибки ложного предсказания 1 %) как для проксимальных промоторов, так и для дистальных сайтов.

Пики ChIP-seq (сайты связывания) для фиксированного ТФ, содержащие в окрестности 100 нт пики другого фактора, последовательно кластеризовали друг с другом. Кластер увеличивался до тех пор, пока новые пики уже невозможно было добавить. Ограничением по длине геномного участка, варьируемой в программе при построении кластеров, был размер 500 нт. Для каждого локуса получено описание, сколько сайтов разных ТФ, заданных пиками ChIP-seq, он содержит; названия факторов; высота пиков; его геномные координаты (см. рис. 2).

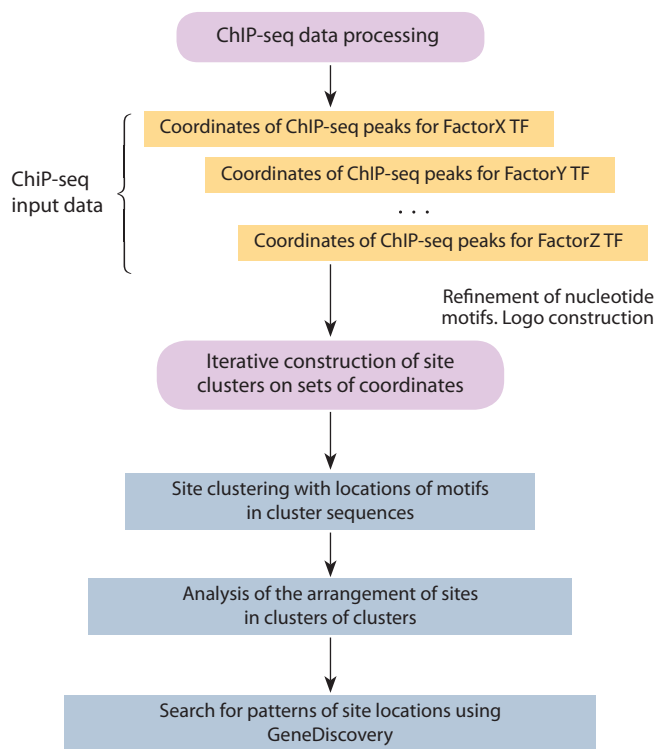


Fig. 1. Data processing: construction of site clusters, search for nucleotide motifs, and analysis of their co-localization.

Рассмотрим геномную локализацию построенных кластеров сайтов связывания транскрипционных факторов. 40 % из них находятся во внутригеномных районах, оставшиеся кластеры располагаются в промоторных районах (37 %) и внутри генов (23 %).

Интересно отметить, что кластеры большего размера, как правило, расположены дистально, лишь менее 20 % кластеров из семи и более ТФ находятся в промоторных районах в сравнении с 40 % кластеров размером не более пяти ТФ. Следовательно, совместная встречаемость ССТФ в кластерах не связана с их расположением в промоторах, где могло бы находиться большинство геномных ССТФ. Разработанная программа позволяет пересчитывать уровень значимости для совместной встречаемости группы сайтов для произвольного набора транскрипционных факторов.

Матрицы совместной локализации сайтов связывания различных транскрипционных факторов и тепловые карты

Разработана программа, позволяющая рассчитать по геномным координатам пересечение расположения сайта связывания транскрипционного фактора с другими сайтами и построить кластеры (рис. 3).

На рис. 3 показано число кластеров в зависимости от размера кластера (числа содержащихся в нем сайтов). Параметр расчета границ расположения в данном случае – 200 нт между координатами сайтов, определенными методом ChIP-seq. Из этого рисунка видно, что число сайтов связывания в кластерах в зависимости от размера кластера

Example of site cluster identification in the genome

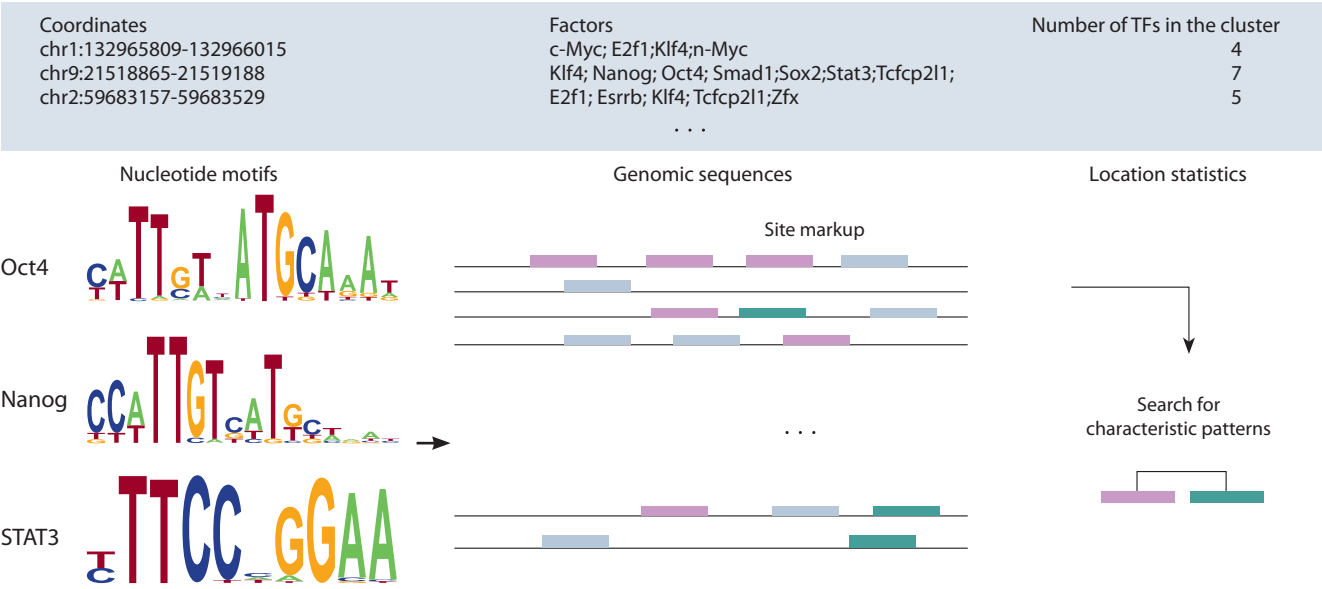


Fig. 2. Schematic presentation of transcription regulation loci containing clusters of transcription factor binding sites recognized according to ChIP-seq data (Chen et al., 2008). Nucleotide motifs of binding sites, their location in regulatory regions, and pattern search.

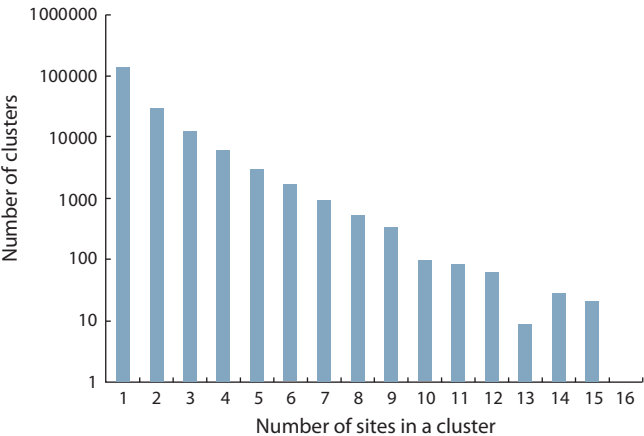


Fig. 3. Distribution of binding site number depending on cluster size.

экспоненциально убывает. Есть единичные кластеры, содержащие одновременно до 15 сайтов связывания различных факторов. В то же время нет геномного района, содержащего одновременно сайты всех 18 факторов, что говорит о функциональных различиях факторов. Подобный результат был получен ранее для набора сайтов связывания 13 факторов, при этом максимальный размер кластера был равен 11 (Orlov et al., 2009) (см. Доп. материалы 2, рис. 1–3).

Для каждой пары транскрипционных факторов был рассчитан коэффициент линейной корреляции Пирсона совместной локализации их сайтов связывания в геноме мыши. Построена матрица частот совместной локализации сайтов различных ТФ по координатам в геноме (рис. 4).

В каждой ячейке такой симметричной матрицы содержится число совпадающих (перекрывающихся) в геномном интервале до 200 нт) сайтов связывания двух факторов. По диагонали располагается число сайтов каждого фактора по отдельности. Далее была построена матрица корреляций. Для каждой пары транскрипционных факторов был подсчитан линейный коэффициент корреляции между строками исходной матрицы. Такая матрица задает меру близости, или ассоциации, между различными транскрипционными факторами. Чем выше коэффициент корреляции, тем ближе располагаются сайты связывания транскрипционных факторов друг к другу относительно других факторов. С использованием коэффициента корреляции расположения сайтов как меру близости факторов с помощью программы в среде R была рассчитана матрица совместной локализации исследованных транскрипционных факторов. Интенсивность такой совместной локализации показана в форме тепловой карты (термокарты) (см. рис. 4).

Подтверждено, что группа кластеров Мус (включая гены с-Мус, n-Мус) характеризуется преимущественно промоторным расположением, а группа Nanog (с другими ключевыми факторами плюрипотентности) – преимущественно дистальным расположением относительно старта транскрипции генов.

Выявление нуклеотидных мотивов в кластерах

Анализ данных ChIP-seq о сайтах связывания нескольких транскрипционных факторов позволяет получить комбинаторную информацию о регуляторном действии сайтов (Chen et al., 2008). Такая информация собирается с помощью определения точного положения нуклеотидных мотивов и их взаимного расположения в регуляторных участках генов (в нашем случае кластеров сайтов).

Проведена обработка исходных данных ChIP-seq, рассчитаны пороги для выделения пиков из профиля и число пиков и выделены их координаты в геноме мыши. Результаты представлены в таблице, в которой показано число сайтов для каждого транскрипционного фактора по отдельности, варьирующее от 1 до 39 тыс. сайтов в геноме. Длина нуклеотидного мотива (весовая матрица) взята из базы данных TRANSFAC (Heinemeyer et al., 1998). Далее был рассчитан процент присутствия нуклеотидных мотивов в пиках ChIP-seq всех анализируемых ТФ.

Процент присутствия вычисляли как отношение числа пиков с хотя бы одним присутствием мотива (выше или равно порога по score, порог был установлен равным 0.8) к общему числу пиков данного ТФ. Процент присутствия мотивов характеризует как информационное содержание мотива, так и качество данных ChIP-seq в наборе. При этом присутствие или отсутствие выраженного нуклеотидного мотива недостаточно для определения качества эксперимента, как показано в ряде работ на том же наборе данных (Kuznetsov et al., 2010). Для пиков ChIP-seq в кластерах, содержащих несколько различных сайтов, процент содержания нуклеотидных мотивов выше, чем для пиков ChIP-seq, расположенных вне кластеров. Таким образом, присутствие других сайтов (пиков ChIP-seq) на близком расстоянии в геноме может свидетельствовать о большей функциональной значимости рассматриваемого регуляторного участка. Дальнейший анализ олигонуклеотидных мотивов в сайтах (Putta et al., 2011) может служить контекстной характеристикой для изучения регуляторных районов транскрипции генов в целом.

Рассмотрим пример уточнения нуклеотидного мотива для сайтов связывания с-Мус. (Полностью данные для набора всех ТФ приведены в Доп. материалах 3.) Пример уточнения нуклеотидных мотивов по сравнению с известными ранее в базах данных (TRANSFAC и JASPAR) показан на рис. 5.

В Доп. материалах 3 (таблица) показано графическое представление мотивов для всех рассмотренных ТФ.

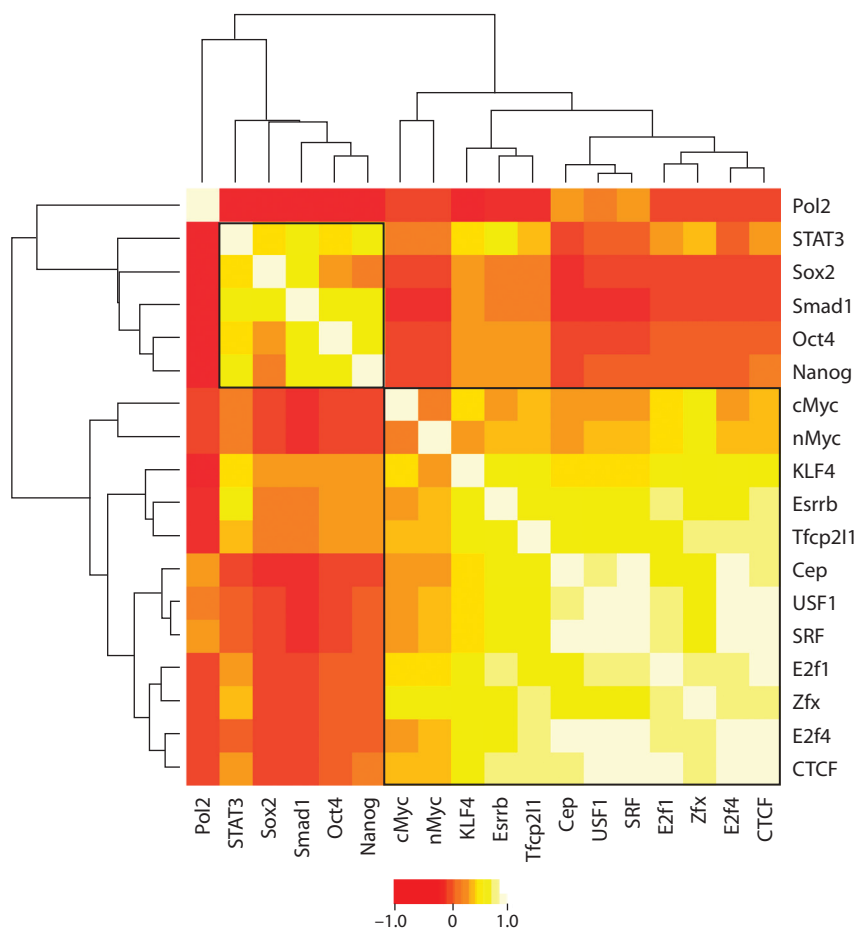


Fig. 4. Heatmap of the colocalization of transcription factor binding sites in the mouse genome according to ChIP-seq data.

Calculation of the colocalization of sites with regard to new factors was performed as in (Chen et al., 2008).

Для сайтов связывания CTCF показана высокая консервативность мотива (Доп. материалы 4, рис. 1). Выполнена качественная оценка присутствия мотивов в зависимости от интенсивности связывания (высоты пиков). Пики ChIP-seq были сортированы по высоте (интенсивности связывания), полученный ранжированный список разбит на квартили. В каждой квартили подсчитаны число и относительная доля найденных мотивов. Показана положительная ассоциация присутствия мотивов в зависимости от квартили (25 % элементов упорядоченного списка). Результаты приведены в Доп. материалах 4, на рис. 2 и 3. Таким образом, анализ нуклеотидных мотивов позволяет повысить качество рассматриваемых данных для дальнейшего исследования при выборе только пиков ChIP-seq, содержащих мотивы.

Были проанализированы состав и предпочтения к совместной локализации транскрипционных факторов в кластерах сайтов. В целом в исследованном наборе выделяются две группы: относящиеся к Мус и относящиеся к Nanog транскрипционные факторы. Среди 18 факторов Nanog, Sox2, Oct4, Smad1, и STAT3 имеют тенденцию встречаться совместно чаще (см. рис. 4). Сайты связывания Zfx, CTCF и E2f1 чаще встречаются совместно в промоторных районах генов, так же, как и кластеры сайтов, содержащие сайты Мус. Недавние исследования (Yanan et al., 2016) показали, что связывание активного CTCF (CCCTC-связывающего фактора), который имеет важное значение для регуляции генов, с CBS (CTCF-сайтом связывания) устанавливает специфическое взаимодействие между энхансерами и промоторами, а три кластера Pcdh $\beta\gamma$ образуют два CCD (CTCF/ когезин-опосредованных домена хрома-

Binding site motifs in ChIP-seq peaks

Transcription factor	Total number of ChIP-seq peaks	Nucleotide motif length	Percentage of present motifs
c-Myc	3 422	19	58.82
CTCF	39 609	8	75.53
E2f1	20 699	12	99.90
E2f4	4 546	10	60.70
Esrrb	21 647	16	83.57
KLF4	10 875	10	27.58
Nanog	10 343	15	58.25
n-Myc	7 182	16	44.45
Oct4	3 761	15	24.42
Pol2	30 877	10	38.57
Smad1	1 126	14	59.42
Sox2	4 526	12	46.02
SRF	2 609	10	43.04
Cep	11 434	10	–
STAT3	2 546	10	–
Tfcp2l1	26 910	10	–
USF1	6 741	10	–
Zfx	10 368	10	–

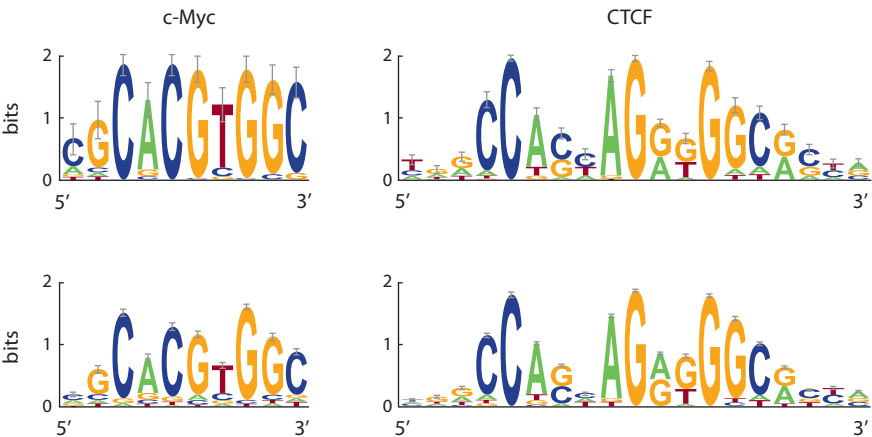


Fig. 5. Known nucleotide motifs (logos) of binding sites of transcription factors c-Myc and CTCF. Above: known motifs of c-Myc and CTCF; below: motifs refined according to nucleotide sequences from ChIP-seq peaks.

тина). Полагается, что CTCF опосредуют специфические взаимодействия между энхансерами с образованием активной площадки транскрипции для экспрессии генов. Кроме того, в ходе исследований (Hutter et al., 2010) было обнаружено умеренное обогащение сайтов связывания фактора CTCF в области геномного импринтинга. Было найдено несколько мотивов в элементах с высокой консервативностью, которые могут выступать в качестве дополнительных регуляторных элементов.

Заключение

Анализ расширенного набора сайтов связывания транскрипционных факторов подтвердил совместную кластеризацию тех сайтов, которые относятся к поддержанию плюрипотентности – Nanog, Sox2, Oct4 – относительно более широкого набора факторов, координаты которых определены экспериментально в статье X. Chen с коллегами (2008). Объединение полногеномных данных картирования сайтов с помощью программ анализа данных ChIP-seq позволяет исследовать генные сети регуляции плюрипотентного состояния стволовых клеток и в дальнейшем рассматривать возможности оптимизации репрограммирования клеток (Orlov et al., 2012).

Использование нуклеотидных последовательностей кластеров сайтов позволяет уточнить нуклеотидные мотивы связывания. С помощью компьютерного исследования комбинаций сайтов связывания по данным ChIP-seq можно определять более сложные закономерности с использованием дополнительных характе-

ристик нуклеотидных последовательностей, в частности оценки сложности текста (Babenko et al., 2015; Safronova et al., 2015), и их функциональную аннотацию, полученную с помощью других пакетов, таких как ICGenomics (Орлов и др., 2012).

Анализ расположения кластеров относительно генов с помощью таких инструментов, как GREAT (Genomic Regions Enrichment of Annotations Tool), дает возможность оценить расположение сайтов связывания отдельных факторов и кластеров сайтов относительно генов (Guo et al., 2012), классифицировать их как промоторные и дистальные, что является дополнительной функциональной характеристикой таких кластеров. Дополнительные характеристики нуклеотидных последовательностей для описания кластеров сайтов связывания, таких как участки низкой сложности текста (Orlov, Potapov, 2004; Orlov et al., 2006), нуклеосомная упаковка (Орлов и др., 2006; Goh et al., 2010), позволяют более точно определить регуляторные районы в геноме и выполнять их поиск на основе непрямых данных, без экспериментов ChIP-seq. Интеграция геномных данных средствами UGENE (Vas'kin et al., 2011; Васькин и др., 2012; Golosova et al., 2014) позволит качественно решать новые задачи комбинаторного анализа сайтов по данным проектов ENCODE (<https://genome.ucsc.edu/ENCODE/>) и FactorBook (<http://www.factorbook.org>) в геноме человека. Развитие методов поиска закономерностей взаимного расположения сайтов связывания в регуляторных районах (комбинаций сайтов ChIP-seq в кластерах) по уточненным мотивам способствует определению более тонких закономерностей работы транскрипционных факторов для регуляции транскрипции генов (Boeva, 2016).

Acknowledgments

We are grateful to Drs. Yu. Yu. Vas'kin, K.S. Bekker, N.S. Safronova, and M.S. Evdokimov for fruitful discussion. Computation was done at the Bioinformatics Shared Access Center, SB RAS, and the Siberian Supercomputer Center, SB RAS. The study of transcription factor binding sites was supported by the Russian Foundation for Basic Research, project 14-04-01906. The development of programs for genomic data analysis was supported by the Russian Science Foundation, project 14-24-00123.

Conflict of interest

The authors declare no conflict of interest.

References

- Babenko V.N., Kosarev P.S., Vishnevsky O.V., Levitsky V.G., Basin V.V., Frolov A.S. Investigating extended regulatory regions of genomic DNA sequences. *Bioinformatics*. 1999;15(7-8):644-653. DOI 10.1093/bioinformatics/15.7.644.
- Babenko V.N., Matvienko V.F., Safronova N.S. 19 Implication of transposons distribution on chromatin state and genome architecture in human. *J. Biomol. Struct. Dyn.* 2015;33(1):10-11. DOI 10.1080/07391102.2015.1032559.
- Bieda M., Xu X., Singer M.A., Green R., Farnham P. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res.* 2006;16(5):595-605. DOI 10.1101/gr.4887606.
- Boeva V. Analysis of genomic sequence motifs for deciphering transcription factor binding and transcriptional regulation in eukaryotic cells. *Front. Genet.* 2016;7:24. DOI 10.3389/fgene.2016.00024.

- Boyer L.A., Lee T.I., Cole M.F., Johnstone S.E., Levine S.S., Zuckerman J.P., Guenther M.G., Kumar R.M., Murray H.L., Jenner R.G., Gifford D.K., Melton D.A., Jaenisch R., Young R.A. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005;122(6):947-956. DOI 10.1016/j.cell.2005.08.020.
- Chen X., Xu H., Yuan P., Fang F., Huss M., Vega V.B., Wong E., Orlov Y.L., Zhang W., Jiang J., Loh Y.H., Yeo H.C., Yeo Z.X., Narang V., Govindarajan K.R., Leong B., Shahab A., Ruan Y., Bourque G., Sung W.K., Clarke N.D., Wei C.L., Ng H.H. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008;133(6):1106-1117. DOI 10.1016/j.cell.2008.04.043.
- Goh W.S., Orlov Y., Li J., Clarke N.D. Blurring of high-resolution data shows that the effect of intrinsic nucleosome occupancy on transcription factor binding is mostly regional, not local. *PLoS Comput. Biol.* 2010;6(1):e1000649. DOI 10.1371/journal.pcbi.1000649.
- Golosova O., Henderson R., Vas'kin Yu., Gabrieli A., Grekhov G., Nagarajan V., Oler A.J., Quiñones M., Hurt D., Fursov M., Huyen Y. Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. *Peer J*. 2014;2:e644. DOI 10.7717/peerj.644.
- Guo Y., Mahony S., Gifford D.K. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.* 2012;8(8):e1002638. DOI 10.1371/journal.pcbi.1002638.
- Han J., Yuan P., Yang H., Zhang J., Soh B.S., Li P., Lim S.L., Cao S., Tay J., Orlov Y.L., Lufkin T., Ng H.H., Tam W.L., Lim B. Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature*. 2010;463(7284):1096-1100.
- He X., Cicek A.E., Wang Y., Schulz M.H., Le H.-S., Ziv B.-J. De novo ChIP-seq analysis. *Genome Biol.* 2015;16(1):205. DOI 10.1186/s13059-015-0756-4.
- Heinemeyer T., Wingender E., Reuter I., Hermjakob H., Kel A.E., Kel O.V., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., Kolpakov F.A., Podkolodny N.L., Kolchanov N.A. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res.* 1998;26(1):362-367. DOI 10.1093/nar/26.1.362.
- Heng J.C., Feng B., Han J., Jiang J., Kraus P., Ng J.H., Orlov Y.L., Huss M., Yang L., Lufkin T., Lim B., Ng H.H. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell*. 2010;6(2):167-174. DOI 10.1016/j.stem.2009.12.009.
- Hutter B., Bieg M., Helms V., Paulsen M. Imprinted genes show unique patterns of sequence conservation. *BMC Genomics*. 2010;11:649. DOI 10.1186/1471-2164-11-649.
- Ignatieva E.V., Podkolodnaya O.A., Orlov Yu.L., Vasiliev G.V., Kolchanov N.A. Regulatory genomics: Combined experimental and computational approaches. *Rus. J. Genet.* 2015;51(4):334-352. DOI 10.1134/S1022795415040067.
- Ivanova N., Dobrin R., Lu R., Kotenko I., Levorse J., DeCoste C., Schafer X., Lun Yi., Lemischka I.R. Dissecting self-renewal in stem cells with RNA interference. *Nature*. 2006;442(7102):533-538. DOI 10.1038/nature04915.
- Kuznetsov V.A., Orlov Yu.L., Wei C.L., Ruan Y. Computational analysis and modeling of genome-scale avidity distribution of transcription factor binding sites in chip-pet experiments. *Genome Inform.* 2007;19:83-94.
- Kulakova E.V., Spitsina A.M., Orlova N.G., Dergilev A.I., Svichkarov A.V., Safronova N.S., Chernykh I.G., Orlov Yu.L. Programs for the analysis of genomic sequence data obtained by ChIP-seq, ChIA-PET, and Hi-C technologies. *Programmye Sistemy: Teoriya i Prilozheniya = Program Systems: Theory and Applications*. 2015;6(2(25)):129-148. (in Russian)
- Kuznetsov V.A., Singh O., Jenjaroenpun P. Statistics of protein-DNA binding and the total number of binding sites for a transcription factor in the mammalian genome. *BMC Genomics*. 2010;11(1):S12. DOI 10.1186/1471-2164-11-S1-S12.
- Kuzniowska B., Nader K., Dabrowski M., Kaczmarek L., Kalita K. Adult deletion of SRF increases epileptogenesis and decreases ac-

- tivity-induced gene expression. *Mol. Neurobiol.* 2015;1-16. DOI 10.1007/s12035-014-9089-7.
- Lee K.L., Lim S.K., Orlov Y.L., Yit le Y., Yang H., Ang L.T., Poellinger L., Lim B. Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLoS Genet.* 2011;7(6):e1002130. DOI 10.1371/journal.pgen.1002130.
- Li G., Cai L., Chang H., Hong P., Zhou Q., Kulakova E.V., Kolchanov N.A., Ruan Y. Chromatin interaction analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics.* 2014;15(12):S11. DOI 10.1186/1471-2164-15-S12-S11.
- Loh Y.H., Wu Q., Chew J.L., Vega V.B., Zhang W., Chen X., Bourque G., George J., Leong B., Liu J., Wong K.Y., Sung K.W., Lee C.W., Zhao X.D., Chiu K.P., Lipovich L., Kuznetsov V.A., Robson P., Stanton L.W., Wei C.L., Ruan Y., Lim B., Ng H.H. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* 2006;38(4):431-440. DOI 10.1038/ng1760.
- Orlov Yu.L. Computer-assisted study of the regulation of eukaryotic gene transcription on the base of data on chromatin sequencing and precipitation. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding.* 2014;18(1):193-206. (in Russian)
- Orlov Yu.L., Bragin A.O., Medvedeva I.V., Gunbin K.V., Demenkov P.S., Vishnevsky O.V., Levitsky V.G., Oshchepkov D.Y., Podkolodnyy N.L., Afonnikov D.A., Grosse I., Kolchanov N.A. IC-Genomics: Software for analysis of symbol genomics sequences. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding.* 2012;16(4/1):732-741. (in Russian)
- Orlov Yu.L., Huss M.E., Joseph R., Xu H., Vega V.B., Lee Y.K., Goh W.S., Thomsen J.S., Cheung E.C., Clarke N.D., Ng H.H. Genome-wide statistical analysis of multiple transcription factor binding sites obtained by ChIP-seq technologies. *Proc. 1st ACM Workshop on Breaking Frontiers of Computational Biology (CompBio '09).* ACM, New York, N.Y., 2009;11-18.
- Orlov Yu.L., Levitsky V.G., Smirnova O.G., Podkolodnaya O.A., Khlebodarova T.M., Kolchanov N.A. Statistical analysis of nucleosome formation sites. *Biofizika = Biophysics.* 2006;51(4):608-614. (in Russian)
- Orlov Yu.L., Potapov V.N. Complexity: an internet resource for analysis of DNA sequence complexity. *Nucleic Acids Res.* 2004;32:W628-W633. DOI 10.1093/nar/gkh466.
- Orlov Yu.L., Te Boekhorst R., Abnizova I.I. Statistical measures of the structure of genomic sequences: entropy, complexity, and position information. *J. Bioinform. Comput. Biol.* 2006;4:523-536. DOI 10.1142/S0219720006001801.
- Orlov Yu., Xu H., Afonnikov D., Lim B., Heng J.C., Yuan P., Chen M., Yan J., Clarke N., Orlova N., Huss M., Gunbin K., Podkolodnyy N., Ng H.H. Computer and statistical analysis of transcription factor binding and chromatin modifications by ChIP-seq data in embryonic stem cell. *J. Integr. Bioinform.* 2012;9(2):211. DOI 10.2390/biecoll-jib-2012-211.
- Panne D., Maniatis T., Harrison S.C. An atomic model of the interferon-beta enhanceosome. *Cell.* 2007;129(6):1111-1123. DOI 10.1016/j.cell.2007.05.019.
- Polunin D.A., Shtager I.A., Efimov V.M. JACOBI 4 software for multivariate analysis of microarray data. *Vestnik NGU. Ser. Informatsionnye tekhnologii = Novosibirsk State University Journal of Information Technologies.* 2014;12(2):90-98. (in Russian)
- Putta P., Orlov Yu.L., Podkolodnyy N.L., Mitra C.K. Relatively conserved common short sequences in transcription factor binding sites and miRNA. *Vavilov Journal Genetics and Breeding.* 2011;15(4):750-756.
- Safronova N.S., Babenko V.N., Orlov Yu.L. 117 Analysis of SNP containing sites in human genome using text complexity estimates. *J. Biomol. Struct. Dyn.* 2015;33(1):73-74. DOI 10.1080/07391102.2015.1032750.
- Sirito M., Lin Q., Deng J.M., Behringer R.R., Sawadogo M. Overlapping roles and asymmetrical cross-regulation of the USF proteins in mice. Overlapping roles and asymmetrical cross-regulation of the USF proteins in mice. *Proc. Natl. Acad. Sci. USA.* 1998;95(7):3758-3763.
- Spitsina A.M., Orlov Yu.L., Podkolodnaya N.N., Svichkarev A.V., Dergilev A.I., Chen M., Kuchin N.V., Chernych I.G., Glinskiy B.M. Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA sequencing. *Programmye Sistemy: Teoriya i Prilozheniya = Program Systems: Theory and Applications.* 2015;6(1(23)):157-174. (in Russian)
- Takahashi K., Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006;126(4):663-676. DOI 10.1016/j.cell.2006.07.024.
- Vas'kin Yu., Khomicheva I.V., Ignatieva E.V., Vityaev E.E. Expert discovery and UGENE integrated system for intelligent analysis of regulatory regions of genes. *In Silico Biol.* 2011-2012;11(3-4):97-108. DOI 10.3233/ISB-2012-0448.
- Vas'kin Yu.Yu., Khomicheva I.V., Ignatieva E.V., Vityaev E.E. Sequence analysis of regulatory regions of genes with the Expert Discovery relational system built in package UGENE. *Vestnik NGU. Ser. Informatsionnye tekhnologii = Novosibirsk State University Journal of Information Technologies.* 2012;10(1):73-86. (in Russian)
- Vityaev E.E. Izvlechenie znaniy iz dannyykh. Kompyuternoe poznanie. Modeli kognitivnykh protsessov [Data mining: Computerized cognition. Models of cognitive processes]. Novosibirsk: Novosibirsk State University Publ., 2006. (in Russian)
- Vityaev E.E., Orlov Yu.L., Vishnevsky O.V., Belenok A.S., Kolchanov N.A. Computer system "Gene Discovery" to search for patterns in eukaryotic regulatory nucleotide sequences. *Molekulyarnaya biologiya = Molecular Biology (Moscow).* 2001;35(6):952-960. (in Russian)
- Xu D., Wei G., Lu P., Luo J., Chen X., Skogerb G., Chen R. Analysis of the p53/CEP-1 regulated non-coding transcriptome in *C. elegans* by an NSR-seq strategy. *Protein Cell.* 2014;5(10):770-782. DOI 10.1007/s13238-014-0071-y.
- Yanan Z., Quan X., Ya G., Qiang W. Characterization of a cluster of CTCF-binding sites in a protocadherin regulatory region. *Yi Chuan.* 2016;38(4):323-336. DOI 10.16288/j.yczs.16-037.
- Zhang Y., Wang P. A fast cluster motif finding algorithm for ChIP-Seq data sets. *Biomed. Res. Int.* 2015;2015:218068. DOI 10.1155/2015/218068.



Оценка трансляционной значимости характеристик нуклеотидной последовательности мРНК млекопитающих на основе данных рибосомного профилирования

О.А. Волкова¹✉, Ю.В. Кондрахин^{2, 3}, Р.Н. Шарипов^{2, 3, 4}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное бюджетное научное учреждение «Институт вычислительных технологий Сибирского отделения Российской академии наук», Новосибирск, Россия

³ Общество с ограниченной ответственностью «Институт системной биологии», Новосибирск, Россия

⁴ Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Известно, что характеристики 5'-нетранслируемой последовательности (5'-НТП) мРНК могут оказывать влияние на эффективность и специфичность инициации трансляции. Ранее знания о характеристиках 5'-НТП были получены теоретически и в экспериментах *in vitro* для мРНК отдельных генов, что не давало возможности оценить реальную трансляционную значимость ее параметров. Для выявления трансляционно-значимых характеристик 5'-НТП необходимо проанализировать их связь с трансляционной активностью соответствующих мРНК. Однако до недавнего времени доступные технологии не позволяли получить широкогеномные экспериментальные данные по эффективности трансляции. Благодаря появившейся технологии профилирования рибосом такие данные были получены для мРНК ряда эукариот. Использование их позволяет оценивать и выявлять трансляционно-значимые параметры мРНК, а также предсказывать эффективность трансляции мРНК на основании характеристик ее нуклеотидной последовательности. Цель нашей работы – определение трансляционной значимости отдельных характеристик нуклеотидной последовательности 5'-НТП мРНК на основании соответствующих экспериментальных данных по эффективности трансляции, рибосомному профилированию. Проведен статистический анализ отдельных характеристик нуклеотидных последовательностей мРНК человека и мыши; выявлена их взаимосвязь с соответствующими данными рибосомного профилирования. Были отобраны трансляционно-значимые параметры мРНК, тенденция влияния на эффективность трансляции которых наиболее значима и одинакова для всех трех проанализированных выборок: пурин в –3-позиции стартового кодона, вышележащие стартовые кодоны AUG в 5'-НТП, и комплементарные нуклеотиды G+C в составе 5'-НТП снижают эффективность трансляции; олигонуклеотид CCGCCA в районе 5'-НТП и олигонуклеотиды AAGAAA, AAGAAG, AAGCAG, AAAAAG в составе белок-кодирующей последовательности – усиливают. Разработаны с помощью платформы BioUML набор инструментов, позволяющий анализировать трансляционную значимость отдельных 5'-НТП мРНК, и программа для предсказания эффективности трансляции мРНК на основании ее нуклеотидной последовательности.

Ключевые слова: профилирование рибосом; RiboSeq; 5'-НТП; мРНК; инициация трансляции.

Estimation of translational importance of mammalian mRNA nucleotide sequence characteristics based on ribosome profiling data

O.A. Volkova¹✉, Yu.V. Kondrakhin^{2, 3}, R.N. Sharipov^{2, 3, 4}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Institute of Computational Technologies SB RAS, Novosibirsk, Russia

³ Institute of Systems Biology, Ltd, Novosibirsk, Russia

⁴ Novosibirsk State University, Novosibirsk, Russia

It is known that the 5' untranslated region (5' UTR) mRNA characteristics can influence translation initiation efficiency and specificity. Previous knowledge about 5' UTR characteristics was obtained theoretically and *in vitro* for mRNA of individual genes. It did not allow systematic analysis of mRNA translationally important parameters. To identify the above mentioned 5' UTR characteristics, it is necessary to analyze their relationships with the translational activity of the corresponding mRNAs. Until recently, there were no experimental data on translation efficiency. Thanks to ribosome profiling technology, genome-wide experimental data of translation efficiency have been obtained for many eukaryotic mRNAs. Now it seems to be possible to reveal translationally important mRNA parameters and predict translation efficiency based on their nucleotide sequences. The aim of this study was to determine the translational significance of individual 5' UTR characteristics in accordance with experimental ribosome profiling data. A statistical analysis was carried out for revealing relationships between the human and mouse mRNA nucleotide sequence characteristics and ribosome profiling data. Some of the mRNA parameters influencing translation efficiency were most significant, and the same trends for all three samples analyzed were revealed: a purine at start codon context position –3, upstream AUG pre-

sence and G+C complementary nucleotide concentration reduce translation efficiency; whereas gexonucleotides CCGCCA (5' UTR) and AAGAAA, AAGAAG, AAGCAG, AAAAAAG (CDS) increase translation efficiency. A toolkit that allows analyzing the importance of 5' UTR characteristics and a program for prediction of translation efficiency were developed on the base of the BioUML platform.

Key words: ribosome profiling; RiboSeq; 5' UTR; mRNA; translation initiation.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Волкова О.А., Кондрахин Ю.В., Шарипов Р.Н. Оценка трансляционной значимости характеристик нуклеотидной последовательности мРНК млекопитающих на основе данных рибосомного профилирования. Вавиловский журнал генетики и селекции. 2016;20(6):779-786. DOI 10.18699/VJ16.195

HOW TO CITE THIS ARTICLE:

Volkova O.A., Kondrakhin Yu.V., Sharipov R.N. Estimation of translational importance of mammalian mRNA nucleotide sequence characteristics based on ribosome profiling data. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):779-786. DOI 10.18699/VJ16.195

Трансляция оказывает существенное влияние на регуляцию экспрессии генов. Наиболее значима стадия инициации, а элонгация, трансляционные паузы и другие этапы не играют существенной роли (Ingolia et al., 2011). Эффективность трансляции является консервативной в плане эволюции. Это предполагает, что эффективность трансляции обусловлена первичной последовательностью мРНК и ее взаимодействиями с соответствующими трансляционными факторами и 40S субъединицами рибосом.

мРНК эукариотических генов различаются по интенсивности синтеза белка в процессе трансляции. 5'-нетранслируемая последовательность мРНК (5'-НТП, лидерная последовательность) участвует во взаимодействии с факторами аппарата трансляции и 40S субъединицами рибосом (Hinnebusch, Lorsch, 2012; Singh et al., 2012). Показано, что характеристики нуклеотидной последовательности 5'-НТП оказывают существенное влияние на эффективность инициации трансляции (Kozak, 2005; Kochetov, 2008). Были выявлены некоторые из таких функционально значимых характеристик, в частности стабильная вторичная структура (Ding et al., 2012; Li et al., 2012) и потенциальные стартовые кодоны в районе 5'-НТП, которые могут ингибировать трансляцию (Kozak, 2005). Однако имеющаяся информация о структурно-функциональной организации 5'-НТП недостаточна и требует дальнейшего уточнения.

Контекст стартового кодона трансляции в эукариотических мРНК. Считается, что распознавание триплета AUG в качестве стартового кодона зависит от нуклеотидного окружения (контекста). Экспериментально было подтверждено, что аденин или гуанин в позиции -3 (A/G $^{-3}$) и, в меньшей степени, гуанин в позиции $+4$ (G $^{+4}$) относительно кодона AUG способствуют распознаванию его как стартового (оптимальный контекст: R $^{-3}$ NNAUGG $^{+4}$, R=A или G) (Kozak, 1981, 1997). Показано, что у млекопитающих в дискриминации между AUG кодонами в оптимальном и неоптимальном контекстах участвует эукариотический фактор инициации трансляции eIF1 (Pestova, Kolupaeva, 2002; Takacs et al., 2011). На основе экспериментальных данных предполагается, что за узнавание 43S инициаторным комплексом пуринового нуклеотида в позиции -3 ответственно взаимодействие

данного нуклеотида с α -субъединицей eIF2, а за узнавание гуанина в позиции $+4$, возможно, ответственно его взаимодействие с нуклеотидами A1818-A1819 в спирали 44 18S рРНК (Pisarev et al., 2006).

Оценки влияния нуклеотидов белок-кодирующей последовательности (БКП) на распознавание стартового кодона варьируют от существенного (Grunert, Jackson, 1994; Niimura et al., 2003; Nakagawa et al., 2008) и ограничено значимого (только G в позиции $+4$; (Kozak, 1997)) до незначительного (Harkins et al., 2005; Xia et al., 2007). Показано, что существуют устойчивые комбинации нуклеотидов в 5'- и 3'-частях контекста стартового кодона, к которым относятся варианты оптимального контекста AnnAUGn и GnnAUGG (Volkova, Kochetov, 2010).

Встречаемость триплетов AUG в составе 5'-НТП. В рамках модели линейного сканирования считается, что в составе 5'-НТП эукариотических мРНК не должны содержаться триплеты AUG (upstream AUG, uAUG), поскольку часть рибосом может распознавать их как стартовые кодоны и это будет снижать эффективность трансляции основной рамки считывания (Kozak, 2005; Kochetov, 2008). Ранее показано, что от 19 до 48 % 5'-НТП мРНК разных видов эукариот содержат uAUG, что значительно реже относительно промоторов, трейлеров и интронов (Rogozin et al., 2001; Volkova, Kochetov, 2010).

Контекстные особенности эукариотических 5'-НТП способствуют формированию менее стабильной вторичной структуры. Считается, что повышенное содержание G и C коррелирует со способностью нуклеотидной последовательности РНК формировать стабильную вторичную структуру, поскольку комплементарное взаимодействие G-C-пар наиболее энергетически выгодно. Вторичная структура в составе 5'-НТП препятствует сканированию мРНК 40S субъединицами рибосом и снижает эффективность инициации трансляции (Kozak, 2005). Так, обнаружено, что 5'-НТП характеризуются даже более высоким содержанием G+C в сравнении с другими функциональными районами мРНК генов. Однако 5'-НТП человека и мыши характеризовались большим дисбалансом в содержании комплементарных нуклеотидов по сравнению с другими функциональными районами генов (Volkova, Kochetov, 2010). Также показано, что вторичная структура, лежащая ниже стартового кодона

в неоптимальном контексте в позициях 13–17, может способствовать его распознаванию 40S субъединицами рибосом (Kochetov et al., 2007).

Ранее было найдено (Kochetov et al., 1998), что трансляционно-значимые характеристики мРНК (контекст старт-кодона, размер 5'-НТП, потенциал формирования вторичной структуры) у выборки высокоэкспрессирующихся генов млекопитающих оптимизированы. мРНК с оптимальным контекстом стартового кодона характеризуются более короткими 5'-НТП, отсутствием (или меньшим числом) uAUG, и они менее склонны к формированию стабильной вторичной структуры в лидерном районе. Однако данные о характеристиках 5'-НТП в основном были получены теоретически и в экспериментах *in vitro* для мРНК отдельных генов, что не позволяло оценить реальную трансляционную значимость ее параметров.

В последние несколько лет благодаря технологии профилирования рибосом (ribosome profiling technique, RiboSeq) (Siwiak, Zielenkiewicz, 2010; Ingolia et al., 2011; Schwanhaussner et al., 2011; Weiss, Atkins, 2011; Lee et al., 2012; Michel et al., 2012; и др.) были получены экспериментальные данные по эффективности трансляции мРНК ряда эукариот. Профилирование рибосом – «глобальный снимок всех рибосом, загруженных на мРНК в клетке в одно мгновение», – производится путем количественного преобразования в ДНК защищенных рибосомами от рибонуклеазы участков мРНК и секвенирования всех ДНК.

Процедура рибосомного профилирования осуществляется с помощью соответствующих ингибиторов инициации трансляции (например, харрингтонин) или элонгации (например, циклогексимида) или сначала инициации, а затем элонгации. Обработка харрингтонином ингибирует формирование первой пептидной связи и, таким образом, способствует позиционированию рибосом на стартовых кодонах; циклогексимид ингибирует элонгацию путем подавления перемещения пептидил-тРНК из акцепторного в донорный сайт большой субъединицы рибосомы.

Ранее в мировой практике оценивали уровень трансляции мРНК с помощью анализа микрочипов мРНК, выделенных из полисом (Johannes et al., 1999; Zong et al., 1999; Na, Lee, 2010), и профилирования с использованием аффинной очистки эпитопом-связанных рибосом (Heiman et al., 2008; Sanz et al., 2009). Однако эти методы не дают позиционной и количественной информации в отличие от метода профилирования рибосом. Данная техника важна для аннотирования генома и исследования регуляции экспрессии.

Рибосомное профилирование широко используется для исследования регуляции трансляции мРНК. Так, технология RiboSeq позволила обнаружить новые транслируемые мРНК и альтернативные сайты инициации трансляции (Liu et al., 2011; Michel et al., 2012), в том числе в зависимости от фазы клеточного цикла (Brar et al., 2012), в условиях стресса (Gerashchenko et al., 2012), локализации трансляции в клетке (Reid, Nicchitta, 2012), апоптозе (Zhu et al., 2014). Однако существуют и недостатки данной технологии (Gerashchenko, Gladyshev, 2014).

Выявление трансляционно-значимых характеристик 5'-НТП мРНК важно для предсказания трансляционной

активности мРНК (Simon, Ruckenstein, 1966; Heinrich, Rapoport, 1980; Godefroy-Colburn, Thach, 1981; Zouridis, Hatzimanikatis, 2007; Dimelow, Wilkinson, 2009), особенностей контроля экспрессии генов на пост-транскрипционном уровне, а также связи между мутациями в этом районе гена и патологическими состояниями. В этой работе мы используем данные RiboSeq для выявления трансляционно-значимых характеристик мРНК и предсказания эффективности трансляции мРНК по ее нуклеотидной последовательности.

Материалы и методы

Для выявления взаимосвязи между отдельными трансляционно-значимыми характеристиками последовательностей мРНК и соответствующими экспериментально найденными с помощью технологии рибосомного профилирования RiboSeq эффективностями трансляции (количество ридов), известными из литературы и базы данных GEO (<http://www.ncbi.nlm.nih.gov/geo/>), нами были составлены три следующие выборки мРНК с обработанными данными RiboSeq: GSE30839 (Set 1, *Mus musculus*, эмбриональные стволовые клетки) (Ingolia et al., 2011); GSE31539 (Set 2, *Homo sapiens*, HEK293) (Reid, Nicchitta, 2012); GSE37744 (Set3, *Homo sapiens*, HEK293) (Ingolia et al., 2012).

В качестве трансляционно-значимых характеристик мРНК мы использовали: 1) длину 5'-НТП; 2) наличие оптимального контекста аденина или гуанина в позиции –3 контекста стартового кодона, A/G–3; 3) наличие гуанина в позиции +4 контекста стартового кодона, G⁺4; 4) содержание кодонов AUG в 5'-НТП, [uAUG]; 5) концентрацию отдельных гексануклеотидов в районах [–100, –1] и [3, 103] относительно стартового кодона; 6) содержание комплементарных нуклеотидов гуанина и цитозина в 5'-НТП, ([G+C]); 7) индексы комплементарности (IC) в позициях [–15, 1] и [1, 20] относительно стартового кодона; 8) длину 3'-НТП; 9) концентрацию комплементарных нуклеотидов C+G в 3'-НТП; 10) длину мРНК; 11) концентрацию комплементарных нуклеотидов C+G в мРНК. Для оценки эффективности трансляции использовали данные RiboSeq: нормированное количество ридов, *gkpm* (reads per kilobase of transcript per million mapped reads) после обработки харрингтонином.

Были отобраны два специфических набора гексануклеотидов, которые наиболее часто встречались в выборке мРНК, с высокими значениями *gkpm* и, соответственно, высокой эффективностью инициации трансляции: GCCGCC, CGCCGC, CCGCCA, CCGCCG, CCGCGC, CCCGCC, CTCCGC, CGCGCC и AAGAAG, GAAGAA, CAAGAA, AGAAGC, CCAAGA, AAGCAG, AGAAGA, AAGAAA, AAAAAG, AAGAAC для районов [–100, –1] и [3, 103] относительно стартового кодона соответственно (см. табл. 4).

Индексы комплементарности IC[–15, 1] и IC[1, 20] вычисляли в районах [–15, 1] и [1, 20] соответственно, относительно стартовых кодонов. Комплементарный индекс IC определяют как коэффициент корреляции между частотами комплементарных триплетов. Предполагается, что чем выше индекс IC, тем более стабильная вторичная структура формируется исследуемым районом.

Для оценки и предсказания влияния характеристик нуклеотидной последовательности 5'-НТП мРНК млекопитающих на ее трансляционную эффективность, полученную с помощью рибосомного профилирования, были использованы методы регрессионного, дискриминантного и кластерного анализов.

Для оценки качества предсказания линейной регрессии, построенной методом наименьших квадратов, использовали коэффициент корреляции между наблюдаемыми и предсказанными эффективностями трансляции. Необходимо отметить, что для построения наиболее эффективной регрессионной функции нет необходимости использования всех отобранных трансляционно-значимых характеристик, например, вследствие наличия мультиколлинеарности (т. е. характеристики мРНК могут коррелировать между собой).

Работа выполнялась на основе платформы BioUML (<http://www.biouml.org>). Были использованы следующие ее возможности: поддержка работы со многими базами данных; импорт/экспорт данных в разных форматах, работа с таблицами и выборками; поддержка DAS протокола (<http://www.biodas.org>) для доступа к последовательностям и их аннотациям; геномный браузер, обеспечивающий широкие возможности интерактивной визуализации последовательностей, их аннотаций и данных NGS (поддержка и визуализация данных в форматах SAM/BAM); разнообразные встроенные методы анализа данных; интеграция с R/Bioconductor и Galaxy (<https://main.g2.bx.psu.edu/>) позволяют использовать внутри BioUML множество других программ, в первую очередь для выравнивания ридов – BWA, Bowtie и др.

Результаты

Предсказание эффективности трансляции на основании трансляционно-значимых характеристик с помощью дискриминантного анализа. Для проведения дискриминантного анализа мы отобрали по 1000 мРНК с высокими и низкими значениями *grkm* (с высоким и низким уровнем трансляции соответственно) из выборки Set 1. Использовали следующие характеристики мРНК: [G+C] в 5'-НТП, длина 5'-НТП, A/G⁻³, IC [-15, 1], IC [1, 20], [uAUG], G⁺⁴, описанные в литературе как трансляционно-значимые (табл. 1). Коэффициенты линейной дискриминантной функции определяли стандартным образом.

Среди 2000 проанализированных мРНК из выборки Set 1 с высокими и низкими трансляционными эффективностями при проведении дискриминантного анализа удалось правильно классифицировать 74.8 % (табл. 2).

Регрессионный анализ характеристик 5'-НТП мРНК, связанных с функционированием сигнала инициации трансляции. Проведен регрессионный анализ с целью выявить трансляционно-значимые характеристики 5'-НТП мРНК на основании данных рибосомного профилирования (нормированное количество ридов, *grkm*) для выборок Set 1, Set 2 и Set 3 (табл. 3). При построении регрессионной функции были отобраны характеристики мРНК, определенные теоретически или в отдельных экспериментах как трансляционно-значимые. Исходный набор характеристик содержал все (см. Материалы и методы)

Table 1. Fisher linear discriminant function coefficients for translationally important characteristics of mRNAs

mRNA characteristics	Fisher coefficient
A/G ⁻³	-5.8×10^{-5}
G ⁺⁴	-1.4×10^{-4}
5'-UTR length	1.7×10^{-7}
IC [1, 20]	2.0×10^{-4}
IC [-15, 1]	4.4×10^{-4}
[C,G] in 5'-UTR	0.002
[uAUG]	0.075

Table 2. The accuracy of translation efficiency classification based on translationally important mRNA characteristics

mRNA translation efficiency	Percentage of classified mRNAs	
	correctly	incorrectly
High	0.828	0.172
Low	0.668	0.332
All mRNAs	0.748	0.252

предсказывающие переменные, регрессоры. На каждом последующем регрессионном шаге удаляли наименее значимую характеристику с максимальным *p*-value.

В результате окончательная функция регрессии для выборки Set 1 содержала 14 наиболее значимых характеристик. Коэффициент корреляции между наблюдаемыми и предсказанными значениями достиг максимального значения 0.377. Для предсказания значений из выборки Set 2 было отобрано 19 характеристик мРНК, для Set 3 – 17. Коэффициенты корреляции между наблюдаемыми и предсказанными значениями эффективности трансляции достигли максимальных значений – 0.382 и 0.305 для Set 2 и Set 3 соответственно.

Использование функции линейной регрессии $y = \log(R1)$, где *R1* – количество нормированных ридов, *grkm*, после обработки харрингтонином, позволяет на основании экспериментальных данных по эффективности трансляции RiboSeq и по теоретически выявленным статистически значимым переменным предсказывать эффективность трансляции мРНК. На базе платформы BioUML мы разработали набор инструментов, позволяющий предсказывать эффективность трансляции по данным нуклеотидной последовательности мРНК (<http://micro.biouml.org/bioumlweb/#de=analyses/Methods/Binding%20regions/Ribo-Seq%20and%20mRNA%20features>). Доступ к данному набору осуществляется по запросу к авторам статьи.

Анализ частот олигонуклеотидов в районах, прилежащих к стартовому кодону. При формировании набора трансляционно-значимых характеристик мРНК для регрессионного анализа был проведен предварительный

Table 3. Regression coefficients for mRNA characteristics: Set 1, Set 2, Set 3

mRNA characteristics, Set 1	Regression coefficient	Z-score	p-value*
Set 1			
[uAUG]	-18.7745	12.4538	4.8×10^{-15}
5'-UTR length	-0.0061	9.9547	6.4×10^{-15}
mRNA length	7.0×10^{-8}	10.6784	1.5×10^{-13}
AAAAAG [3, 103]	0.1036	6.1414	4.1×10^{-10}
[G+C] in 5' UTR	-0.6468	5.7385	5.2×10^{-10}
CCGCCA [-100, -1]	0.2425	6.0110	5.9×10^{-10}
IC [-15, 1]	-0.2041	4.2438	1.1×10^{-6}
CCCGCC [-100, -1]	-0.1341	4.1337	1.8×10^{-6}
A/G ⁻³	0.1320	4.0518	3.0×10^{-6}
[G+C] in mRNA	1.2758	5.3203	5.5×10^{-6}
3'-UTR length	4.0×10^{-6}	3.5822	1.7×10^{-5}
AAGCAG [3, 103]	0.0592	3.9310	5.0×10^{-5}
AAGAAA [3, 103]	0.1008	2.6775	0.0037
AAGAAG [3, 103]	0.0321	2.0705	0.0192
Set 2			
[G+C] in 5' UTR	1.5874	45.4915	6.7×10^{-18}
[G+C] in mRNA	-2.0886	43.6240	5.4×10^{-18}
mRNA length	6.2×10^{-7}	27.5748	8.8×10^{-17}
3'-UTR length	-5.6×10^{-6}	15.8597	9.1×10^{-17}
CGCCGC [-100, -1]	0.1456	11.1288	8.3×10^{-15}
G ⁺⁴	0.0491	7.0671	8.0×10^{-13}
CAAGAA [3, 103]	0.1017	6.7531	7.3×10^{-12}
AAAAAG [3, 103]	0.1036	6.1414	4.1×10^{-10}
A/G ⁻³	0.0429	5.0215	2.5×10^{-7}
GAAGAA [3, 103]	0.0659	4.7225	1.1×10^{-6}
AGAAGC [3, 103]	0.0664	4.6729	1.4×10^{-6}
CTCCGC [-100, -1]	0.0724	4.4143	5.0×10^{-6}
CCGCCA [-100, -1]	0.0717	3.9105	6.0×10^{-6}
AAGCAG [3, 103]	0.0592	3.9310	5.0×10^{-5}
5'-UTR length	6.3×10^{-7}	3.1167	9.0×10^{-5}
[uAUG]	-1.0117	-3.2702	0.0005
IC [-15, 1]	-0.0638	-2.9720	0.0015
AAGAAG [3, 103]	0.0491	2.0705	0.0142
AGAAGA [3, 103]	0.0253	1.8736	0.0310
Set 3			
mRNA length	-9.97×10^{-6}	24.3408	$< 10^{-20}$
[uAUG]	-8.9533	10.3665	$< 10^{-20}$
A/G ⁻³	0.1905	10.1875	$< 10^{-18}$
3'-UTR length	5.1×10^{-6}	8.1743	2.2×10^{-16}
CCGCCA [-100, -1]	0.1569	5.4900	2.0×10^{-9}
IC [-15, 1]	-0.1584	4.6798	1.5×10^{-7}
[G+C] in 5' UTR	-0.2258	4.3418	7.0×10^{-7}
AAGCAG [3, 103]	0.0903	3.3412	4.2×10^{-5}
[G+C] in 3' UTR	0.3403	3.2927	4.5×10^{-5}
CTCCGC [-100, -1]	0.0788	3.2086	6.0×10^{-5}

End of Table 3

mRNA characteristics, Set 1	Regression coefficient	Z-score	p-value*
CGCCGC [-100, -1]	0.0559	2.9725	0.0014
CAAGAA [3, 103]	0.0829	2.9264	0.0017
AAGAAC [3, 103]	0.0914	2.8492	0.0023
5'-UTR length	-7.0×10^{-8}	2.6223	0.0044
AAGAAG [3, 103]	0.0524	2.3658	0.0090
CCAAGA [3, 103]	0.0527	1.7994	0.0360
[G + C] in mRNA	-0.4714	3.2485	0.5825

* Regressors are arranged in the descending order of statistical significance, p-value.

Table 4. Hexanucleotides tending to occur in [-100, -1] and [+3, +103] mRNA regions characterized by high efficiency of translation in samples Set1, Set2, and Set3; $F_{high}/F_{low} > 1$

Hexanucleotide	Set1			Set2			Set3		
	F_{low}	F_{high}	F_{high}/F_{low}	F_{low}	F_{high}	F_{high}/F_{low}	F_{low}	F_{high}	F_{high}/F_{low}
[-100, -1] 5'-UTR									
GCCGCC	0.070	0.187	2.671	0.139	0.253	1.820	0.063	0.169	2.68
CGCCGC	0.080	0.161	2.013	0.129	0.240	1.860	0.051	0.148	2.90
CCGCCA	0.046	0.126	2.739	0.040	0.177	4.425	0.037	0.105	2.838
CCGCCG	0.063	0.141	2.238	0.127	0.224	1.764	0.050	0.136	2.720
CCGCCG	0.048	0.110	2.292	0.110	0.148	1.345	0.048	0.110	2.292
CCCGCC	0.083	0.123	1.482	0.117	0.190	1.624	0.067	0.119	1.776
CTCCGC	0.048	0.084	1.750	0.057	0.118	2.070	0.042	0.079	1.881
CGCGCC	0.055	0.094	1.709	0.092	0.150	1.630	0.046	0.098	2.130
[+3, +103] CDS									
AAGAAA	0.031	0.060	1.935	0.046	0.059	1.283	0.027	0.061	2.259
AAAAAG	0.060	0.097	1.617	0.066	0.095	1.439	0.058	0.091	1.569
AAGAAC	0.037	0.054	1.459	0.024	0.071	2.958	0.031	0.050	1.613
AAGAAG	0.071	0.134	1.887	0.090	0.161	1.789	0.073	0.134	1.836
AAGCAG	0.053	0.068	1.283	0.054	0.100	1.852	0.054	0.072	1.333
AGAAGA	0.076	0.105	1.382	0.095	0.127	1.337	0.081	0.115	1.420
AGAAGC	0.056	0.095	1.696	0.082	0.102	1.244	0.056	0.090	1.607
CAAGAA	0.050	0.097	1.940	0.053	0.130	2.453	0.053	0.100	1.887

анализ частот гексануклеотидов в районах, прилежащих к стартовому кодону.

мРНК выборок Set1, Set2 и Set3 были ранжированы по значению нормированного количества ридов (rpkm) и разбиты на группы с высокой и низкой эффективностью трансляции. Был проведен сравнительный анализ частот гексануклеотидов в районах [-100, -1] 5'-НТП и [3, 100] БКП. Для дальнейшего анализа были отобраны гексануклеотиды, частоты которых значимо выше в последовательностях мРНК с высокой эффективностью трансляции (F_{high}) относительно аналогичных частот в последовательностях мРНК, с низкой эффективностью трансляции (F_{low}) (табл. 4). Данные гексануклеотиды, таким образом, являются характеристиками, дискрими-

нирующими мРНК с высокой и низкой эффективностью трансляции. Затем анализ проводился только для гексануклеотидов, чаще встречающихся в последовательностях мРНК, характеризующихся высокой эффективностью трансляции, т. е. $F_{high}/F_{low} > 1$.

Обсуждение

Предсказание эффективности трансляции на основании трансляционно-значимых характеристик с помощью дискриминантного анализа. Проведен предварительный дискриминантный анализ с характеристиками мРНК: [C+G] в 5' НТП, длина 5' НТП, A/G⁻³, IC [-15, 1], IC [1, 20], [uAUG], G⁺⁴, описанными в литературе как трансляционно-значимые. С использованием функции

Фишера (см. табл. 1) нам удалось с точностью 74.8 % классифицировать 2000 мРНК из выборки Set 1 как характеризующиеся высокой и низкой эффективностью трансляции (см. табл. 2).

Далее с помощью регрессионного анализа мы решили выявить наиболее трансляционно-значимые характеристики мРНК, используя экспериментальные данные рибосомного профилирования как параметры эффективности трансляции.

На основании результатов проведенного регрессионного анализа мы сделали ряд выводов.

1. Полученные с помощью регрессионного анализа результаты не однозначны для трех проанализированных выборок. Тенденции влияния отдельных трансляционно-значимых параметров разнонаправлены. Возможно, это связано с погрешностями эксперимента, различными условиями и особенностями регуляции трансляции для разных типов клеток.

2. Тем не менее можно выделить трансляционно-значимые параметры мРНК, тенденция влияния на эффективность трансляции которых наиболее значима и одинакова для всех трех выборок:

A/G⁻³, [uAUG], IC [-15, 1] снижают эффективность трансляции;

CCGCCA [-100, -1], AAGAAA [3, 103], AAGAAG [3, 103], AAGCAG [3, 103], AAAAAG [3, 103] усиливают эффективность трансляции.

Обнаруженные закономерности хорошо укладываются в общую схему инициации трансляции в рамках модели «линейного сканирования» (Kozak, 2005) и согласуются с полученными теоретически данными (Volkova, Kochetov, 2010).

3. Регрессионный анализ позволяет выявлять наиболее трансляционно-значимые параметры мРНК и с использованием их предсказывать эффективность трансляции.

4. Данные рибосомного профилирования позволяют выявлять трансляционно-значимые параметры мРНК и использовать полученные данные для предсказания эффективности трансляции мРНК и при клонировании мРНК эукариот методами генной инженерии.

Следует отметить, что гексануклеотиды, отобранные как значимо коррелирующие с эффективностью трансляции на левых флангах стартовых кодонов, существенно отличаются по своему нуклеотидному составу от олигонуклеотидов, отобранных на правых флангах мРНК, характеризующихся высокой эффективностью трансляции, и часто содержат в 5'-НТП GC-богатые гексануклеотиды с гуанином в каждой третьей позиции, в БКП – A-, G- и C-богатые (U избегается) гексануклеотиды. Отметим, что гексануклеотиды, часто расположенные в 5'-НТП мРНК, характеризующихся высокой эффективностью трансляции, имеют сходство с оптимальным контекстом млекопитающих GCCRCCAUGG (Kozak, 1986, 1987a, b; De Angioletti et al., 2004). Показано, что олигонуклеотид GCCGCCGCC (с G в каждой третьей позиции) усиливает эффективность трансляции на AUG и неAUG кодонов в отсутствии G в +4 позиции (Kozak, 1987a, 1989). По-видимому, высокая частота встречаемости данных гексануклеотидов в 5'-НТП мРНК, характеризующихся высокой эффективностью трансляции, связана с их спо-

собностью усиливать эффективность инициации трансляции как стартовых кодонов.

Отметим, что олигонуклеотиды БКП не позиционированы относительно рамки считывания. Представленные комбинации нуклеотидов приводят к образованию триплетов, кодирующих аминокислоты Ala, Asn, Arg, Ser, Gln, Glu и Pro. Ранее было показано, что данные аминокислоты перепредставлены во 2–4-й позициях белков, и считается, что они способствуют образованию первой пептидной связи (Volkova, Kochetov, 2010). Уридин не встречается в комбинации нуклеотидов в мРНК, характеризующихся высокой эффективностью трансляции. Это позволяет избегать стартового и стоп-кодонов, а также уридина в +4-позиции, который, как показано, уменьшает эффективность трансляции (Kozak, 1997).

Acknowledgments

This work was supported by the Russian Foundation for Basic Research, project 4-04-01284

The authors are grateful to the engineer of the Institute of Computational Technologies SB RAS Ivan Evshin for assistance in data processing.

Conflict of interest

The authors declare no conflict of interest.

References

- Brar G.A., Yassour M., Friedman N., Regev A., Ingolia N.T., Weissman J.S. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*. 2012;335:552-557. DOI 10.1126/science.1215110.
- De Angioletti M., Lacerra G., Sabato V., Carestia C. Beta+45 G--> C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence. *Br. J. Haematol.* 2004;124(2):224-231. DOI 10.1046/j.1365-2141.2003.04754.x.
- Dimelow R.J., Wilkinson S.J. Control of translation initiation: a model-based analysis from limited experimental data. *J. R. Soc. Interface*. 2009;6:51-61. DOI 10.1098/rsif.2008.0221.
- Ding Y., Shah P., Plotkin J.B. Weak 5'-mRNA secondary structures in short eukaryotic genes. *Genome Biol. Evol.* 2012;4(10):1046-1053. DOI 10.1093/gbe/evs082.
- Gerashchenko M., Gladyshev V. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucl. Acids Res.* 2014; 42(17):e134. DOI 10.1093/nar/gku671.
- Gerashchenko M.V., Lobanov A.V., Gladyshev V.N. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci. USA*. 2012;109:17394-17399. DOI 10.1073/pnas.1120799109.
- Godefroy-Colburn T., Thach R.E. The role of mRNA competition in regulating translation. IV. Kinetic model. *J. Biol. Chem.* 1981;256: 11762-11773.
- Grünert S., Jackson R.J. The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. *EMBO J.* 1994;13(15):3618-3630.
- Harkins S., Cornell C.T., Whitton J.L. Analysis of translational initiation in coxsackievirus B3 suggests an alternative explanation for the high frequency of R+4 in the eukaryotic consensus motif. *J. Virol.* 2005;79(2):987-996. DOI 10.1128/JVI.79.2.987-996.2005.
- Heiman M., Schaefer A., Gong S., Peterson J.D., Day M., Ramsey K.E., Suárez-Fariñas M., Schwarz C., Stephan D.A., Surmeier D.J., Greengard P., Heintz N. A translational profiling approach for the molecular characterization of CNS cell types. *Cell*. 2008;135(4):738-748. DOI 10.1016/j.cell.2008.10.028.
- Heinrich R., Rapoport T.A. Mathematical modelling of translation of mRNA in eukaryotes; steady-states, time dependent processes and

- application to reticulocytes. *J. Theor. Biol.* 1980;86:279-313. DOI 10.5936/csbj.201204002.
- Hinnebusch A.G., Lorsch J.R. The Mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harb. Perspect. Biol.* 2012;4:a011544. DOI 10.1101/cshperspect.a011544.
- Ingolia N.T., Brar G.A., Rouskin S., McGeachy A.M., Weissman J.S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 2012;7(8):1534-1550. DOI 10.1038/nprot.2012.086.
- Ingolia N.T., Lareau L.F., Weissman J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011;147(4):789-802. DOI 10.1016/j.cell.2011.10.002.
- Johannes G., Carter M.S., Eisen M.B., Brown P.O., Sarnow P. Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc. Natl. Acad. Sci. USA*. 1999;96(23):13118-13123. DOI 10.1073/pnas.96.23.13118.
- Kochetov A.V. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *BioEssays*. 2008;30(7):683-691. DOI 10.1002/bies.20771.
- Kochetov A.V., Palyanov A., Titov I.I., Grigorovich D., Sarai A., Kolchanov N.A. AUG_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics*. 2007;8(1):318. DOI 10.1186/1471-2105-8-318.
- Kozak M. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 1981;9(20):5233-5252. DOI 10.1093/nar/9.20.5233.
- Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. 1986;44:283-292. DOI 10.1038/308241a0.
- Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 1987a;15(20):8125-8148. DOI 10.1093/nar/15.20.8125.
- Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* 1987b;196:947-950. DOI 10.1016/0022-2836(87)90418-9.
- Kozak M. Context effects and inefficient initiation at non-AUG codons in eukaryotic cell-free translation systems. *Mol. Cell. Biol.* 1989;9(11):5073-5080. DOI 10.1128/MCB.9.11.5073.
- Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J.* 1997;16(9):2482-2492. DOI 10.1093/emboj/16.9.2482.
- Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*. 2005;361:13-37. DOI 10.1016/j.gene.2005.06.037.
- Lee S., Liu B., Lee S., Huang S.X., Shen B., Qian S.B. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. USA*. 2012;109(37):E2424-E2432. DOI 10.1073/pnas.1207846109.
- Li F., Zheng Q., Vandivier L.E., Willmann M.R., Chen Y., Gregory B.D. Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome. *Plant Cell*. 2012;24:4346-4359. DOI 10.1105/tpc.112.104232.
- Liu W., Zhao Y., Cui P., Lin Q., Ding F., Xin C., Tan X., Song S., Yu J., Hu S. Thousands of novel transcripts identified in mouse cerebrum, testis, and ES cells based on ribo-minus RNA sequencing. *Front. Genet.* 2011;2:93. DOI 10.3389/fgene.2011.00093.
- Michel A.M., Choudhury K.R., Firth A.E., Ingolia N.T., Atkins J.F., Baranov P.V. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.* 2012;22:2219-2229. DOI 10.1101/gr.133249.111.
- Na D., Lee S., Lee D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Syst. Biol.* 2010;4:71. DOI 10.1186/1752-0509-4-71.
- Nakagawa S., Niimura Y., Gojobori T., Tanaka H., Miura K. Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Res.* 2008;36(3):861-871. DOI 10.1093/nar/gkm1102.
- Niimura Y., Terabe M., Gojobori T., Miura K. Comparative analysis of the base biases at the gene terminal portions in seven eukaryote genomes. *Nucleic Acids Res.* 2003;31(17):5195-5201. DOI 10.1093/nar/gkg701.
- Pestova T.V., Kolupaeva V.G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* 2002;16:2906-2922. DOI 10.1101/gad.1020902.
- Pisarev A.V., Kolupaeva V.G., Pisareva V.P., Merrick W.C., Hellen C.U.T., Pestova T.V. Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* 2006;20:624-636. DOI 10.1101/gad.1397906.
- Reid D.W., Nicchitta C.V. Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *J. Biol. Chem.* 2012;287:5518-5527. DOI 10.1074/jbc.M111.312280.
- Rogozin I.B., Kochetov A.V., Kondrashov F.A., Koonin E.V., Milanesi L. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics*. 2001;17(10):890-900. DOI 10.1093/bioinformatics/17.10.890.
- Sanz E., Yang L., Su T., Morris D.R., McKnight G.S., Amieux P.S. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl. Acad. Sci. USA*. 2009;106(33):13939-13944. DOI 10.1073/pnas.0907143106.
- Schwanhäusser B., Busse D., Li N., Dittmar G., Schuchhardt J., Wolf J., Chen W., Selbach M. Global quantification of mammalian gene expression control. *Nature*. 2011;473(7347):337-342. DOI 10.1038/nature10098.
- Simon Z., Ruckenstein E. Regulation and synthesis processes in the living cell II. Kinetics of protein synthesis. *J. Theor. Biol.* 1966;11:299-313. DOI 10.1016/0022-5193(66)90167-6.
- Singh C.R., Watanabe R., Chowdhury W., Hiraishi H., Murai M.J., Yamamoto Y., Miles D., Ikeda Y., Asano M., Asano K. Sequential eukaryotic translation initiation factor 5 (eif5) binding to the charged disordered segments of eif4g and eif2β stabilizes the 48S preinitiation complex and promotes its shift to the initiation mode. *Mol. Cell Biol.* 2012;32(19):3978-3989. DOI 10.1128/MCB.00376-12.
- Siwiak M., Zielenkiewicz P.A. A comprehensive, quantitative, and genome-wide model of translation. *PLoS Comp. Biol.* 2010;6(7):e1000865. DOI 10.1371/journal.pcbi.1000865.
- Takacs J.E., Neary T.B., Ingolia N.T., Saini A.K., Martin-Marcos P., Pelletier J., Hinnebusch A.G., Lorsch J.R. Identification of compounds that decrease the fidelity of start codon recognition by the eukaryotic translational machinery. *RNA*. 2011;17:439-452. DOI 10.1261/rna.2475211.
- Volkova O.A., Kochetov A.V. Interrelations between the nucleotide context of human start AUG codon, N-end amino acids of the encoded protein and initiation of translation. *J. Biomol. Struct. Dyn.* 2010;27(5):611-618. DOI 10.1080/07391102.2010.10508575.
- Weiss R.B., Atkins J.F. Translation goes global. *Science*. 2011;334(6062):1509-1510. DOI 10.1126/science.1216974.
- Xia X., Huang H., Carullo M., Betran E., Moriyama E.N. Conflict between translation initiation and elongation in vertebrate mitochondrial genomes. *PLoS ONE*. 2007;2:e227. DOI 10.1371/journal.pone.0000227.
- Zhu L.H., Xu J.X., Zhu S.W., Cai X., Yang S.F., Chen X.L., Guo Q. Gene expression profiling analysis reveals weaning-induced cell cycle arrest and apoptosis in the small intestine of pigs. *J. Anim. Sci.* 2014;92:996-1006. DOI 10.2527/jas.2013-7551.
- Zong Q., Schummer M., Hood L., Morris D.R. Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. USA*. 1999;96(19):10632-10636. DOI 10.1073/pnas.96.19.10632.
- Zouridis H., Hatzimanikatis V. A model for protein translation: polyzome self-organization leads to maximum protein synthesis rates. *Biophys. J.* 2007;92:717-730. DOI 10.1529/biophysj.106.087825.



Кандидатные SNP-маркеры социального доминирования, способные влиять на сродство ТАТА-связывающего белка к промоторам генов человека

И.В. Чадаева¹, Д.А. Рассказов¹, Е.Б. Шарыпова¹, Л.К. Савинкова¹, П.М. Пономаренко², М.П. Пономаренко^{1,3}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Детский Госпиталь Лос-Анжелеса, Университет Южной Калифорнии, США

³ Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Предложена эвристическая гипотеза, согласно которой, если избыток какого-либо белка в ряде органов животных был экспериментально установлен как физиологический маркер повышенной агрессивности и если некоторый полиморфизм (SNP) человека вызывает суперэкспрессию гена, гомологичного гену этого белка у животных, то этот полиморфизм может быть кандидатным SNP-маркером предрасположенности к социальному доминированию, тогда как случаю дефицитной экспрессии может соответствовать социальное подчинение и наоборот. В рамках этой гипотезы проанализировали 21 ген человека: *ADORA2A*, *BDNF*, *CC2D1A*, *CC2D1B*, *ESR2*, *FEV*, *FOS*, *GH1*, *GLTSCR2*, *GRIN1*, *HTR1B*, *HTR1A*, *HTR2A*, *HTR2C*, *LGI4*, *LEP*, *MAOA*, *SLC17A7*, *SLC6A3*, *SNCA*, *TH*, – которые представляют функции белков, известных как физиологические маркеры агрессивного поведения животных: гормоны и их рецепторы, ферменты биосинтеза и рецепторы нейромедиаторов, транскрипционные и нейротропные факторы. Эти белки могут играть важную роль при установлении иерархических отношений у социальных видов животных. С использованием созданного нами ранее Web-сервиса SNP_TATA_Comparator (<http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl>) мы проанализировали 381 SNP в районе [–70; –20] перед стартами белок-кодирующих транскриптов (район связывания ТАТА-связывающего белка, TBP) из базы данных dbSNP (выпуск № 147). Было найдено 45 и 47 кандидатных SNP-маркеров доминирования и подчинения соответственно (например, rs373600960 и rs747572588). В рамках предложенной эвристической гипотезы и выпуска № 147 базы данных dbSNP мы получили статистически достоверные ($\alpha < 10^{-5}$) свидетельства о действии естественного отбора как против дефицитной экспрессии генов, способных влиять на предрасположенность к доминированию, так и в пользу того, что подчинение и доминирование могут характеризовать норму реакции агрессивности (отличие незначимо: $\alpha > 0.35$). Предложенная гипотеза, выявленные на ее основе кандидатные SNP-маркеры и закономерности их влияния естественного отбора на геном человека обсуждаются в контексте литературных данных: могут ли они иметь какое-либо отношение к социальному доминированию у людей. Сделано заключение – эти результаты нуждаются в экспериментальной проверке.

Ключевые слова: ТАТА-связывающий белок (TBP); сайт TBP-связывания (ТАТА-бок); промотор; сродство TBP/промотор; ген; нуклеотидный полиморфизм (SNP); изменение экспрессии гена; достоверность; SNP-маркер; доминант; субординант.

Candidate SNP markers of social dominance, which may affect the affinity of the TATA-binding protein for human gene promoters

I.V. Chadaeva¹, D.A. Rasskazov¹, E.B. Sharypova¹, L.K. Savinkova¹, P.M. Ponomarenko², M.P. Ponomarenko^{1,3}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Children's Hospital Los Angeles, University of Southern California, USA

³ Novosibirsk State University, Novosibirsk, Russia

The following heuristic hypothesis has been proposed: if an excess of a protein in several animal organs was experimentally identified as physiological marker of increased aggressiveness and if a polymorphism (SNP) can cause superexpression of the human gene homologous of the animal gene encoding this protein, then this polymorphism can be a candidate SNP marker of social dominance, whereas a deficient expression corresponds to subordinate and vice versa. Within this hypothesis, we analyzed 21 human genes – *ADORA2A*, *BDNF*, *CC2D1A*, *CC2D1B*, *ESR2*, *FEV*, *FOS*, *GH1*, *GLTSCR2*, *GRIN1*, *HTR1B*, *HTR1A*, *HTR2A*, *HTR2C*, *LGI4*, *LEP*, *MAOA*, *SLC17A7*, *SLC6A3*, *SNCA*, *TH* – which represent the functions of proteins known as physiological markers of aggressive behavior in animals: hormones and their receptors, biosynthetic enzymes and receptors of neurotransmitters, transcription and neurotrophic factors. These proteins may play an important role in determining hierarchical relationships in social animals. Using our previously developed Web-service SNP_TATA_Comparator (<http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl>), we analyzed 381 SNPs within the region of [–70; –20] relative to the start protein-coding transcripts, which is the region of the all known TATA-binding protein (TBP) binding sites. We took them from the database dbSNP, v.147. As a result, we found 45 and 47 candidate SNP markers of dominance and submission, respectively (e.g., rs373600960 and rs747572588). Within the framework of the proposed heuristic hypotheses and database dbSNP v.147, we found statistically significant ($\alpha < 10^{-5}$)

evidence of the effects of natural selection against the deficient expression of genes, which can affect the predisposition to dominate, as well as in favor of both subordination and domination behavior as a norm of reaction of aggressiveness (difference not significant: $\alpha > 0.35$). The proposed hypothesis, the candidate SNP markers predicted and the observed regularities of effects of natural selection for the human genome are discussed in comparison with published data: whether they can have any relation to social dominance in human. It was concluded that these results require experimental verification.

Key words: TATA-binding protein (TBP); TBP-binding site (TATA box); promoter; TBP-promoter affinity; gene; single nucleotide polymorphism (SNP); gene expression change; significance; SNP marker; dominant; subordinate.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Чадаева И.В., Рассказов Д.А., Шарыпова Е.Б., Савинкова Л.К., Пономаренко П.М., Пономаренко М.П. Кандидатные SNP-маркеры социального доминирования, способные влиять на сродство TATA-связывающего белка к промоторам генов человека. Вавиловский журнал генетики и селекции. 2016;20(6):787-796. DOI 10.18699/VJ16.196

HOW TO CITE THIS ARTICLE:

Chadaeva I.V., Rasskazov D.A., Sharypova E.B., Savinkova L.K., Ponomarenko P.M., Ponomarenko M.P. Candidate SNP markers of social dominance, which may affect the affinity of the TATA-binding protein for human gene promoters. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):787-796. DOI 10.18699/VJ16.196

Социальным доминированием называется иерархическая структура взаимоотношений между особями в популяции, которая является необходимым условием их сосуществования, способствующим уменьшению агрессии среди них, упорядочиванию конкуренции за ограниченные ресурсы среды и поддержанию высокого репродуктивного потенциала (Rowell, 1974; Hinde, 1970).

Иерархические отношения между особями одного вида устанавливаются путем агрессии, т.е. агонистического поведения. Оно определяется этологами как врожденная форма поведения по защите самого себя, потомства, территории, жилища (Lorenz, 2002). Эксперименты по одомашниванию (Belyaev, 1979) и искусственной селекции животных на агрессивность (Kulikov et al., 2016) доказали вклад генетических факторов в фенотипические проявления агрессивности (Ehrman, Parsons, 1981; Moore, 2013). Уже известны работы полногеномного поиска генетических факторов, проводимые на модельных животных (Zapata et al., 2016). Однако референсный геном человека (Colonna et al., 2014) и его однонуклеотидный полиморфизм (single nucleotide polymorphism, SNP) – результат не искусственного, а естественного отбора. Тем не менее у социальных видов (в том числе у человека) границы допустимой агрессии между особями возникают при установлении иерархических отношений «доминант–субординант». Эти границы контролируются естественным отбором как норма реакции агрессивного поведения (Eldakar, Gallup, 2011), а ранг особи в социальной иерархии влияет на условия, качество и продолжительность ее жизни (Michoropoulos et al., 2012). В микропопуляциях лабораторной мыши были описаны репродуктивные корреляты социальной иерархии (Осадчук и др., 2007), а также обнаружено наследование предрасположенности к социальному доминированию (Серова и др., 1991). Это свидетельствует о возможности вклада агрессии между особями одного вида в приспособленность. Предвестником поиска генетических факторов агрессии у человека

стало обнаружение биомедицинских SNP-маркеров – отличий между референсным геномом человека и геномом индивида с определенным фенотипом, – для агрессивности как осложнения при терапии антипсихотическими препаратами (например, оланзапином (Ellingrod et al., 2005)).

Если в «догеномную эру» открытие корреляции между SNP и фенотипическим признаком (например, патологией) было редкой случайной удачей, то в настоящее время это одна из главных целей крупнейшего научного проекта современности «1000 геномов» (Colonna et al., 2014). Официальным источником найденных и тщательно верифицированных в рамках «1000 геномов» SNP является база данных dbSNP (Sherry et al., 2001). На ее основе корректируют референсный геном человека (Colonna et al., 2014), который собирается из самых частых (анцестральных) нуклеотидов в каждой позиции ДНК. Общепринятый подход к поиску SNP-маркеров – оценка достоверности различия частоты встречаемости целевого SNP в представительных когортах особей с целевым фенотипическим признаком (например, с импульсивностью поведения как формой агрессивности) и без этого признака, что является весьма трудоемкой и дорогостоящей процедурой (Abbas et al., 2006). Предварительный компьютерный анализ известных SNP может удешевить и сделать прицельным поиск SNP-маркеров путем выбраковки нейтральных SNP, которых абсолютное большинство (по теории нейтральной эволюции (Kimura, 1968)). Точность биоинформатического подхода достигает своего максимума для SNP-маркеров белок-кодирующих районов генов в силу инвариантности дефектов белков, а также минимума для SNP-маркеров регуляторных районов генов из-за отсутствия дефектных белков (Amberger et al., 2015). Большинство известных биомедицинских регуляторных SNP-маркеров ($\approx 10\%$) занимают узкий район $[-70; -20]$ относительно стартов транскрипции генов и изменяют связывание TATA-связывающего белка с TATA-боксом (Ponomarenko et al.,

2015), необходимым для запуска транскрипции генов (Martianov et al., 2002). Было экспериментально показано, что родство TBP/TATA-бокс достоверно позитивно коррелирует с экспрессией соответствующих генов (Mogno et al., 2010).

Мы фокусировали наши предыдущие исследования на районе $[-70; -20]$ перед стартами белок-кодирующих транскриптов в геноме человека, куда проецируются известные сайты связывания TBP, самые изученные регуляторные сигналы генома (Ponomarenko et al., 2013). Была создана вычислительная модель трехшагового связывания TBP/TATA-бокс (Ponomarenko et al., 2008): скольжение TBP вдоль ДНК $\rightarrow$ остановка TBP на TATA-боксе $\rightarrow$ стабилизация комплекса TBP/TATA изгибом спирали ДНК, как это годом позже наблюдали в опыте R.F. Delgadillo с коллегами (2009). На основе этой модели мы создали Web-сервис SNP_TATA_Comparator (<http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl>) (Ponomarenko et al., 2015) для оценки значимости изменения родства TBP к минорному аллелю промотора относительно нормы (анцестральный аллель) и применили его к предсказанию кандидатных SNP-маркеров для ожирения (Arkova et al., 2015), хронопатологий (Ponomarenko P. et al., 2016), аутоиммунных патологий (Ponomarenko M. et al., 2016) и агрессивности как осложнения при ряде заболеваний (Chadaeva et al., 2016).

Цель этой работы – прогноз на основе референсного генома человека кандидатных SNP-маркеров предрасположенности к социальному доминированию и ее отсутствию (что обуславливает субординантный (подчиненный) ранг особи). Для ее достижения мы предложили гипотезу о возможности такого прогноза на основе учета физиологических маркеров агрессивности животных. В рамках этой гипотезы мы применили Web-сервис (Ponomarenko et al., 2015) к анализу генов человека, гомологичных генам животных для широкого круга функций белков-маркеров, связанных с агрессивным поведением: гормоны, ферменты биосинтеза и рецепторы нейромедиаторов, транскрипционные и нейротропные факторы. В результате нашли 92 маркера предрасположенности к доминированию и подчинению у людей, которые могут быть актуальной информацией при планировании экспериментально-компьютерных исследований социального поведения человека.

Материалы и методы

Базовая гипотеза. Поскольку геном человека как биологического вида закреплен в результате действия естественного, а не искусственного отбора, в частности по агрессивному поведению, как в случае экспериментов на лабораторных животных (см., например, (Kulikov et al., 2016)), то для его анализа мы предлагаем эвристическую гипотезу: если избыток какого-либо белка в ряде органов животных был экспериментально установлен как физиологический маркер повышенной агрессивности и если некоторый SNP человека вызывает суперэкспрессию гена, гомологичного гену этого белка у животных, то этот полиморфизм может быть кандидатным SNP-маркером предрасположенности к социальному доминированию, тогда как случаю дефицитной экспрессии может соответствовать поведение подчинения и наоборот.

Исследуемые гены человека. Для исследования предложенной гипотезы в рамках референсного генома человека (Colonna et al., 2014) мы сформировали выборку из 21 гена человека (*ADORA2A*, *BDNF*, *CC2D1A*, *CC2D1B*, *ESR2*, *FEV*, *FOS*, *GH1*, *GLTSCR2*, *GRIN1*, *HTR1B*, *HTR1A*, *HTR2A*, *HTR2C*, *LG14*, *LEP*, *MAOA*, *SLC17A7*, *SLC6A3*, *SNCA*, *TH*), представляющую различные функции белков, которые экспериментально известны в качестве маркеров агрессивного поведения животных: гормоны и их рецепторы, фермент биосинтеза и рецепторы нейромедиаторов, транскрипционные и нейротропные факторы (таблица).

Компьютерный анализ SNP-последовательностей ДНК промоторов генов. Для анализируемых генов человека мы взяли из базы данных Ensembl (Zerbino et al., 2015), (http://www.ensembl.org/Homo_sapiens/) номера всех белок-кодирующих транскриптов и с помощью Web-сервиса UCSC Genome Browser (Haeussler et al., 2015), (<http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38>) – номера всех SNP в районе $[-70; -20]$ относительно стартов каждого из этих транскриптов. Отобранные таким путем SNP были проанализированы с помощью нашего Web-сервиса (Ponomarenko et al., 2015), как показано на рисунке для случая SNP rs3813929 в промоторе к транскрипту 001 гена *HTR2C* человека. В окне Base sequence (см. рисунок) показана последовательность ДНК этого промотора, взятая из базы данных Ensembl (Zerbino et al., 2015) с помощью стандартной библиотеки BioPerl доступа к референсному геному человека (Colonna et al., 2014). Она была использована как анцестральный вариант SNP rs3813929 (норма). Минорный вариант этого SNP был сделан вручную по его описанию в базе данных dbSNP ((Sherry et al., 2001), <https://www.ncbi.nlm.nih.gov/snp/>), как показано в окне Editable sequence. С помощью этих последовательностей ДНК мы оценили, согласно работе М. Ponomarenko с коллегами (2015), величины 16.42 ± 0.07 и 16.70 ± 0.07 натуральных логарифмических единиц родства TBP к анцестральному и минорному аллелям данного промотора соответственно (окно Result), которые даны также в таблице как 73 и 56 nM. Различие между этими аллелями по родству TBP/промотор оценили Z-статистикой (отношение разности к корню суммы квадратов стандартных отклонений) и взяли уровень значимости из стандартного статистического пакета R, как показано на рисунке ($Z = 6$, $\alpha < 10^{-6}$). Это означает, что рассматриваемый SNP rs3813929 может значимо увеличить родство TBP к минорному варианту рассматриваемого промотора относительно нормы (анцестральный вариант) и, таким образом, вызвать суперэкспрессию гена *HTR2C*, согласно (Mogno et al., 2010). Этот прогноз Web-сервиса (Ponomarenko et al., 2015) в строке DECISION на рисунке согласуется с известными данными об SNP-маркере rs3813929 устойчивости к оланзапину (Landrum et al., 2014) вследствие избытка серотонинового рецептора 2C (Ellingrod et al., 2005), кодируемого этим геном.

Наконец, поскольку эксперимент (Porova et al., 2010) выявил избыток этого рецептора в гиппокампе и в лобной коре головного мозга в качестве физиологического маркера неагрессивных крыс в сравнении с агрессивными крысами, то в рамках предложенной нами гипотезы rs3813929 может быть кандидатным SNP-маркером для

Candidate SNP markers for social dominance predicted by the Web-service SNP_TATA_Comparator (Ponomarenko et al., 2015)

Human gene	dbSNP, rel. 147 (Sherry et al., 2001)	5'-flank	wt mut	3'-flank	K _D , nM				Features: known (see Reference) and/or predicted (this work)	Reference	
					wt mut	Δ	Z	α			
ADORA2A	rs199593618	tgcccgccat	$\frac{c}{g}$	tgagggaggt	$\frac{45}{32}$	↑	7	10 ⁻⁶	Subordination	Jegou et al., 2003	
	rs200817204	ggtgccaatc	$\frac{c}{a}$	ttcacgtgac	$\frac{45}{25}$	↑	13	10 ⁻⁶			
	rs539804305	caatccttca	$\frac{c}{t}$	gtgacgtgac	$\frac{45}{20}$	↑	15	10 ⁻⁶			
	rs201534466	aatccttcac	$\frac{g}{a}$	tgacgtgacg	$\frac{45}{21}$	↑	15	10 ⁻⁶			
	rs200174085	cttcacgtga	$\frac{c}{t}$	gtgacgtggc	$\frac{45}{32}$	↑	7	10 ⁻⁶			
	rs143744701	ccgtccgtca	$\frac{gtcc}{-}$	gtccgtccgt	$\frac{100}{47}$	↑	13	10 ⁻⁶			
	rs561125568	gccgcaggta	$\frac{g}{c}$	cgggcgggga	$\frac{47}{55}$	↓	2	0.05			Dominance
BDNF	rs549649514	cttaattaga	$\frac{g}{t}$	attttaagc	$\frac{9}{7}$	↑	3	0.01	Dominance	Bondar et al., 2009; Kudryavtseva et al., 2010; Ilchibaeva et al., 2015	
	rs774974933	agtatcactt	$\frac{a}{c}$	attagagatt	$\frac{9}{8}$	↑	2	0.05			
	rs554116130	atatcaagta	$\frac{t}{a}$	cacttaatta	$\frac{9}{6}$	↑	5	10 ⁻³			
	rs368334830	tctgtttaag	$\frac{c}{t}$	cggcgtgttt	$\frac{10}{76}$	↑	6	10 ⁻⁶			
	rs768585661	cccgggaaac	$\frac{g}{a}$	cacgccctct	$\frac{48}{39}$	↑	4	10 ⁻³			
	rs539103547	gggaagcttg	$\frac{c}{t}$	gccttcagcc	$\frac{142}{66}$	↑	13	10 ⁻⁶			
	rs781368365	aaacagcgag	$\frac{g}{t}$	ttagtcgtcg	$\frac{27}{14}$	↑	11	10 ⁻⁶			
CC2D1A	rs749187303	ctgcagagcc	$\frac{c}{t}$	taggtgggcc	$\frac{41}{32}$	↑	4	10 ⁻³	Subordination	Kondaurova et al., 2016	
	rs774355414	gggcctgctt	$\frac{g}{a}$	tcctccttac	$\frac{41}{21}$	↑	12	10 ⁻⁶			
	rs752806837	ggcctgcttg	$\frac{tcc}{-}$	tccttaccct	$\frac{41}{31}$	↑	5	10 ⁻³			
	rs761077398	tcctccttac	$\frac{c}{t}$	cccgccacca	$\frac{41}{31}$	↑	5	10 ⁻⁶			
	rs372837238	tccttaccct	$\frac{c}{t, (g)}$	gccaccagcc	$\frac{41}{34}$	↑	3	0.01			
	rs750548001	aggaggcagg	$\frac{g}{t}$	caaggccggg	$\frac{121}{81}$	↑	8	10 ⁻⁶			
CC2D1B	rs751009765	cgcagctagc	$\frac{c}{t}$	tctgtgagga	$\frac{39}{33}$	↑	3	0.01	Dominance	Kondaurova et al., 2016	
	rs756613102	gtcgcagcta	$\frac{g}{a}$	cctctgtgag	$\frac{39}{24}$	↑	8	10 ⁻⁶			
	rs752164206	cttgagctcg	$\frac{c}{t}$	agctagcctc	$\frac{39}{18}$	↑	12	10 ⁻⁶			
	rs143901936	ccttgagctc	$\frac{g}{a}$	cagctagcct	$\frac{39}{34}$	↑	2	0.05			
	rs769629145	tgggtttctg	$\frac{g}{t}$	accagttctc	$\frac{39}{9}$	↑	23	10 ⁻⁶			
	rs551992483	caggccctta	$\frac{g}{a}$	ctgaattgtt	$\frac{43}{17}$	↑	16	10 ⁻⁶			
	rs200669825	gcaggccctg	$\frac{g}{t, (a)}$	agtggccagg	$\frac{43}{22}$	↑	10	10 ⁻⁶	Subordination		
	rs549629102	aatggtaggc	$\frac{c}{t, g}$	tagcaggccc	$\frac{43}{36}$	↑	3	0.01			
	rs767975647	gcctctgtga	$\frac{g}{c}$	gagaggcaga	$\frac{39}{50}$	↓	4	10 ⁻³			
	rs764015054	gagtcgcagc	$\frac{tagc}{-}$	ctctgtgagg	$\frac{39}{50}$	↓	4	10 ⁻³			
	rs536919659	agtcgcagct	$\frac{a}{g}$	gcctctgtga	$\frac{39}{50}$	↓	4	10 ⁻³			
	rs375339315	agtgggggtca	$\frac{c}{a}$	gtgttcgcgcg	$\frac{38}{44}$	↓	2	0.05			

Continuation of Table

Human gene	dbSNP, rel. 147 (Sherry et al., 2001)	5'-flank	wt mut	3'-flank	K _D , nM				Features: known (see Reference) and/or predicted (this work)	Reference			
					wt mut	Δ	Z	α					
ESR2	rs35036378	cctctcggtc	$\frac{t}{g}$	ttaaaggaa	$\frac{6}{8}$	↓	5	10 ⁻³	Dominance	Cao et al., 2014; Rosenfeld, 2015			
	rs766797386	ttaaaggaa	$\frac{g}{t}$	aaggggctta	$\frac{6}{7}$	↓	3	10 ⁻²					
FOS	rs547882209	ctggcgccac	$\frac{c}{t}$	gtggttgagc	$\frac{61}{51}$	↑	3	10 ⁻³	Dominance	Savonenko et al., 1999; Lin et al., 2011			
	rs7101	cagctctgct	$\frac{c}{t}$	cacagcgccc	$\frac{62}{46}$	↑	6	10 ⁻⁶					
	rs76195634	tcattcccac	$\frac{g}{a}$	gtcactgcca	$\frac{60}{52}$	↑	3	0.05					
	rs747221475	tcactgccat	$\frac{c}{t}$	tcgaccagtc	$\frac{60}{32}$	↑	12	10 ⁻⁶					
	rs148786753	ctcgaccagt	$\frac{c}{a}$	cggacctgca	$\frac{60}{38}$	↑	8	10 ⁻⁶					
	rs752219516	caggtgagga	$\frac{a}{t}$	ctctagcgta	$\frac{26}{14}$	↑	10	10 ⁻⁶					
	rs757931341	gcgtactctt	$\frac{c}{t}$	ctgggaatgt	$\frac{26}{22}$	↑	3	0.01					
	rs763882586	tactcttcct	$\frac{g}{a}$	ggaatgtggg	$\frac{26}{18}$	↑	6	10 ⁻⁶	Subordination				
	rs180743738	agacgtcagg	$\frac{g}{a}$	atatttataa	$\frac{1.7}{2.1}$	↓	3	10 ⁻³					
	rs776646101	cgtcagggat	$\frac{a}{g}$	tttataacaa	$\frac{1.7}{3.0}$	↓	8	10 ⁻⁶					
	rs761656041	agggatattt	$\frac{a}{g}$	taacaaaccc	$\frac{1.7}{4.5}$	↓	14	10 ⁻⁶					
	rs767588227	aactctagcg	$\frac{t}{c}$	actcttcctg	$\frac{26}{50}$	↓	10	10 ⁻⁶					
	GH1	rs774326004	ccagggtata	$\frac{a}{t}$	aaagggccca	$\frac{1.5}{0.9}$	↑	7			10 ⁻⁶	Dominance	Ji, Findling, 2016
		rs777003420	aaggggccag	$\frac{g}{t}$	gtataaaaag	$\frac{1.5}{1.3}$	↑	3			0.05		
rs768454929		agggtataaa	$\frac{a}{c}$	agggccca	$\frac{1.5}{2.6}$	↓	7	10 ⁻⁶	Subordination				
rs761695685		gccagggtat	$\frac{a}{g}$	aaaagggccc	$\frac{1.5}{5.8}$	↓	19	10 ⁻⁶					
GLTSCR2	rs562807927	agtacgtcac	$\frac{g}{a}$	gaagagggtg	$\frac{16}{14}$	↑	2	0.05	Subordination	Heyne et al., 2014			
	rs779295677	acaaagaaga	$\frac{g}{a}$	aaccaaagtc	$\frac{34}{25}$	↑	6	10 ⁻⁶					
	rs79458115	agaagagaac	$\frac{c}{g}$	aaagtccaga	$\frac{34}{31}$	↑	2	0.05					
	rs776466959	gtcactgctt	$\frac{c}{t}$	tcaagaaacc	$\frac{34}{18}$	↑	12	10 ⁻⁶					
	rs548738286	gggagaagct	$\frac{g}{a}$	gccaaagcagg	$\frac{207}{168}$	↑	4	10 ⁻³					
	rs748396660	gagaagctgg	$\frac{c}{t}$	caagcagggc	$\frac{207}{92}$	↑	16	10 ⁻⁶					
	rs769839944	tggccaagca	$\frac{g}{c}$	ggcgagctgc	$\frac{207}{163}$	↑	5	10 ⁻³					
	rs773436143	ccaagcaggg	$\frac{c}{t}$	gagctgcccc	$\frac{207}{168}$	↑	4	10 ⁻³					
	rs541246368	gggcgagctg	$\frac{c}{a}$	cccgggaggt	$\frac{207}{152}$	↑	6	10 ⁻⁶					
	rs561564168	cgagctgccc	$\frac{c}{t}$	gggaggtgcg	$\frac{207}{182}$	↑	3	0.05					
	rs760081568	gagctgcccc	$\frac{g}{a}$	ggaggtgcgc	$\frac{207}{158}$	↑	5	10 ⁻⁶					
	rs755859075	ccgggaggtg	$\frac{c}{t}$	gcagggccca	$\frac{207}{76}$	↑	16	10 ⁻⁶					
	rs752084705	cgggaggtgc	$\frac{g}{a}$	cagggcccag	$\frac{207}{73}$	↑	18	10 ⁻⁶					
	rs780677965	ccggctgcgc	$\frac{g}{a}$	ggatcaaggc	$\frac{56}{44}$	↑	5	10 ⁻³					
	rs747572588	cggctgcgcg	$\frac{g}{a}$	gatcaaggcc	$\frac{56}{50}$	↑	2	0.05					

End of Table

Human gene	dbSNP, rel. 147 (Sherry et al., 2001)	5'-flank	wt mut	3'-flank	K _D , nM				Features: known (see Reference) and / or predicted (this work)	Reference
					wt mut	Δ	Z	α		
GLTSCR2	rs746705150	aagtcaactgc	$\frac{t}{c}$	tctcaagaaa	$\frac{34}{39}$	↓	2	0.05	Dominance	Heyne et al., 2014
	rs569945779	agtcactgct	$\frac{t}{c}$	ctcaagaaac	$\frac{34}{39}$	↓	3	0.01		
	rs767787052	ggccttgact	$\frac{t}{c}$	gttctctttc	$\frac{24}{52}$	↓	15	10 ⁻⁶		
	rs752987742	gccttgactt	$\frac{g}{c}$	ttctctttcc	$\frac{24}{39}$	↓	10	10 ⁻⁶		
GRIN1	rs372970942	ccgagggccc	$\frac{c}{t}$	gcgttcgcgc	$\frac{203}{153}$	↑	6	10 ⁻⁶	Dominance	Reif et al., 2009
	rs759827008	gagggccccc	$\frac{c}{a}$	gttcgcgcgc	$\frac{203}{147}$	↑	6	10 ⁻⁶		
	rs200150234	gccccgcgtt	$\frac{c}{a}$	gcgccgcgca	$\frac{203}{61}$	↑	19	10 ⁻⁶		
	rs540882401	cgcgttcgcg	$\frac{c}{a}$	cgcgcagagc	$\frac{203}{157}$	↑	5	10 ⁻⁶		
	rs753166193	gcgttcgcgc	$\frac{c}{t}$	gcgcagagcc	$\frac{203}{160}$	↑	5	10 ⁻³		
	rs201254060	caaagacagg	$\frac{g}{a}$	tggtgtggac	$\frac{23}{26}$	↓	2	0.05	Subordination	
HTR1B	rs75032335	tctgcagatc	$\frac{c}{t}$	aaaagcgctc	$\frac{37}{9}$	↑	24	10 ⁻⁶	Subordination	Popova, 2006
HTR2A	rs6316	tgacatacac	$\frac{a}{g}$	tagaggggagg	$\frac{5}{7}$	↓	4	10 ⁻³	Dominance	Popova, 2006
HTR2C	rs3813929 (рисунок)	cccctcatcc	$\frac{c}{t}$	gcttttggcc	$\frac{73}{56}$	↑	6	10 ⁻⁶	Resistance to olanzapine (clinical history) and subordination	Ellingrod et al., 2005; Landrum et al., 2014; Popova et al., 2010
LGI4	rs187763897	ccagcctctt	$\frac{c}{t}$	acctccaact	$\frac{34}{16}$	↑	11	10 ⁻⁶	Dominance	Heyne et al., 2014
	rs538317089	gatgggggct	$\frac{g}{a}$	caattctgga	$\frac{34}{17}$	↑	10	10 ⁻⁶		
LEP	rs200487063	tgatcgggcc	$\frac{g}{a}$	ctataagagg	$\frac{4}{2}$	↑	6	10 ⁻⁶	Dominance	Michopoulos et al., 2012
	rs34104384	ccgctataag	$\frac{a}{t}$	ggggcgggca	$\frac{4}{3}$	↑	4	10 ⁻²		
	rs201381696	tcgggccgct	$\frac{a}{g}$	taagaggggc	$\frac{4}{12}$	↓	17	10 ⁻⁶	Subordination	
SNCA	rs373600960	tgtattttat	$\frac{g}{a,(t)}$	ttttccagt	$\frac{6}{3}$	↑	12	10 ⁻⁶	Dominance	Bondar et al., 2009
	rs578214974	gggcagaagc	$\frac{g}{t}$	ctgacaaatc	$\frac{25}{23}$	↑	2	0.05		
	rs539338282	ctcacgcctt	$\frac{g}{a}$	ccttcaagcc	$\frac{47}{26}$	↑	13	10 ⁻⁶		
	rs748857490	cccgcgagtg	$\frac{t}{c}$	gagcggcgcc	$\frac{51}{26}$	↓	14	10 ⁻⁶	Subordination	
	rs562817680	cgcgagtgtg	$\frac{a}{g}$	gcggcgctg	$\frac{51}{114}$	↓	8	10 ⁻⁶		
	rs530304614	gtgcttaact	$\frac{a}{g}$	atcttccgaa	$\frac{7}{19}$	↓	13	10 ⁻⁶		
	rs777103921	agttctcatt	$\frac{11 \text{ n.o.}^{\$}}{-}$	ttatgttttc	$\frac{6}{10}$	↓	8	10 ⁻⁶		
	rs773252996	cattcaaagt	$\frac{g}{c}$	tattttatgt	$\frac{6}{9}$	↓	7	10 ⁻⁶		
TH	rs772061026	agctggacaa	$\frac{g}{t}$	tgatcatcacc	$\frac{22}{27}$	↓	4	10 ⁻³	Subordination	Bondar et al., 2009

Designations: **wt** and **mut** are the ancestral (reference human genome, release GRCh38(NCBI)/hg38(UCSC)) and minor (Parenthesized are neutral substitutions.) alleles, respectively; # the DNA strand of the gene promoter; \$ rs777103921, the 11-bp caaagtgtatt deletion; K_D, dissociation constant of the TBP/DNA complex *in vitro* (Savinkova et al., 2013); Δ, expression rate change: ↑, increase; ↓, decrease; Z – Z-score; α, significance (α = 1 – p, where p is the probability in the figure). Genes: *ADORA2A*, A2a adenosine receptor; *BDNF*, brain-derived neurotrophic factor; *CC2D1A* and *CC2D1B*, Freud-1 and Freud-2, respectively; *ESR2*, estrogen receptor β; *GH1*, somatotropin (growth hormone 1); *GLTSCR2*, Glioma tumor suppressor candidate region gene 2; *GRIN1*, glutamate ionotropic receptor subunit 1; *HTR1B*, *HTR2A*, and *HTR2C*, serotonin receptors 1B, 2A, and 2C, respectively; *LGI4*, leucine-rich glioma-inactivated gene 4; *LEP*, gene for leptin (satiety hormone); *SNCA*, α-synuclein; *TH*, tyrosine hydroxylase.

предрасположенности людей к подчинению (см. таблицу).

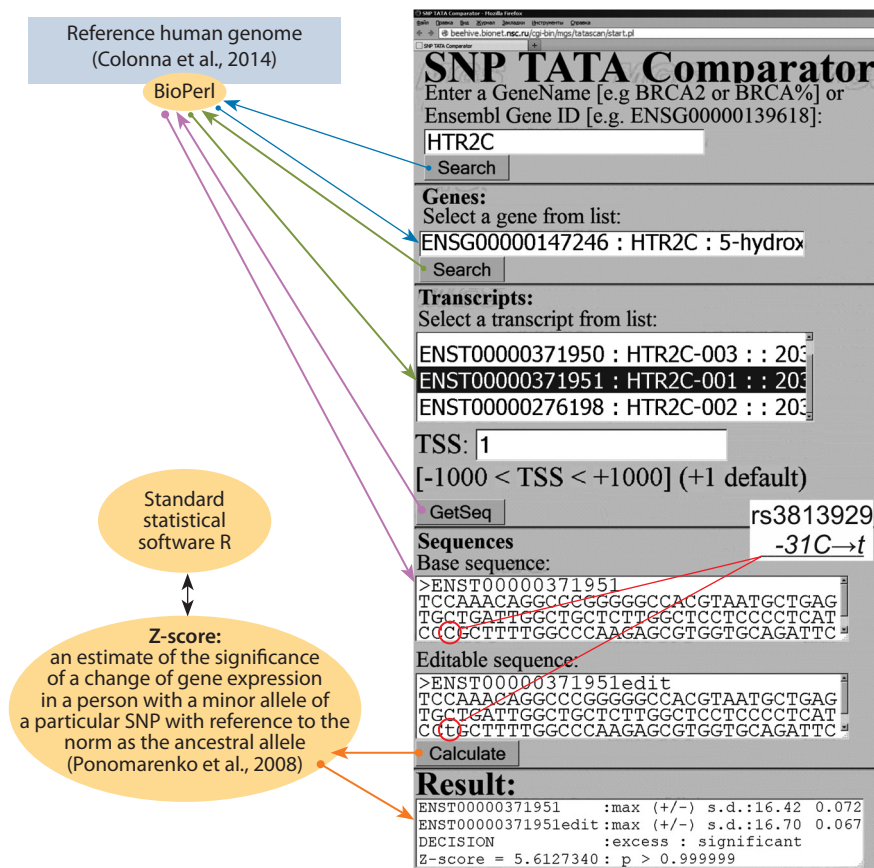
Более детальное описание метода можно найти в статье (Ponomarenko et al., 2015), его верификации – в наших экспериментах *in vivo* на культурах клеточных линий человека с использованием вектора pGL4.23 (Chadaeva et al., 2016), *in vitro* в реальном времени (Arkova et al., 2014), в неравновесных (Drachkova et al., 2014) и равновесных (Savinkova et al., 2013) условиях, а также на данных независимых экспериментов других авторов (см. обзоры (Ponomarenko et al., 2009, 2010; Suslov et al., 2010)).

Результаты

Проанализирован 381 случай SNP-промоторов 21 гена человека и выявлено 92 кандидатных SNP-маркера доминирования и подчинения, которые приведены в таблице, как описано выше на примере SNP rs3813929. Для генов *MAOA*, *HTR1A*, *SLC17A7*, *SLC6A3* и *FEV* мы не нашли таких SNP-маркеров (данные не показаны).

Прежде всего, неожиданным результатом стало выявление 45 и 47 кандидатных SNP-маркеров для предрасположенности к социальному доминированию и подчинению соответственно, в границах 46 ± 2 стандартной ошибки ожидаемого их количества при равновероятном биномиальном распределении $B(n = 92; p = 0.5)$: значимость различия: $\alpha > 0.35$. Это может означать, что для возникновения доминантно-подчиненных (субординантных) отношений может быть достаточным присутствие в социуме особей, предрасположенных как к доминированию, так и к подчинению. Полученный результат статистически значимо отражает результат давления естественного отбора в пользу поддержания обеих стратегий поведения – доминирования и подчинения, т.е. нормы реакции, у человека.

Неожиданностью было также выявление 26 и 66 кандидатных SNP-маркеров, которые могут, соответственно, уменьшить и увеличить сродство ТВР/промотор для генов человека, способных влиять на поведение доминирования и подчинения. Эти количества достоверно отличаются от их ожидаемой оценки при равновероятном биномиальном распределении



Illustrative prediction of a candidate SNP marker for dominant and subordinate behavior in humans with the Web-service (Ponomarenko et al., 2015) in the case of the known SNP marker rs3813929 for resistance to olanzapine (Landrum et al., 2014) determined by an excess of serotonin receptor 2C (Ellingrod et al., 2005).

The DECISION row indicates that this prediction fits clinical data (Ellingrod et al., 2005).

($\alpha < 10^{-5}$). В этом смысле рассматриваемый результат может статистически достоверно отражать возможность давления естественного отбора на геном человека против значимого снижения экспрессии генов, влияющих на поведение доминирования и подчинения у людей.

Обсуждение

Мы проанализировали 381 SNP и выбрали из них 92 кандидатных SNP-маркера, которые способны значимо влиять на сродство ТВР к промоторам генов человека (24 %). Это согласуется как с теорией нейтральной эволюции (Kimura, 1968), так и с текущим представлением о снижении затрат на экспериментальный поиск ценных для практики SNP-маркеров в геноме человека методом «проб и ошибок» за счет предварительного компьютерного анализа известных SNP (Deplancke et al., 2016).

При этом мы учли лишь влияние SNP на связывание ТВР/промотор как самый изученный, облигатный и позиционированный этап регуляции экспрессии генов, вследствие чего любой из оставшихся 289 SNP также может оказаться кандидатным SNP-маркером для поведения доминирования и подчинения у человека при его влиянии на связывание геномной ДНК с другими белками или, например, при изменении им кодона в зоне перекрытия промотора одной РНК с белок-кодирующим экзоном в другой РНК, транскрибируемых с общего для них гена человека (например, SNP rs772061026). Кандидатные SNP-маркеры для социальной иерархии человека могут быть вне узкого района $[-70; -20]$ (длина 50 п.о.) относительно стартов транскрипции белок-кодирующих РНК человека, анализом которого мы ограничили эту работу.

Для компьютерного анализа различных локализаций и проявлений SNP на уровнях реорганизации хроматина, транскрипции, сплайсинга, трансляции и последующих модификаций белковых продуктов генов человека имеются свободно доступные Web-сервисы, которые можно найти, например, в обзоре (Podkolodnyu et al., 2014).

Достоверность предсказанных кандидатных SNP-маркеров мы оценили по их бинарным классификациям: доминирование/подчинение и рост/спад сродства TBP/промотор в рамках традиционного биоинформатического подхода к анализу геномов (Haldane, 1957; Kimura, 1968), например, скорость их эволюции оценивают по отношению частот транзиций и трансверсий (Kimura, 1980). В этой связи интересно сравнение генов, способных влиять на поведение доминирования и подчинения людей, со всем геномом человека в целом, охарактеризованным вдвое более частым уменьшением в сравнении с увеличением сродства транскрипционных факторов к промоторам с SNP (Kasowski et al., 2010; 1000 Genomes Project Consortium, 2012). Указанная полногеномная оценка диаметрально противоположна нашему предсказанию 26 и 66 кандидатных SNP-маркеров для соответственно уменьшения и увеличения сродства TBP к промоторам рассматриваемых генов. Столь высокая значимость ($\alpha < 10^{-12}$) отличия этих генов от всего генома человека в целом согласуется с существующим мнением об адаптивной роли социального поведения человека в его эволюции, в отличие от нейтральной эволюции генома в целом, согласно теории Кимура (Kimura, 1968) и дилемме Холдейна (Haldane, 1957). Это может дополнительно подтверждать достоверность нашего предсказания кандидатных SNP-маркеров для предрасположенности к доминированию или подчинению у человека (см. таблицу).

Исследования на инбредных линиях мышей позволили установить, что поведение доминирования и подчинения в значительной мере детерминировано генотипом (Науменко и др., 1983; Осадчук и др., 2009; Клещев и др., 2013). Это свидетельствует, вероятнее всего, о возможности предсказания кандидатных SNP-маркеров для тех или иных социальных рангов на основе компьютерного анализа последовательностей геномной ДНК. Но, наряду с генетическими, были описаны социальные механизмы передачи иерархических рангов от родителей к потомству, например у макак (Prud'Homme, Chapais, 1993), оленей (Dusek et al., 2007) и гиен (Engh et al., 2000). Кроме того, известно влияние морфометрических параметров, особенностей онтогенеза, гендерного диморфизма и социального опыта на ранг особи (van der Kooij, Sandi, 2015). Это означает, что реальная картина вклада индивидуального генома человека в его социальный ранг может быть гораздо сложнее, многообразнее, богаче, ярче и интереснее в сравнении с нашей предельно упрощенной гипотезой, которая была, однако, необходима как базовый инструмент в рамках традиционных биоинформатических подходов к компьютерному анализу SNP и референсного генома человека.

Заключение

Найдены SNP, способные влиять на сродство TBP к промоторам генов человека, гомологичным генам животных,

которые были ассоциированы с поведением доминирования и подчинения. Высказано предположение, что эти кандидатные SNP-маркеры могут иметь отношение к социальному доминированию и у людей. Как компьютерный прогноз, этот результат нуждается в дальнейшей экспериментальной проверке.

Acknowledgments

DAR acknowledges the support of the Web service development by the Russian Science Foundation, project 14-24-00123. EBSH and LKS acknowledge the support of data collection by the Russian Foundation for Basic Research, project 14-04-00485. MPP acknowledges the support of data analysis by the Russian Federation government, project 14.B25.31.0033 of July 28, 2013 according to Decree 220 of the Russian Federation government of April 9, 2010.

Conflict of interest

The authors declare no conflict of interest.

References

- 1000 Genomes Project Consortium. Abecasis G.R., Auton A., Brooks L.D., DePristo M.A., Durbin R.M., Handsaker R.E., Kang H.M., Marth G.T., McVean G.A. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. DOI 10.1038/nature11632.
- Abbas A., Lechevrel M., Sichel F. Identification of new single nucleotide polymorphisms (SNP) in alcohol dehydrogenase class IV ADH7 gene within a French population. *Arch. Toxicol.* 2006;80(4):201-205. DOI 10.1007/s00204-005-0031-7.
- Amberger J.S., Bocchini C.A., Schiettecatte F., Scott A.F., Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucl. Acids Res.* 2015;43:D789-D798. DOI 10.1093/nar/gku1205.
- Arkova O.V., Kuznetsov N.A., Fedorova O.S., Kolchanov N.A., Savinkova L.K. Real-time interaction between TBP and the TATA box of the human triosephosphate isomerase gene promoter in the norm and pathology. *Acta Naturae*. 2014;6(2):36-40.
- Arkova O.V., Ponomarenko M.P., Rasskazov D.A., Drachkova I.A., Arshinova T.V., Ponomarenko P.M., Savinkova L.K., Kolchanov N.A. Obesity-related known and candidate SNP markers can significantly change affinity of TATA-binding protein for human gene promoters. *BMC Genomics*. 2015;16(Suppl. 13):S5. DOI 10.1186/1471-2164-16-S13-S5.
- Belyaev D.K. The Wilhelmine E. Key 1978 invitational lecture. Destabilizing selection as a factor in domestication. *J. Hered.* 1979;70(5):301-308.
- Bondar N.P., Boyarskikh U.A., Kovalenko I.L., Filipenko M.L., Kudryavtseva N.N. Molecular implications of repeated aggression: Th, Dat1, Snca and Bdnf gene expression in the VTA of victorious male mice. *PLoS ONE*. 2009;4(1):e4190. DOI 10.1371/journal.pone.0004190.
- Cao J., Joyner L., Mickens J.A., Leyrer S.M., Patisaul H.B. Sex-specific ESR2 mRNA expression in the rat hypothalamus and amygdala is altered by neonatal bisphenol A exposure. *Reproduction*. 2014;147(4):537-554. DOI 10.1530/REP-13-0501.
- Chadaeva I.V., Ponomarenko M.P., Rasskazov D.A., Sharypova E.B., Kashina E.V., Matveeva M.Yu., Arshinova T.V., Ponomarenko P.M., Arkova O.V., Bondar N.P., Savinkova L.K., Kolchanov N.A. Candidate SNP markers of aggressiveness-related complications and comorbidities of genetic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics*. 17(Suppl.14):53. DOI 10.1186/s12864-016-3353-3.
- Colonna V., Ayub Q., Chen Y., Pagani L., Luisi P., Pybus M., Garrison E., Xue Y., Tyler-Smith C. 1000 Genomes Project Consortium. Abecasis G.R., Auton A., Brooks L.D., DePristo M.A., Durbin R.M.,

- Handsaker R.E., Kang H.M., Marth G.T., McVean G.A. Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. *Genome Biol.* 2014;15(6):R88. DOI 10.1186/gb-2014-15-6-r88.
- Delgadillo R.F., Whittington J.E., Parkhurst L.K., Parkhurst L.J. The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism. *Biochemistry.* 2009; 48(8):1801-1809. DOI 10.1021/bi8018724.
- Deplancke B., Alpern D., Gardeux V. The genetics of transcription factor DNA binding variation. *Cell.* 2016;166(3):538-554. DOI 10.1016/j.cell.2016.07.012.
- Drachkova I., Savinkova L., Arshinova T., Ponomarenko M., Peltek S., Kolchanov N. The mechanism by which TATA-box polymorphisms associated with human hereditary diseases influence interactions with the TATA-binding protein. *Hum. Mutat.* 2014;35(5):601-608. DOI 10.1002/humu.22535.
- Dusek A., Bartos L., Svecova L. The effect of a mother's rank on her offspring's pre-weaning rank in farmed red deer. *Appl. Anim. Behav. Sci.* 2007;103(1-2):146-155. DOI 10.1016/j.applanim.2006.03.020.
- Ehrman L., Parsons P.A. *Behavior genetics and evolution.* N.Y.: McGraw-Hill, 1981.
- Eldakar O.T., Gallup A.C. The group-level consequences of sexual conflict in multigroup populations. *PLoS ONE.* 2011;6(10):e26451. DOI 10.1371/journal.pone.0026451.
- Ellingrod V.L., Perry P.J., Ringold J.C., Lund B.C., Bever-Stille K., Fleming F., Holman T.L., Miller D. Weight gain associated with the -759C/T polymorphism of the 5HT2C receptor and olanzapine. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 2005;134B(1):76-78. DOI 10.1002/ajmg.b.20169.
- Engh A.L., Esch K., Smale L., Holekamp K.E. Mechanisms of maternal rank "inheritance" in the spotted hyaena, *Crocuta cr. Anim. Behav.* 2000;60(3):323-332. DOI 10.1006/anbe.2000.1502.
- Haeussler M., Raney B.J., Hinrichs A.S., Clawson H., Zweig A.S., Karolchik D., Casper J., Speir M.L., Haussler D., Kent W.J. Navigating protected genomics data with UCSC genome browser in a box. *Bioinformatics.* 2015;31(5):764-766. DOI 10.1093/bioinformatics/btu712.
- Haldane J.B.S. The cost of natural selection. *J. Genet.* 1957;55:511-524.
- Heyne H.O., Lautenschlager S., Nelson R., Besnier F., Rotival M., Cagan A., Kozhemyakina R., Plyusnina I.Z., Trut L., Carlborg O., Petroit E., Kruglyak L., Paabo S., Schoneberg T., Albert F.W. Genetic influences on brain gene expression in rats selected for tameness and aggression. *Genetics.* 2014;198(3):1277-1290. DOI 10.1534/genetics.114.168948.
- Hinde R.A. *Animal Behaviour.* New York: McGraw-Hill, 1970.
- Ilchibaeva T.V., Kondaurova E.M., Tsybko A.S., Kozhemyakina R.V., Popova N.K., Naumenko V.S. Brain-derived neurotrophic factor (BDNF) and its precursor (proBDNF) in genetically defined fear-induced aggression. *Behav. Brain Res.* 2015;290:45-50. DOI 10.1016/j.bbr.2015.04.041.
- Jegou S., El Yacoubi M., Mounien L., Ledent C., Parmentier M., Costentin J., Vaugeois J.M., Vaudry H. Adenosine A2A receptor gene disruption provokes marked changes in melanocortin content and pro-opiomelanocortin gene expression. *J. Neuroendocrinol.* 2003;15(12):1171-1177.
- Ji N.Y., Findling R.L. Pharmacotherapy for mental health problems in people with intellectual disability. *Curr. Opin. Psychiat.* 2016;29: 103-125. DOI 10.1097/YCO.0000000000000233.
- Kasowski M., Grubert F., Heffelfinger C., Hariharan M., Asabere A., Waszak S.M., Habegger L., Rozowsky J., Shi M., Urban A.E., Hong M.Y., Karczewski K.J., Huber W., Weissman S.M., Gerstein M.B., Korbel J.O., Snyder M. Variation in transcription factor binding among humans. *Science.* 2010;328(5975):232-235. DOI 10.1126/science.1183621.
- Kimura M. Evolutionary rate at the molecular level. *Nature.* 1968; 217(5129):624-626. DOI 10.1038/217624a0.
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 1980;16:111-120.
- Kleshchev M.A., Gutorova N., Osadchuk L.V. The genetic characteristics of age dynamics of agonistic behavior in male laboratory mice in a social hierarchy. *Ekologicheskaya genetika = Ecological genetics.* 2013;11(4):64-72. (in Russian)
- Kondaurova E.M., Ilchibaeva T.V., Tsybko A.S., Kozhemyakina R.V., Popova N.K., Naumenko V.S. 5-HT1A receptor gene silencers Freud-1 and Freud-2 are differently expressed in the brain of rats with genetically determined high level of fear-induced aggression or its absence. *Behav. Brain Res.* 2016;310:20-25. DOI 10.1016/j.bbr.2016.04.050.
- Kudryavtseva N.N., Bondar N.P., Boyarskikh U.A., Filipenko M.L. Snca and Bdnf gene expression in the VTA and raphe nuclei of mid-brain in chronically victorious and defeated male mice. *PLoS ONE.* 2010;5(11):e14089. DOI 10.1371/journal.pone.0014089.
- Kulikov A.V., Bazhenova E.Y., Kulikova E.A., Fursenko D.V., Trapezoza L.I., Terenina E.E., Mormede P., Popova N.K., Trapezo O.V. Interplay between aggression, brain monoamines and fur color mutation in the American mink. *Genes Brain Behav.* 2016. DOI 10.1111/gbb.12313.
- Landrum M.J., Lee J.M., Riley G.R., Jang W., Rubinstein W.S., Church D.M., Maglott D.R. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucl. Acids Res.* 2014;42(Database issue):D980-5. DOI 10.1093/nar/gkt1113.
- Lin D., Boyle M.P., Dollar P., Lee H., Lein E.S., Perona P., Anderson D.J. Functional identification of an aggression locus in the mouse hypothalamus. *Nature.* 2011;470(7333):221-226. DOI 10.1038/nature 09736.
- Lorenz K. *On Aggression.* Hove (UK): Psychology Press, 2002.
- Martianov I., Viville S., Davidson I. RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science.* 2002;298(5595):1036-1039. DOI 10.1126/science.1076327.
- Michopoulos V., Higgins M., Toufexis D., Wilson M.E. Social subordination produces distinct stress-related phenotypes in female rhesus monkeys. *Psychoneuroendocrinology.* 2012;37(7):1071-1085. DOI 10.1016/j.psyneuen.2011.12.004.
- Mogno I., Vallania F., Mitra R.D., Cohen B.A. TATA is a modular component of synthetic promoters. *Genome Res.* 2010;20(10):1391-1397. DOI 10.1101/gr.106732.110.
- Moore A.J. Genetic influences on social dominance: cow wars. *Heredity.* 2013;110(1):1-2. DOI 10.1038/hdy.2012.85.
- Naumenko E.V., Osadchuk A.V., Serova L.I., Shishkina G.T. *Genetiko-fiziologicheskie mekhanizmy regulatsii funktsiy semennikov [Genetic and Physiological Mechanisms of the Regulation of Testicular Functions].* Novosibirsk, Nauka Publ., 1983. (in Russian)
- Osadchuk L.B., Bragin A.V., Osadchuk A.V. Interstrain differences in social and time patterns of agonistic behavior in male laboratory mice. *Zhurnal vysshey nervnoy deyatel'nosti im. I.P. Pavlova = I.P. Pavlov Journal of Higher Nervous Activity.* 2009;59(4):473-481. (in Russian)
- Osadchuk L.B., Salomacheva I.N., Bragin A.V., Osadchuk A.V. The reproductive correlates of social hierarchy in laboratory male mice. *Zhurnal vysshey nervnoy deyatel'nosti im. I.P. Pavlova = I.P. Pavlov Journal of Higher Nervous Activity.* 2007;57(5):604-612. (in Russian)
- Podkolodnyy N.L., Afonnikov D.A., Vaskin Yu.Yu., Bryzgalov L.O., Ivanisenko V.A., Demenkov P.S., Ponomarenko M.P., Rasskazov D.A., Gunbin K.V., Protsyuk I.V., Shutov I.Yu., Leontyev P.N., Fursov M.Yu., Bondar N.P., Antontseva E.V., Merkulova T.I., Kolchanov N.A. Program complex SNP-MED for analysis of single-nucleotide polymorphism (SNP) effects on the function of genes associated with socially significant diseases. *Russ. J. Genet. Appl. Res.* 2014;4(3):159-167. DOI 10.1134/S2079059714030034.
- Ponomarenko M., Arkova O., Rasskazov D., Ponomarenko P., Savinkova L., Kolchanov N. Candidate SNP markers of gender-biased autoimmune complications of monogenic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front. Immunol.* 2016;7:130. DOI 10.3389/fimmu.2016.00130.
- Ponomarenko M., Mironova V., Gunbin K., Savinkova L. *Hogness Box.* Eds. S. Maloy, K. Hughes. *Brenner's Encyclopedia of Genet-*

- ics. San Diego: Acad. Press, Elsevier Inc. 2013;3:491-494. DOI 10.1016/B978-0-12-374984-0.00720-8.
- Ponomarenko M., Rasskazov D., Arkova O., Ponomarenko P., Suslov V., Savinkova L., Kolchanov N. How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter. *Biomed. Res. Int.* 2015;2015:35983004625. DOI 10.1155/2015/359835.
- Ponomarenko P.M., Ponomarenko M.P., Drachkova I.A., Lysova M.V., Arshinova T.V., Savinkova L.K., Kolchanov N.A. Prediction of the affinity of the TATA-binding protein to TATA boxes with single nucleotide polymorphisms. *Mol. Biol. (Mosk.)*. 2009;43(3):472-479. DOI 10.1134/S0026893309030157.
- Ponomarenko P., Rasskazov D., Suslov V., Sharypova E., Savinkova L., Podkolodnaya O., Podkolodny N., Tverdokhle N., Chadaeva I., Ponomarenko M., Kolchanov N. Candidate SNP markers of chronopathologies are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BioMed Res. Int.* 2016;2016:8642703. DOI 10.1155/2016/8642703.
- Ponomarenko P.M., Savinkova L.K., Drachkova I.A., Lysova M.V., Arshinova T.V., Ponomarenko M.P., Kolchanov N.A. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. *Dokl. Biochem. Biophys.* 2008;419:88-92. DOI 10.1134/S1607672908020117.
- Ponomarenko P.M., Suslov V.V., Savinkova L.K., Ponomarenko M.P., Kolchanov N.A. A precise equilibrium equation for four steps of binding between TBP and TATA-box allows for the prediction of phenotypic expression upon mutation. *Biophysics*. 2010;55(3):358-369. DOI 10.1134/S0006350910030036.
- Popova N.K. From genes to aggressive behavior: the role of serotonergic system. *Bioessays*. 2006;28(5):495-503. DOI 10.1002/bies.20412.
- Popova N.K., Naumenko V.S., Kozhemyakina R.V., Plyusnina I.Z. Functional characteristics of serotonin 5-HT_{2A} and 5-HT_{2C} receptors in the brain and the expression of the 5-HT_{2A} and 5-HT_{2C} receptor genes in aggressive and non-aggressive rats. *Neurosci. Behav. Physiol.* 2010;40(4):357-361. DOI 10.1007/s11055-010-9264-x.
- Prud'Homme J., Chapais B. Aggressive interventions and matrilineal dominance relations in semifree-ranging Barbary macaques (*Macaca sylvanus*). *Primates*. 1993;34(3):271-283. DOI 10.1007/BF02382621.
- Reif A., Jacob C.P., Rujescu D., Herterich S., Lang S., Gutknecht L., Baehne C.G., Strobel A., Freitag C.M., Giegling I., Romanos M., Hartmann A., Rosler M., Renner T.J., Fallgatter A.J., Retz W., Ehli A.C., Lesch K.P. Influence of functional variant of neuronal nitric oxide synthase on impulsive behaviors in humans. *Arch. Gen. Psychiat.* 2009;66(1):41-50. DOI 10.1001/archgenpsychiatry.2008.510.
- Rosenfeld C.S. Bisphenol A and phthalate endocrine disruption of parental and social behaviors. *Front. Neurosci.* 2015;9:57. DOI 10.3389/fnins.2015.00057.
- Rowell T.E. The concept of social dominance. *Behav. Biol.* 1974;11:131-154.
- Savinkova L., Drachkova I., Arshinova T., Ponomarenko P., Ponomarenko M., Kolchanov N. An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein. *PLoS ONE*. 2013;8(2):e54626. DOI 10.1371/journal.pone.0054626.
- Savonenko A., Filipkowski R.K., Werka T., Zielinski K., Kaczmarek L. Defensive conditioning-related functional heterogeneity among nuclei of the rat amygdala revealed by c-Fos mapping. *Neuroscience*. 1999;94(3):723-733.
- Serova L.I., Kozlova O.N., Naumenko E.V. The significance of the genotype and individual behavioral characteristics for the manifestation of the dominant phenotype in mouse micropopulations. *Zhurnal vysshey nervnoy deyatel'nosti im. I.P. Pavlova = I.P. Pavlov Journal of Higher Nervous Activity*. 1991;41(1):79-84. (in Russian)
- Sherry S.T., Ward M.H., Kholodov M., Baker J., Phan L., Smigielski E.M., Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucl. Acids Res.* 2001;29(1):308-311. DOI 10.1093/nar/29.1.308.
- Suslov V.V., Ponomarenko P.M., Ponomarenko M.P., Drachkova I.A., Arshinova T.V., Savinkova L.K., Kolchanov N.A. TATA box polymorphisms in genes of commercial and laboratory animals and plants associated with selectively valuable traits. *Russ. J. Genet.* 2010;46(4):394-403. DOI 10.1134/S1022795410040022.
- Van der Kooij M.A., Sandi C. The genetics of social hierarchies. *Cur. Opin. Behav. Sci.* 2015;2:52-57. DOI 10.1016/j.cobeha.2014.09.001.
- Zapata I., Serpell J.A., Alvarez C.E. Genetic mapping of canine fear and aggression. *BMC Genomics*. 2016;17:572. DOI 10.1186/s12864-016-2936-3.
- Zerbino D.R., Wilder S.P., Johnson N., Juettemann T., Flicek P.R. The Ensembl regulatory build. *Genome Biol.* 2015;16:56. DOI 10.1186/s13059-015-0621-5.

Выявление новых регуляторных SNP, ассоциированных с предрасположенностью к развитию колоректального рака

Е.Ю. Леберфарб , Л.О. Брызгалов, И.И. Брусенцов, Т.И. Меркулова

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

Разработан новый подход к поиску регуляторных SNP (rSNP), основанный на анализе ChIP-seq и RNA-seq данных проекта ENCODE. Подход успешно применен для выявления rSNP, связанных с колоректальным раком. В качестве исходных данных были взяты результаты 15 независимых ChIP-Seq экспериментов, выполненных на клеточной линии колоректального рака HCT-116, что позволило сформировать пул из 7985 SNP, расположенных в регуляторных районах генов. Для дальнейшего отбора регуляторных SNP из этого пула использована выявляемая в экспериментах ChIP-seq аллель-специфичность распределения гистоновых меток и негистоновых белков хроматина в местах локализации гетерозиготных SNP. Это позволило выявить 775 SNP, которые потенциально могут влиять на уровень экспрессии генов в клетках HCT-116. В дальнейшем была проведена оценка асимметричности экспрессии аллелей на основании данных RNA-seq, полученных на той же клеточной линии. В результате была подтверждена функциональность 231 SNP, которые были классифицированы как rSNP. Для отбора из них тех, что могут иметь отношение к развитию колоректального рака, были взяты rSNP, находящиеся в одной группе сцепления (± 10 т. п. н.) с SNP, ассоциированными с этим заболеванием, по данным GWAS и ClinVar. Функциональная аннотация генов, содержащих выбранные таким образом SNP, подтвердила полученные ранее данные о роли генов *BAIAP2L1* и *BUB3* в формировании предрасположенности к раку толстого кишечника. Также были обнаружены новые гены-кандидаты колоректального рака, *RRAGD* и *FZD6*, белковые продукты которых являются компонентами RAS/MAPK- и WNT-сигнальных путей, сопряженных с развитием данной патологии. Кроме того, выявлено 14 новых потенциальных генов-кандидатов колоректального рака, перспективных для дальнейшего изучения.

Ключевые слова: регуляторные SNP (rSNP); колоректальный рак; полногеномные данные; проект ENCODE.

Detection of new regulatory SNPs associated with colorectal cancer predisposition

E.Yu. Leberfarb , L.O. Bryzgalov, I.I. Brusentsov, T.I. Merkulova

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

A new approach to the search for regulatory SNPs (rSNPs) based on the use of ENCODE project data on ChIP-seq and RNA-seq experiments was developed. The approach was successfully used for the detection of rSNPs associated with colorectal cancer susceptibility. To start out with, we used raw sequence data of 15 independent ChIP-seq experiments run on colorectal cancer cell line HCT-116, which allowed us to generate the initial pool of 7985 SNPs located in regulatory regions. For further selection of functional SNPs, we used the ChIP-seq binding bias analysis and revealed 775 SNPs that are more likely to influence transcription regulation in HCT-116 cells. Then the RNA-seq bias analysis in HCT-116 cells was performed. As a result, we confirmed the functionality of 231 SNPs, which were classified as rSNPs. In order to select rSNPs potentially associated with colorectal cancer we chose those in strong linkage disequilibrium with SNPs associated with this pathology according to GWAS and ClinVar data. Functional annotation analysis of genes containing the rSNPs selected confirmed the involvement of *BAIAP2L1* and *BUB3* genes in colorectal cancer predisposition. We also found two genes (*RRAGD* and *FZD6*) playing a role in the RAS/MAPK and WNT signaling pathways. Although the involvement of the RAS/MAPK and WNT pathways in colon cancer is a well-known fact, these two genes are still unknown candidates. Moreover, we found 14 new candidate genes with promise for further study of colorectal cancer predisposition.

Key words: regulatory SNPs (rSNP); colorectal cancer; next generation sequencing; ENCODE project.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Леберфарб Е.Ю., Брызгалов Л.О., Брусенцов И.И., Меркулова Т.И. Выявление новых регуляторных SNP, ассоциированных с предрасположенностью к развитию колоректального рака. Вавиловский журнал генетики и селекции. 2016;20(6):797-803. DOI 10.18699/VJ16.197

HOW TO CITE THIS ARTICLE:

Leberfarb E.Yu., Bryzgalov L.O., Brusentsov I.I., Merkulova T.I. Detection of new regulatory SNPs associated with colorectal cancer predisposition. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):797-803. DOI 10.18699/VJ16.197

ORIGINAL ARTICLE

Received 11.03.2016

Accepted for publication 26.04.2016

© AUTHORS, 2016

 e-mail: Leberfarb@mail.ru

Колоректальный рак является одним из наиболее распространенных онкологических заболеваний и вносит существенный вклад в показатели смертности населения по всему миру. В настоящее время в структуре онкологических заболеваний на эту патологию приходится приблизительно 9.7 % (Ferlay et al., 2015).

Колоректальный рак представляет собой мультифакторное заболевание с четко выраженной генетической компонентой. По современным оценкам, от 15 до 35 % риска развития этой патологии можно объяснить наследственными факторами (Taylor et al., 2010). При этом известно, что до 6 % случаев возникновения колоректального рака связано с тремя хорошо охарактеризованными наследственными заболеваниями: семейным аденоматозным полипозом (familial adenomatous polyposis, FAP), MUTYH-ассоциированным полипозом (MUTYH-associated polyposis, MAP) и наследственным неполипозным колоректальным раком (HNPCC, или синдром Линча) (Kaz, Brentnall, 2006; Jaspersion, 2014). Причиной FAP являются мутации в гене *APC*, кодирующем важный регулятор Wnt-сигнального пути, MAP – мутации в гене *MUTYH*, кодирующем один из ключевых ферментов системы эксцизионной репарации ДНК, а синдром Линча обусловлен мутациями в генах *MLH1*, *MSH2*, *MSH6* и *PMS2*, кодирующих ферменты системы репарации ошибочно спаренных нуклеотидов (Al-Tassan et al., 2002; Castells et al., 2009; Jaspersion et al., 2010; Fearon, 2011). В то же время многие случаи колоректального рака демонстрируют связь с пока неизвестными генетическими причинами (Esteban-Jurado, 2014). Кроме того, можно предполагать, что ряд спорадических форм этого заболевания также может быть связан с редкими и до сих пор не выявленными мутациями в как уже известных, так и новых кандидатных генах. В этой связи нельзя переоценить значение работ по поиску новых молекулярных маркеров предрасположенности к колоректальному раку для разработки критериев формирования группы диспансерного учета.

Однонуклеотидные полиморфизмы (single nucleotide polymorphism, SNP) – самый распространенный тип геномных вариаций. В результате прогресса в технологиях секвенирования число известных SNP в геноме человека из года в год увеличивается и в настоящее время составляет более 150 млн (<http://www.ncbi.nlm.nih.gov/feed/rss.cgi>). Предполагается, что большинство SNP относятся к нейтральным и не оказывают заметного влияния на фенотип, и только малая доля замен характеризуется фенотипическими проявлениями посредством изменения функции гена или уровня его экспрессии (Chung, Chanock, 2011; Zhao et al., 2011). В настоящее время основным методом выявления связи между генетическими вариациями и признаком является полногеномный поиск ассоциаций (genome-wide association studies, GWAS). Согласно данным каталога NIH GWAS, уже более 15 тыс. SNP, а также мелких делеций/инсерций ассоциировано с болезнями и другими фенотипическими проявлениями с помощью этого метода (Welter et al., 2014). Важно отметить, что около 90 % таких вариаций находятся в некодирующих районах генома и могут быть связаны с регуляцией экспрессии генов (Manolio et al., 2009; Maurano et al., 2012). Однако, по результатам GWAS, невозможно от-

личить действительно связанный с болезнью вариант множества маркерных, выявленных в исследованиях за счет неравновесия по сцеплению (Edwards et al., 2013). Кроме того, установление ассоциаций тех или иных генетических вариантов с заболеванием не дает сведений о функциональности этих вариантов, и, соответственно, о молекулярных механизмах развития патологии, что тормозит разработку эффективных методов ее лечения и профилактики (Lappalainen, 2015).

В решении этих проблем может существенно помочь информация о регуляторных элементах генома (ChIP-seq) и транскрипционных профилях (RNA-seq), накапливаемая в базах данных консорциума ENCODE (The Encyclopedia of DNA Elements). В частности, данные ChIP-seq экспериментов об аллель-специфичности в связывании различных транскрипционных факторов можно рассматривать как указание на регуляторную функцию SNP, расположенных в участках связывания этих факторов (Kasowski et al., 2010; Reddy et al., 2012; Rozowsky et al., 2014). Таким же указанием на функциональность SNP могут служить данные по аллель-специфичности распределения гистоновых меток активного хроматина в местах локализации SNP (Trynka et al., 2015). А включение в анализ данных по транскриптомам позволит связать такие потенциально регуляторные SNP с уровнем экспрессии конкретных генов (Reddy et al., 2012; Rozowsky et al., 2014). Мы использовали в работе результаты экспериментов ChIP-seq и RNA-seq из проекта ENCODE, полученные на клеточной линии колоректального рака HCT-116, для выявления новых регуляторных SNP, которые могут быть связаны с предрасположенностью к развитию этого заболевания.

Материалы и методы

Были использованы материалы ChIP-seq (с антителами к ряду гистоновых меток и негистоновых белков) (таблица) и RNA-seq (SRX159835) экспериментов, выполненных на линии HCT-116, депонированные в SRA (sequencing read archive NCBI, <http://www.ncbi.nlm.nih.gov/sra/>).

Предпроцессинг секвенсов проводили с использованием программы Trimmomatic-3.2.2 (Bolger et al., 2014). Данные секвенирования выравнивали на референсный геном человека GRCh37/Hg19 с помощью программы Bowtie2 (Langmead, Salzberg, 2012). Потенциальные ПЦР-дубликаты были удалены с помощью программы SAMtools (Li et al., 2009). Для уменьшения ошибок секвенирования и выравнивания использовали только уникально картированные риды с QMAP > 25 (SAMtools). Для поиска SNP и оценки их покрытия в районах OTFR (Bryzgalov et al., 2013) применяли программу SAMtools mpileup (Li et al., 2009) и программы, написанные нами на perl. В дальнейшем анализе не использовались SNP, попадающие в тандемные повторы (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/simpleRepeat.txt>), в половые хромосомы, а также микроделеции и инсерции. Для исключения соматических мутаций в дальнейшую работу были взяты только те SNP, которые аннотированы в dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>).

Для коррекции смещения величины покрытия в сторону референсного аллеля мы сконструировали альтернативный геном, в который были введены найденные

The data used in the work

Antibody	Experiment
H3K27me3	SRX105258, SRX157411, SRX157410
H3K4me1	SRX360379, SRX360378
H3K4me2	SRX360381, SRX360380
H3K4me3	SRX105259, SRX360383, SRX360382, SRX157409, SRX157408
MLL4	SRX360376
Sp1	SRX113706
p300	SRX360377

нами в HCT-116 полиморфизмы. После этого был проведен повторный раунд выравнивания уже на этот альтернативный геном. Затем на основании выравниваний на референсный и альтернативный геномы было оценено покрытие каждого аллеля (Ni et al., 2012). В результате из всего массива найденных SNP были отобраны те, у которых покрытия аллелей достоверно различаются ($p < 0,01$) – aSNP (асимметричные SNP). Затем по данным аннотации генома человека HG19/GRCh37.75 (<http://ftp.ensembl.org/>) были выбраны гены, у которых найденные нами aSNP были локализованы в области от –1 800 п. н. до старта транскрипции или в теле гена.

Данные RNAseq были использованы для оценки экспрессии аллелей. Для этого были найдены SNP, присутствующие в гетерозиготе в линии HCT116, которые в дальнейшем использовали как маркерные. Был построен альтернативный геном, содержащий данные SNP. Затем на основании выравниваний на референсный и альтернативный геномы была оценена асимметрия представленности аллелей в транскриптом. Из генов, содержащих aSNP, были отобраны гены с достоверным различием ($p < 0,01$) экспрессии аллелей. aSNP, связанные с такими генами, были классифицированы как регуляторные (rSNP).

Для поиска rSNP, ассоциированных с онкологическими заболеваниями, были отобраны те, которые находятся в одной группе сцепления (± 10 т. п. н.) с SNP, ассоциированными с этими заболеваниями, по данным GWAS (<http://www.gwascentral.org/>) и ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>). Функциональный анализ генов, содержащих отобранные rSNP, проводили с использованием аннотаций GO, основанных на экспериментальных данных (Inferred from Experiment (EXP), Inferred from Direct Assay (IDA), Inferred from Physical Interaction (IPI), Inferred from Mutant Phenotype (IMP), Inferred from Genetic Interaction (IGI), Inferred from Expression Pattern (IEP)) (<http://geneontology.org/>).

Обработка данных была проведена в среде программирования R.

Результаты

На рис. 1 приведена схема предложенного нами подхода к поиску потенциально регуляторных SNP, связанных с развитием колоректального рака.

В качестве исходного материала использованы данные проекта ENCODE, включающие 15 экспериментов

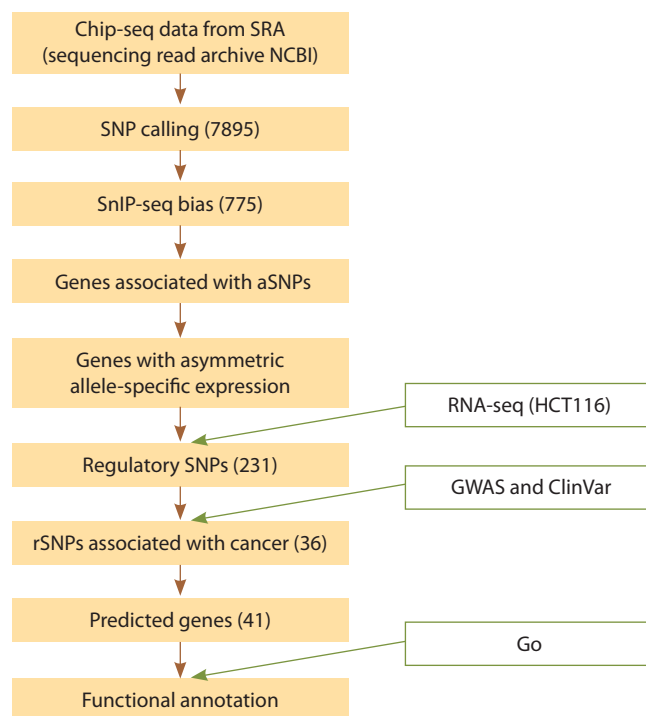


Fig. 1. Genome-wide search for regulatory SNPs.

ChIP-seq для клеточной линии колоректального рака HCT-116. Эксперименты были выполнены с применением антител к различным модификациям гистона H3 (H3K4me1, H3K4me2, H3K4me3 и H3K27me3), а также к транскрипционному фактору Sp1, коактиватору p300 и H3K4 метилтрансферазе MLL4.

Известно, что обнаружение гистоновых меток активно-го хроматина, H3K4me1, H3K4me2, H3K4me3, H3K27me3 и других, так же, как и коактиваторных белков, в определенных районах генома указывает на регуляторную роль этих районов и возможную регуляторную функцию локализованных в них SNP (Ernst et al., 2011; Sahu et al., 2013). Однако наши предыдущие исследования показали, что более универсальным и надежным маркером регуляторных районов генома является их обогащение участками связывания факторов транскрипции (пиками ChIP-seq), которое определяется в результате анализа множества ChIP-seq экспериментов по полногеномному распределению мест связывания этих белков (Bryzgalov et al., 2013). Нами было найдено около 200000 таких регуляторных районов в геноме человека, которые получили название OTFR (overlapping TF binding regions). Поэтому для более надежного определения регуляторных SNP мы проводили их поиск в участках, одновременно являющихся местами локализации исследованных гистоновых меток и негистоновых белков в клетках HCT-116 и найденными нами ранее OTFR.

В результате проведенного поиска было обнаружено 7895 SNP. Для обнаружения функциональных полиморфизмов мы провели анализ данных ChIP-seq на предмет асимметричной представленности аллелей в преципитатах. Для избежания асимметрии, возникающей в ре-

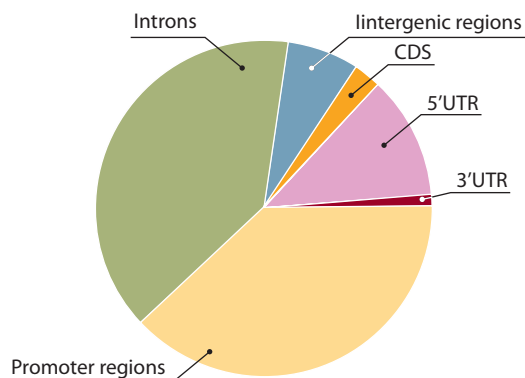


Fig. 2. Distribution of detected aSNPs over genome regions.

зультате предпочтительного картирования референсного аллеля на геном, мы построили альтернативный геном, содержащий альтернативные варианты всех выявленных SNP. После выравнивания на альтернативный геном был произведен подсчет покрытия каждого аллеля и выявлено 775 SNP с асимметричной представленностью аллелей в преципитатах (aSNP). Данные aSNP могут влиять на связывание транскрипционных факторов и состояние хроматина и, как следствие, – на активность определенных генов. При оценке распределения aSNP в геноме мы установили, что большинство их расположено в интронах и промоторной области (от –1800 п. н. до старта транскрипции) (рис. 2).

В межгенные промежутки попадает 81 (6.85 %) aSNP, что не позволяет связать их с экспрессией какого-либо конкретного гена. Они были исключены из дальнейшего анализа.

Для определения влияния остальных aSNP на экспрессию генов были использованы данные RNA-seq для линии HCT-116. В экзонах генов, связанных с aSNP, были найдены маркерные гетерозиготные полиморфизмы, которые позволили оценить асимметричность экспрессии аллелей. В ряде случаев эти маркеры совпадали с aSNP, расположенными в 5'- и 3'-UTR. В итоге был найден 231 rSNP, для которых мы выявили асимметричную экспрессию аллелей связанных с ними генов.

Для выяснения связи найденных rSNP с онкологическими заболеваниями использовали базы данных GWAS и ClinVar. Были отобраны rSNP, которые находятся в одной группе сцепления (± 10 т. п. н.) с SNP, ассоциированными с этими заболеваниями. Таким образом, было выявлено 36 rSNP в регуляторных районах 41 гена. Из них семь SNP могут оказывать влияние на два и более генов. Также имеется шесть генов, содержащих несколько полиморфизмов. При этом ни один из найденных нами rSNP не совпал с SNP, ранее зарегистрированными в этих базах и ассоциированными с онкозаболеваниями.

Первичное описание найденных генов проводили с использованием аннотаций GO, основанных на экспериментальных данных (<http://geneontology.org/>). Оказалось, что 18 из 41 гена не имеют аннотации в GO. Для 12 из 23 оставшихся генов показана их роль в канцерогенезе. Среди них следует отметить ген *RRAGD*, продукт которого яв-

ляется компонентом RAS/MAPK сигнального пути, и ген *FZD6*, продукт которого – компонент WNT-сигнального пути. Известно, что мутации ряда генов, кодирующих компоненты WNT- и RAS/MAPK-сигнальных путей, способствуют развитию колоректального рака (Fearon, 2011; The Cancer Genome Atlas Network, 2013). В нашем исследовании обнаружены новые гены-кандидаты развития колоректального рака, белковые продукты которых принадлежат данным сигнальным путям. Кроме того, найден rSNP в гене *BALAP2LI*, который, по данным литературы, относится к кандидатным генам развития рака толстого кишечника (Cohen, 2003). Белковый продукт этого гена принадлежит к Rho/Rac семейству, участвует в подавлении апоптоза за счет убиквитинилирования p53 и способствует пролиферации клеток в результате активации EGFR-ERK пути (Wang et al., 2013). К известным генам-кандидатам развития колоректального рака относится и *BUB3*. Показано, что мутации в кодирующей части гена *BUB3* увеличивают риск развития этой патологии в молодом возрасте (De Voer et al., 2013). Мы нашли связанный с колоректальным раком SNP в регуляторной области этого гена. Для восьми генов (*ZFP276*, *ACADVL*, *PAFAH1B2*, *MC1R*, *ZNF652*, *CDK10*, *TUBB3*, *P4HB*) связь с колоректальным раком пока не показана, но участие продуктов этих генов в развитии различных злокачественных опухолей (Wong et al., 2003; Pan et al., 2010; Yu et al., 2012; Sun et al., 2013; Kanojia et al., 2015; Kohnz et al., 2015; Laitinen et al., 2015; Tagliabue et al., 2015) дает основание предполагать их роль в развитии и данной патологии.

Еще для 11 генов, содержащих выявленные нами rSNP, связь с канцерогенезом в литературе не описана. Однако среди них присутствуют шесть генов, продукты которых участвуют в таких процессах, как клеточная пролиферация (*COMMD5* (Solban et al., 2000)) и дифференцировка (*RNF114* (Han et al., 2013)), апоптоз (*TARDBP* (Li et al., 2015), *UBE2Z* (Babu et al., 2011), *SCO2* (Madan et al., 2013)), репарация ДНК (*ENDOV* (Zhang et al., 2015)). Это также дает основание предполагать их возможную вовлеченность в развитие злокачественных опухолей, включая колоректальный рак.

Обсуждение

Генетическая предрасположенность играет существенную роль в развитии рака толстого кишечника (Taylor et al., 2010). Выявлен целый ряд генов, генеративные мутации в которых приводят (или способствуют) к развитию данной патологии (*APC*, *MLH1*, *MLH2*, *PTEN* и др.). Обнаружено также множество соматических мутаций, с высокой частотой представленных в клинических образцах колоректального рака. Прежде всего, это мутации генов, кодирующих компоненты WNT-, TGF- β и RAS-MAPK-сигнальных путей, и гены, кодирующие ферменты репарации повреждений ДНК (Fearon, 2011). Важно отметить, что и генеративные, и соматические мутации в большинстве случаев затрагивают одни и те же гены. Более того, часть найденных в опухолях мутаций соответствуют SNP, с повышенной частотой встречающихся в геномах больных раком толстого кишечника. Объединение всех этих данных позволяет в общих чертах наметить картину нарушений генных взаимодействий, ведущих к

колоректальному раку. Однако, несмотря на значительные усилия многих лабораторий в мире, полная реконструкция данной картины весьма далека от завершения.

В последние годы с развитием технологий NGS (next generation sequencing) широкое распространение получили полногеномные исследования. Они открывают широкие возможности для идентификации новых генов, ответственных за предрасположенность к тем или иным патологиям, и для обнаружения генетических вариаций, в первую очередь SNP, повышающих риск развития заболеваний. В настоящее время для поиска таких вариаций наиболее часто используемой технологией NGS является полноэкзомное секвенирование, позволяющее с высокой надежностью и эффективностью проводить исследования большого количества биологических образцов. Так, применение этой технологии к изучению экзотов более 200 образцов опухолей толстого кишечника и прилегающей нормальной ткани, наряду с подтверждением мутаций в уже известных кандидатных генах WNT-, TGF- β и RAS-MAPK-сигнальных путей и ферментов репарации, позволило выявить ряд новых генов, мутации в которых связаны с колоректальным раком. В частности, среди них был обнаружен ген *Sox9*, кодирующий важный регулятор плюрипотентности, который ранее не был ассоциирован ни с одним типом рака (The Cancer Genome Atlas Network, 2013). С помощью полноэкзомного секвенирования была также выявлена вызывающая колоректальный рак генеративная нонсенс-мутация в гене системы эксцизионной репарации *NTHL1* (Weren et al., 2015); обнаружены ассоциированные с этим заболеванием миссенс-варианты в генах *CENPE* и *KIF23* (DeRycke et al., 2013); в финской когорте больных с семейным раком толстого кишечника найдены 11 генов с редкими мутациями, приводящими к синтезу укороченных белков (Gylfe et al., 2013); показано, что rs10795668 в гене *FLJ3802842* и rs4631962 в *CCND2* достоверно ассоциированы с риском развития колоректального рака в тайваньской популяции (Yang et al., 2014) и т.д.

Полноэкзомное секвенирование представляет собой технологию таргетного секвенирования, охватывающего всю совокупность белок-кодирующих последовательностей (~1 % генома), фланкирующие экзоны, районы интронов и некоторые некодирующие РНК (Ng et al., 2010). Его основным преимуществом по сравнению с полногеномным секвенированием является достижение существенно лучшего покрытия при меньших затратах, что позволяет с высокой надежностью и эффективностью идентифицировать генетические вариации, включая SNP. В настоящее время в отношении регуляторных районов специальных технологий таргетного секвенирования еще не разработано. Однако многочисленные ChIP-Seq эксперименты, направленные на выявление и изучение регуляторных районов генома, «попутно» осуществляют таргетное секвенирование этих районов.

В настоящей работе в качестве данных таргетного секвенирования регуляторных районов использованы результаты 15 независимых ChIP-Seq экспериментов, выполненных на клеточной линии колоректального рака HCT-116, что позволило сформировать исходный пул из 7985 потенциально регуляторных SNP. Для дальнейшего отбора регуляторных SNP мы основывались на выявля-

емой в экспериментах ChIP-seq аллель-специфичности распределения гистоновых меток и негистоновых белков хроматина в местах локализации гетерозиготных SNP, которая рассматривается в качестве важного указания на регуляторную функцию таких SNP (Kasowski et al., 2010; Reddy et al., 2012; Trynka et al., 2015). Это позволило выявить 775 SNP, которые потенциально могут влиять на уровень экспрессии генов в клетках HCT-116. Для проверки этого предположения была проведена оценка аллель-специфичности экспрессии генов с использованием данных RNA-seq, полученных на той же клеточной линии. В результате была подтверждена функциональность 231 SNP (rSNP). Для выбора полиморфизмов, относящихся к развитию колоректального рака, были отобраны rSNP, находящиеся в одной группе сцепления (± 10 т.п.н.) с SNP, ассоциированными с этим заболеванием, по данным GWAS и ClinVar. С помощью этого подхода было выявлено 36 rSNP в регуляторных районах 41 гена. Функциональная аннотация этих генов подтвердила полученные ранее данные о роли генов *BALP2L1* и *BUB3* в формировании предрасположенности к раку толстого кишечника (Cohen, 2003). Причем если ранее для гена *BUB3* были найдены связанные с колоректальным раком мутации в его кодирующей части (Cohen, 2003), с помощью предложенного подхода такие мутации были найдены в районе, соответствующем 5'-UTR. Также были получены новые данные, подтверждающие вовлеченность в развитие колоректального рака RAS/MAPK и WNT-сигнальных путей. В нашем исследовании обнаружены новые гены-кандидаты развития колоректального рака, *RRAGD* и *FZD6*, белковые продукты которых являются компонентами данных сигнальных путей. Кроме того, выявлено 14 новых потенциальных генов-кандидатов колоректального рака, перспективных для дальнейшего изучения.

Таким образом, разработан новый подход к поиску регуляторных SNP, основанный на анализе данных проекта ENCODE (The Encyclopedia of DNA Elements, The ENCODE Project Consortium 2012) по полногеномному распределению мест локализации гистоновых меток и негистоновых белков, а также определению транскриптома. Подход успешно опробован при использовании такого комплекса данных, полученных на клеточной линии колоректального рака HCT116, для поиска rSNP, связанных с этим заболеванием. При условии достаточного объема полногеномных данных этот подход может быть применен к поиску rSNP, вовлеченных в развитие любой мультифакторальной патологии.

Acknowledgments

This work was supported by the Siberian Branch of the Russian Academy of Sciences, Integrated Program, project 0324-2015-0029, and the Russian Foundation for Basic Research, project 15-04-05780.

Conflict of interest

The authors declare no conflict of interest.

References

Al-Tassan N., Chmiel N.H., Maynard J., Fleming N., Livingston A.L., Williams G.T., ... Cheadle J.P. Inherited variants of MYH associated

- with somatic G:C→T:A mutations in colorectal tumors. *Nat. Genet.* 2002;30(2):227-232. <http://doi.org/10.1038/ng828>.
- Babu S.G., Ponia S.S., Kumar D., Saxena S. Cellular oncomiR orthologue in EBV oncogenesis. *Comput. Biol. Med.* 2011;41(10):891-898. <http://doi.org/10.1016/j.compbiomed.2011.07.007>.
- Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120. <http://doi.org/10.1093/bioinformatics/btu170>.
- Bryzgalov L.O., Antontseva E.V., Matveeva M.Y., Shilov A.G., Kashina E.V., Mordvinov V.A., Merkulova T.I. Detection of regulatory SNPs in human genome using ChIP-seq ENCODE data. *PLoS ONE.* 2013;8(10):e78833. <http://doi.org/10.1371/journal.pone.0078833>.
- Castells A., Castellvi-Bel S., Balaguer F. Concepts in familial colorectal cancer: where do we stand and what is the future? *Gastroenterology.* 2009;137(2):404-409. <http://doi.org/10.1053/j.gastro.2009.06.015>.
- Chung C.C., Chanock S.J. Current status of genome-wide association studies in cancer. *Hum. Genet.* 2011;130(1):59-78. <http://doi.org/10.1007/s00439-011-1030-9>.
- Cohen M.M. Molecular dimensions of gastrointestinal tumors: Some thoughts for digestion. *Am. J. Med. Gen.* 2003;122A(4):303-314. <http://doi.org/10.1002/ajmg.a.20473>.
- De Voer R.M., Geurts van Kessel A., Weren R.D.A., Ligtenberg M.J.L., Smeets D., Fu L., ... Kuiper R.P. Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. *Gastroenterology.* 2013;145(3):544-547. <http://doi.org/10.1053/j.gastro.2013.06.001>.
- DeRycke M.S., Gunawardena S.R., Middha S., Asmann Y.W., Schaid D.J., McDonnell S.K., ... Goode E.L. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Co-sponsored by the American Society of Preventive Oncology.* 2013; 22(7):1239-1251. <http://doi.org/10.1158/1055-9965.EPI-12-1226>.
- Edwards S.L., Beesley J., French J.D., Dunning A.M. Beyond GWASs: Illuminating the dark road from association to function. *Am. J. Hum. Genet.* 2013;93(5):779-797. <http://doi.org/10.1016/j.ajhg.2013.10.012>.
- Ernst J., Kheradpour P., Mikkelsen T.S., Shores N., Ward L.D., Epstein C.B., ... Bernstein B.E. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011;473(7345):43-49. <http://doi.org/10.1038/nature09906>.
- Esteban-Jurado C. New genes emerging for colorectal cancer predisposition. *World J. Gastroenterol.* 2014;20(8):1961. <http://doi.org/10.3748/wjg.v20.i8.1961>.
- Fearon E.R. Molecular genetics of colorectal cancer. *Ann. Rev. Pathol.: Mech.* 2011;6(1):479-507. <http://doi.org/10.1146/annurev-pathol-011110-130235>.
- Ferlay J., Soerjomataram I., Dikshit R., Eser S., Mathers C., Rebelo M., ... Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer.* 2015;136(5):E359-E386. <http://doi.org/10.1002/ijc.29210>.
- Gylfe A.E., Katainen R., Kondelin J., Tanskanen T., Cajuso T., Hänninen U., ... Aaltonen L.A. Eleven candidate susceptibility genes for common familial colorectal cancer. *PLoS Gen.* 2013;9(10):e1003876. <http://doi.org/10.1371/journal.pgen.1003876>.
- Han J., Kim Y.-L., Lee K.-W., Her N.-G., Ha T.-K., Yoon S., ... Chi S.-G. ZNF313 is a novel cell cycle activator with an E3 ligase activity inhibiting cellular senescence by destabilizing p21(WAF1). *Cell Death Differ.* 2013;20(8):1055-1067. <http://doi.org/10.1038/cdd.2013.33>.
- Jaspersion K.W. Hereditary colorectal cancer: More common than you think. *Curr. Prob. Cancer.* 2014;38(6):249-261. <http://doi.org/10.1016/j.cuprocancer.2014.10.005>.
- Jaspersion K.W., Tuohy T.M., Neklasen D.W., Burt R.W. Hereditary and familial colon cancer. *Gastroenterology.* 2010;138(6):2044-2058. <http://doi.org/10.1053/j.gastro.2010.01.054>.
- Kanojia D., Morshed R.A., Zhang L., Miska J.M., Qiao J., Kim J.W., ... Ahmed A.U. β III-tubulin regulates breast cancer metastases to the brain. *Mol. Cancer Ther.* 2015;14(5):1152-1161. <http://doi.org/10.1158/1535-7163.MCT-14-0950>.
- Kasowski M., Grubert F., Heffelfinger C., Hariharan M., Asabere A., Waszak S.M., ... Snyder M. Variation in transcription factor binding among humans. *Science.* 2010;328(5975):232-235. <http://doi.org/10.1126/science.1183621>.
- Kaz A.M., Brentnall T.A. Genetic testing for colon cancer. *Nat. Clin. Pract. Gastroenterol. Hepatol.* 2006;3(12):670-679. <http://doi.org/10.1038/ncpgasthep0663>.
- Kohn R.A., Mulvihill M.M., Chang J.W., Hsu K.-L., Sorrentino A., Cravatt B.F., ... Nomura D.K. Activity-based protein profiling of oncogene-driven changes in metabolism reveals broad dysregulation of PAFAH1B2 and 1B3 in cancer. *ACS Chem. Biol.* 2015;10(7):1624-1630. <http://doi.org/10.1021/acscembio.5b00053>.
- Laitinen V.H., Rantaperä T., Fischer D., Vuorinen E.M., Tammela T.L.J., PRACTICAL Consortium, ... Schleutker J. Fine-mapping the 2q37 and 17q11.2-q22 loci for novel genes and sequence variants associated with a genetic predisposition to prostate cancer. *Int. J. Cancer.* 2015;136(10):2316-2327. <http://doi.org/10.1002/ijc.29276>.
- Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 2012;9(4):357-359. <http://doi.org/10.1038/nmeth.1923>.
- Lappalainen T. Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Res.* 2015;25(10):1427-1431. <http://doi.org/10.1101/gr.190983.115>.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England).* 2009;25(16):2078-2079. <http://doi.org/10.1093/bioinformatics/btp352>.
- Li Q., Yokoshi M., Okada H., Kawahara Y. The cleavage pattern of TDP-43 determines its rate of clearance and cytotoxicity. *Nat. Commun.* 2015;6:6183. <http://doi.org/10.1038/ncomms7183>.
- Madan E., Gogna R., Kuppusamy P., Bhatt M., Mahdi A.A., Pati U. SCO2 induces p53-mediated apoptosis by Thr845 phosphorylation of ASK-1 and dissociation of the ASK-1-Trx complex. *Mol. Cell. Biol.* 2013;33(7):1285-1302. <http://doi.org/10.1128/MCB.06798-11>.
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorf L.A., Hunter D.J., ... Visscher P.M. Finding the missing heritability of complex diseases. *Nature.* 2009; 461(7265):747-753. <http://doi.org/10.1038/nature08494>.
- Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E., Wang H., ... Stamatoiyannopoulos J.A. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190-1195. <http://doi.org/10.1126/science.1222794>.
- Ng S.B., Buckingham K.J., Lee C., Bigham A.W., Tabor H.K., Dent K.M., ... Bamshad M.J. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* 2010;42(1):30-35. <http://doi.org/10.1038/ng.499>.
- Ni Y., Weber Hall A., Battenhouse A., Iyer V.R. Simultaneous SNP identification and assessment of allele-specific bias from ChIP-seq data. *BMC Gen.* 2012;13(1):46. <http://doi.org/10.1186/1471-2156-13-46>.
- Pan Z., Chen S., Pan X., Wang Z., Han H., Zheng W., ... Shao R. Differential gene expression identified in Uigur women cervical squamous cell carcinoma by suppression subtractive hybridization. *Neoplasma.* 2010;57(2):123-128. <http://www.ncbi.nlm.nih.gov/pubmed/20099975>.
- Reddy T.E., Gertz J., Pauli F., Kucera K.S., Varley K.E., Newberry K.M., ... Myers R.M. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res.* 2012;22(5):860-869. <http://doi.org/10.1101/gr.131201.111>.
- Rozowsky J., Abyzov A., Wang J., Alves P., Raha D., Harmanici A., ... Gerstein M. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* 2014;7(1):522-522. <http://doi.org/10.1038/msb.2011.54>.
- Sahu A.D., Aniba R., Chang Y.-P.C., Hannenhalli S. Epigenomic model of cardiac enhancers with application to genome wide association studies. *Pacific Symposium on Biocomputing.* 2013;22(5):92-102. <http://doi.org/23424115>.
- Solban N., Jia H.P., Richard S., Tremblay S., Devlin A.M., Peng J., ... Tremblay J. HcARG, a novel calcium-regulated gene coding for a nuclear

- protein, is potentially involved in the regulation of cell proliferation. *J. Biol. Chem.* 2000;275(41):32234-32243. <http://doi.org/10.1074/jbc.M001352200>.
- Sun S., Lee D., Ho A.S.W., Pu J.K.S., Zhang X.Q., Lee N.P., ... Leung G.K.K. Inhibition of prolyl 4-hydroxylase, beta polypeptide (P4HB) attenuates temozolomide resistance in malignant glioma via the endoplasmic reticulum stress response (ERSR) pathways. *Neuro-Oncology*. 2013;15(5):562-577. <http://doi.org/10.1093/neuonc/not005>.
- Tagliabue E., Fagnoli M.C., Gandini S., Maisonneuve P., Liu F., Kayser M., ... M-SKIP study group. MC1R gene variants and non-melanoma skin cancer: a pooled-analysis from the M-SKIP project. *British J. Cancer*. 2015;113(2):354-363. <http://doi.org/10.1038/bjc.2015.231>.
- Taylor D.P., Burt R.W., Williams M.S., Haug P.J., Cannon-Albright L.A. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology*. 2010;138(3):877-885. <http://doi.org/10.1053/j.gastro.2009.11.044>.
- The Cancer Genome Atlas Network. 2013. <http://cancergenome.nih.gov>.
- Trynka G., Westra H.-J., Slowikowski K., Hu X., Xu H., Stranger B.E., ... Raychaudhuri S. Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. *Am. J. Hum. Genet.* 2015;97(1):139-152. <http://doi.org/10.1016/j.ajhg.2015.05.016>.
- Wang Y.-P., Huang L.-Y., Sun W.-M., Zhang Z.-Z., Fang J.-Z., Wei B.-F., ... Han Z.-G. Insulin receptor tyrosine kinase substrate activates EGFR/ERK signalling pathway and promotes cell proliferation of hepatocellular carcinoma. *Cancer Lett.* 2013;337(1):96-106. <http://doi.org/10.1016/j.canlet.2013.05.019>.
- Welter D., MacArthur J., Morales J., Burdett T., Hall P., Junkins H., ... Parkinson H. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucl. Acids Res.* 2014;42(D1):D1001-D1006. <http://doi.org/10.1093/nar/gkt1229>.
- Weren R.D.A., Ligtenberg M.J.L., Kets C.M., de Voer R.M., Verwiel E.T.P., Spruijt L., ... Hoogerbrugge N. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat. Genet.* 2015;47(6):668-671. <http://doi.org/10.1038/ng.3287>.
- Wong J.C.Y., Gokgoz N., Alon N., Andrulis I.L., Buchwald M. Cloning and mutation analysis of ZFP276 as a candidate tumor suppressor in breast cancer. *J. Hum. Genet.* 2003;48(12):668-671. <http://doi.org/10.1007/s10038-003-0088-1>.
- Yang C.-Y., Lu R.-H., Lin C.-H., Jen C.-H., Tung C.-Y., Yang S.-H., ... Lin C.-H. Single nucleotide polymorphisms associated with colorectal cancer susceptibility and loss of heterozygosity in a Taiwanese population. *PloS ONE*. 2014;9(6):e100060. <http://doi.org/10.1371/journal.pone.0100060>.
- Yu J.-H., Zhong X.-Y., Zhang W.-G., Wang Z.-D., Dong Q., Tai S., ... Cui Y.-F. CDK10 functions as a tumor suppressor gene and regulates survivability of biliary tract cancer cells. *Oncol. Rep.* 2012;27(4):1266-1276. <http://doi.org/10.3892/or.2011.1617>.
- Zhang Z., Jia Q., Zhou C., Xie W. Crystal structure of E. coli endonuclease V, an essential enzyme for deamination repair. *Sci. Rep.* 2015;5:12754. <http://doi.org/10.1038/srep12754>.
- Zhao Y., Clark W.T., Mort M., Cooper D.N., Radivojac P., Mooney S.D. Prediction of functional regulatory SNPs in monogenic and complex disease. *Hum. Mutat.* 2011;32(10):1183-1190. <http://doi.org/10.1002/humu.21559>.

Эволюция CpG-островов путем тандемных дупликаций

В.Н. Бабенко^{1,2}✉, Ю.Л. Орлов^{1,2}, Ж.Т. Исакова³, Д.А. Антонов⁴, М.И. Воевода^{1,5}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Научно-исследовательский институт молекулярной биологии и медицины при Национальном центре кардиологии и терапии им. Мирсаида Миррахимова при Министерстве здравоохранения Кыргызской Республики, Бишкек, Кыргызская Республика

⁴ Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт экспериментальной и клинической медицины», Новосибирск, Россия

⁵ Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт терапии и профилактической медицины» Сибирского отделения Российской академии медицинских наук, Новосибирск, Россия

CpG-богатые острова (далее CpG-острова; CpG-islands, или CGI) являются важными функциональными элементами генома позвоночных. В частности, они инициируют транскрипцию как промоторы в большинстве (>50 %) генов позвоночных, в ряде случаев двунаправленную, в силу самокомплементарности сг-динуклеотидов; формируют глобальный ландшафт метилирования; выступают как «выключатель» транскрипции через метилирование. Вырожденная природа CpG-островов (смещенный сг-состав) предполагает увеличение вероятности тандемных повторов и палиндромов внутри CpG-острова. Данная работа посвящена идентификации тандемных дупликаций полных CpG-островов, т.е. мегамономеров длиной 400–5000 п.н., в геноме человека. Были найдены меж- и внутригенные тандемные дупликации CpG-островов. Найденные межгенные CGI дупликации опосредовались через CG-богатые субцентромерные и теломерные сателлиты, а также SINE элементы. Высокое сходство мономеров тандемных в отдельных случаях предполагает существование давления отбора на структуру таких локусов. Исследование контекста межгенных тандемных CGI повторов указывает на их возможную роль в выравнивании сг-состава в геномном сегменте. Найденные тандемные CGI были транскрипционно активными в широком диапазоне тканей и клеточных линий. Рассмотренный феномен кластерной организации CGI наиболее выражен при рассмотрении хромосомы 19, известной своим избытком сегментных дупликаций и генных экспансий. К уникальным геномным сегментам относится также мегасателлит DXZ4 на q плече хромосомы X, который также попадает в категорию CpG-островов, порожденных тандемными дупликациями.

Ключевые слова: CpG-острова; сг-состав; метилирование; эпигенетика, генные дупликации; хромосома 19.

Evolution of CpG-islands by means of tandem duplications

V.N. Babenko^{1,2}✉, Y.L. Orlov^{1,2}, Zh.T. Isakova³,
D.A. Antonov⁴, M.I. Voevoda^{1,5}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Molecular Biology and Medicine at National Center of Cardiology and Therapy named after academician Mirsaid Mirrahimov under the Ministry of Health of the Kyrgyz Republic, Bishkek, Kyrgyzstan

⁴ Institute of Clinical and Experimental Medicine SB RAS, Novosibirsk, Russia

⁵ Research Institute of Internal and Preventive Medicine SB RAMS, Novosibirsk, Russia

CpG-rich islands (CpG-islands, or CGI) are important functional elements in a genome of vertebrates. In particular, they: a) initiate transcription as promoters in most (> 50 %) genes of vertebrates, in some cases bi-directional, due to self-complement feature of cg dinucleotides; b) form a global methylation landscape; c) act as a transcription “switch” via methylation. The degenerate nature of CpG-island (elevated CG composition) implies an increase in the probability of tandem repeats and palindromes within CpG-island. This work is devoted to the identification of tandem duplications of complete CpG-islands, i.e. considering mega monomers of size 400–5000 bp, in the human genome. We found a range of inter- and intragenic tandem duplications of CpG-islands. Intergenic CpGi duplication mediates through CG-rich telomeric satellites, as well as elements of the SINE. One of the most pronounced tandems are located in chromosome 19, known for its abundance of segment duplications and gene expansion. We also underline the unique genomic segment, which is DXZ4 mega satellite, in q arm of chromosome X, also falling into the category of CpG-islands which evolved by tandem duplications rounds.

Key words: CpG-islands; tandem repeats; macro-satellites; human genome; methylation; CpG-island clusters.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Бабенко В.Н., Орлов Ю.Л., Исакова Ж.Т., Антонов Д.А., Воевода М.И. Эволюция CpG-островов путем тандемных дупликаций. Вавиловский журнал генетики и селекции. 2016;20(6):804-815. DOI 10.18699/VJ16.198

HOW TO CITE THIS ARTICLE:

Babenko V.N., Orlov Y.L., Isakova Zh.T., Antonov D.A., Voevoda M.I. Evolution of CpG-islands by means of tandem duplications. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6): 804-815. DOI 10.18699/VJ16.198

ORIGINAL ARTICLE

Received 19.05.2015

Accepted for publication 28.12.2015

© AUTHORS, 2016

Одна из характеристик геномов позвоночных – наличие множества так называемых CpG-островов, характеризующихся избытком CG динуклеотидов (Gardiner-Garden, Frommer, 1987). Формальное определение CpG-островов включает ряд параметров: процентное содержание cg-состава более 60 %; отношение $cg_obs/cg_exp > 0.6$; длина острова > 300 п. н. Упомянутое отношение видоспецифично, в частности у мыши плотность cg динуклеотидов в CGI меньше, чем у человека (Illingworth, Bird, 2009). В геноме человека их число составляет примерно 26 тыс. (www.genome.ucsc.edu).

CpG-острова перекрываются с промоторной частью 50–70 % генов. Такие промоторы называют CpG промоторами (Babenko et al., 1999; Deaton, Bird, 2011). В генах гомеобоксов (HOX, PAX), а также примерно в 5 % других случаях CpG-островов промоторов CGI являются тканеспецифичной мишенью для метилтрансфераз: при метилировании промотора ген репрессируется.

CpG сайты в составе CpG-острова часто гомогенны по метилированию. Поэтому в качестве проб для платформы Illumina methyl 450 в ряде случаев берутся только один-два сайта острова (Sandoval et al., 2011).

Недавно было показано, что для репрессии гена *TTP* в клетках рака печени достаточно метилирования лишь одного CpG динуклеотида, находящегося на границе (берегу – shore) CpG-острова (Sohn et al., 2010). Возможно, такой мощный эффект метилирования одного сайта обусловлен разрушением пространственной структуры хроматина путем ликвидации сайта связывания инсулятора *ctcf* (Wang et al., 2012; Kang et al., 2015). Не исключена также ситуация, что состояние сайта CpG на границе CpG-острова связано со статусом метилирования острова в силу кооперативного метилирования (Haerter et al., 2014; Grandi et al., 2015).

Остальные 30–40 % островов, не связанных с промоторами генов, находятся во внутригенных районах (20 %) и в межгенных областях (12–15 %) (Deaton, Bird, 2011). Метилирование внутригенного CpG-острова часто усиливает экспрессию гена, предположительно за счет исключения образования спонтанного иницирующего Pol II-комплекса (Deaton, Bird, 2011), а также выраженного отсутствия нуклеосом из-за слабой гибкости ДНК (Vinogradov, 2005; Fenouil et al., 2012). При этом наблюдается градиент корреляции метилирования и экспрессии с максимальным положительным значением от 3' конца гена.

В настоящее время известны семейства протяженных (более 10 т. п. н.) tandemных cg-богатых межгенных повторов человека, которые были названы мега- или макросателлитами (Chadwick, 2008; Tremblay et al., 2010). Одновременно они являются tandemно повторенными CpG-островами, поэтому мы остановимся на них более подробно. Мегасателлиты имеют специфическую номенклатуру, которая состоит из буквы D (duplication) или RS (repeated sequence) и последующего индекса хромосомы, после чего обычно идет буква Z. Они включают в себя DXZ4 на Xq23-24 (Giacalone et al., 1992; Horakova et al., 2012) и D4Z4 на 4q35/10q24 (Hewitt et al., 1994; Chadwick, 2008; Das, Chadwick, 2016). В базу данных RepeatMasker они не включены, поскольку не являются диспергированными и часто аннотированы как «простые повторы».

Макросателлиты характеризуются двумя общими признаками: высоким cg-составом и значительной (несколько т. п. н.) длиной мономера. Сами мономеры, как правило, не имеют межсателлитной гомологии, а также проявляют индивидуальную вариацию числа копий мономеров (Horakova et al., 2012), что используется как сигнатура популяции по аналогии с микросателлитами (Tremblay et al., 2010).

При наличии ряда межгенных tandemных CGI повторов большинство найденных в работе CGI повторов связаны с рекуррентной tandemной дупликацией генов умеренной длины, каковыми являются, в частности, гены *KRAB-ZNF*, и их промоторов (Lukic et al., 2014). Как правило, число мономеров в таких tandemах немногочисленно, но в некоторых случаях (кластеры некодирующих РНК) может достигать до 30–40.

В контексте генных дупликаций следует отдельно сказать о макросателлите (до недавнего времени генов в них не обнаруживалось) RS447 (4q17), не относящемся к CG-богатым сателлитам, но являющемся классической рекуррентной генной дупликацией. Его мономер содержит ген убиквитин-специфичной пептидазы 17 (11 генов USP17L). Небольшой кластер USP17L (2 гена) расположен также на хромосоме 8p23.1 (chr8:11.974.676-12.006.269). Гомология между мономерами макросателлита RS447 составляет более 99 %, поэтому различить отдельные копии современными методами (кроме клонирования) не представляется возможным (Treangen, Salzberg, 2012). Именно поэтому кластер, хотя он активно экспрессируется в ряде клеток, не аннотирован по состоянию хроматина и экспрессии: NGS секвенирование не может быть однозначно картировано. Это же касается и других мегасателлитов с высокогомологичными мономерами.

Задачей данной работы были поиск и анализ регулярных локальных повторов CpG-островов, представляющих как tandemные, так и диспергированные повторы.

Материалы и методы

Набор из 26 412 CpG-островов взят нами из таблицы *cpGIslandExt* (www.genome.ucsc.edu; version hg19). Список аннотированных повторов в зоне tandemных регулярных CpG-островов заимствован из базы данных (трека) результатов работы RepeatMasker (Smit, Hubley, 2008).

Оценка достоверности кластеризации CGI. Для поиска кластеров CGI геном был разбит на 10 т. п. н. сегменты (243 785). Частота попадания в сегмент острова (λ) равна $26412/243785 = 0.1$. Оценка достоверно неслучайного числа островов в сегменте выполнялась на основе распределения Пуассона

$$P(X=k) = \exp(-\lambda) \frac{\lambda^k}{k!}.$$

Вычисление вероятности P значения проводилось по формуле

$$P(X > k) = 1 - \sum_{n=0}^k P(X = n).$$

При ожидаемом числе CGI в сегменте $\lambda = 0.1$ интегральные вероятности (P -значения) следующие:

$P(X > 0) = 4.7E-3$; $P(X > 1) = 1.6E-4$; $P(X > 2) = 3.8E-6$; $P(X > 3) = 7.7E-8$.

Таким образом, для четырех и более CGI в сегменте вероятность составляет, с учетом поправки на множе-

ственные сравнения, порядка 1Е–3, т. е. может считаться достоверным порогом.

Результаты

При рассмотрении распределения длины CGI в исходной выборке выяснилось, что медианная длина островов составляет 1 т. п. н. В этой связи нами выдвинуто предположение, что длинные CGI являются tandemным расширением островов мономеров. В частности, в исходной выборке CGI наблюдалось 16 островов длиннее 10 т. п. н. Они рассмотрены в отдельном разделе в конце статьи.

По остальным CGI был произведен поиск их кластеров по сегментам, согласно описанной методике. Всего выявлено 25 кластеров с общим числом CGI 141. Наибольшее число кластеров находилось на 19, 4, 2, 7 и 10 хромосомах (табл. 1).

Тандемные кластеры CpG-островов на хромосомах 4 и 19, опосредованные диспергированными повторами,

относятся к макросателлитам и включают 64 острова. Их характеристика приведена далее. Ген-ориентированные CpG-острова (141 – 64 = 77) рассмотрены ниже.

Большинство кластеров CpG-островов расположено в районах генов гомеобокса (развития). Найденные 25 кластеров с числом CpG-островов более 3 на 10 т. п. н. были проаннотированы вручную (табл. 2). Обнаружено, что большинство кластеров, содержащих в общей сложности 141 CpG-остров, относятся к генам гомеобокса (*HOXa,c*) и нервного развития (*PAX2–5*). Число островов, расположенных в районах этих генов, составляет 45. Заметим, что ряд генов гомеобокса находятся в единичных, но протяженных (более 2 т. п. н.) CpG-островах, которые мы не рассматривали по методическим причинам. Например, ген *NKX6-2* на хромосоме 10 расположен внутри CpG-острова длиной более 3 т. п. н., который метилирован в большинстве тканей. О наличии множества CpG-островов в районах генов гомеобокса говорилось и ранее

Table 1. Prevalence of CGI clusters and CGIs on chromosomes, clusters having more than three monomers per 10 kb being taken into consideration

Numbers of	Chromosome													
	2	4	5	6	7	9	10	11	12	14	15	17	19	Total
CGI	13	20	9	4	13	8	12	8	4	4	6	8	32	141
Clusters	3	2	2	1	1	2	3	2	1	1	1	2	4	25

Table 2. Genes associated with CGI clusters

Gene class /location	Gene	Chromosome	Number of CGIs	Position
Homeobox	ALX4	chr11	4	44 325 657
Promoter	Cyp26A1_C1	chr10	4	94 825 546
Homeobox	DLX1	chr2	5	1.73E+08
Homeobox	EN1	chr2	4	1.2E+08
Promoter	FAM89 (и AS)	chr11	4	65 337 213
Intragenic	HMHA1	chr19	4	1 066 866
Homeobox	HOXa	chr7	13	27 146 069
Homeobox	HOXc12	chr12	4	54 348 695
Membrane peptide	HS3ST3B1	chr17	4	14 200 579
Homeobox	MSX1	chr4	4	4 858 389
Myosin (promoter)	MYO1C	chr17	4	1 388 686
Homeobox	NKX2-1	chr14	4	36 986 362
Homeobox	NKX2-3	chr10	4	1.01E+08
Homeobox	NKX2-5	chr5	5	1.73E+08
Homeobox	NRN1	chr6	4	5 996 185
Homeobox (optical)	PAX2	chr10	4	1.02E+08
Homeobox (fetus)	PAX3	chr2	4	2.23E+08
Homeobox (CNS)	PAX5	chr9	4	37 025 492
Neurospecific gene (*protocadherin)	PCDHGA1	chr5	4	1.41E+08
Intragenic	PTB1	chr19	4	807 131

Rows with nonhomeobox genes are filled yellow.

Table 3. Locations of CGI tandem segments with > 10 monomers

Chromosome	Location (Mb)	Number of monomers	Gene	Annotated repeats	Segment features	PMID ID
19	36.7	34	ZNF	SST1	Promoter	21078170
4	132.6	17	–	SST1	–	21078170
X	115	45–112	–	Simple	DXZ4	Chadwick, 2008
X	120	12	CT47A	AluS	1 coding exon, 12 genes	16382448
X	3	11	PPP2R3B	AluSx1	Single gene	11173861
Y	2.5	11	PPP2R3B	AluSx1	Single gene	11173861

(Branciamore et al., 2010). В ряде злокачественных опухолей наблюдается гипометилирование данных генов. Такая этиология (гипометилирование *HOX*, *PAX* генов с последующей злокачественной пролиферацией) возникает, в частности, из-за мутаций метилтрансферазы DNMT3A (Qu et al., 2014).

Кроме пяти генов, кодирующих избыточные белки «домашнего хозяйства» и содержащих кластеры CpG-островов в промоторной области, такие как фактор сплайсинга PTB1, миозин MYO1C, гепарансульфат (мембранный белок) HS3ST3B1, транскрипционный фактор FAM89B, ген гистосовместимости *HMHA1*, цитохром P Cyp26A1_C1 (см. табл. 2). Стоит выделить отдельно ген протокадерина, содержащего сг-богатые кассеты экзонов и опосредующего специфическую адгезию нейронов.

Ген протокадерина (*PCDH*). На хромосоме 5 находится locus гена протокадерина (*PCDH*; см. табл. 2). Locus содержит 22 (аннотированных) транскрипта, разделенных на три подсемейства. Подсемейство А содержит 12 генов, подсемейство В – 7 генов и 2 псевдогена, а более отдаленное семейство G – 3 гена. Тандемный массив 22 CGI-содержащих экзонов, составляющих варибельную часть, завершается константной, общей для всех подсемейств частью, включающей в себя три экзона. Постоянная часть гена кодирует цитоплазматическую часть. В каждом варибельном экзоне имеется внеклеточный домен, состоящий из пяти кадериновых эктодоменов и трансмембранного района. Предполагается, что данные белки клеточной адгезии играют критическую роль в установлении и функционировании межклеточных связей в мозге.

Интересно, что благодаря CpG-островам каждый экзон варибельного кластера также содержит промоторную часть для генов семейства. Эволюция этого гена происходила путем дупликации сегмента из двух соседних экзонов. Длина такого мономера превышала 7 т. п. н., длина каждого экзона в нем достигала 2400–2600 п. н. Подробнее механизм генерации транскриптов в генах этого уникального семейства рассмотрен в работах (Ong, Corces, 2014; Guo et al., 2015).

Сателлитные тандемы CGI, опосредованные аннотированными повторами. В отдельную группу были выделены повторы, имеющиеся в составе мономеров повторы, аннотированные в базе данных RepBase (Smit, Hubley, 2008–2015) (табл. 3).

Table 4. Distribution of SST1 repeats over chromosomes according to RepBase

Chromosome	Number of SST1 monomers	Mean length
chr12	10	710.3
chr16	15	522.7
chr17	27	805.9
chr19	56	1228.5
chr20	49	775.7
chr4	19	1519.9
chr6	3	345.0
chr7	35	757.6
chr9	50	885.9
chrY	321	295.5

Rows for macrosatellites on chromosomes 4 (4SST1) and 9 (19SST11, 19SST12) are filled with yellow.

Повторы, опосредованные SST1. Название умеренного SST1 происходит от рестрикционного энзима Sst1, гидролизующего мономер в одном сайте и приводящего к наблюдению спектра фрагментов, кратных 2.5 т. п. н., при разрезании тандемных кластеров этих повторов. Данный энзим имеет гомологию с аденовирусом (Epstein et al., 1987). Повтор аннотирован в RepBase как центромерный, является G/C богатым (g/c состав 67 %) и имеет наибольшую длину на хромосомах 4 и 19 (табл. 4).

Распределение числа SST1 повторов по хромосомам оценено нами с использованием аннотации RepeatMasker (Smit et al., 2008). Как можно видеть, геном человека содержит 590 SST1 повторов, большинство из которых являются «старыми» и имеют малую длину (см. табл. 4). Множество их следов находится на хромосоме Y.

Два «целевых» кластера макросателлитов SST1 в хромосомах 19 и 4, содержащих в своем мономере CpG-остров, были описаны ранее (Tremblay et al., 2010). Мономеры макросателлитов 4SST1 и 19SST11, 19SST12 на 4 и 19 хромосомах соответственно (Tremblay et al., 2010) являются трехсегментными и включают в себя мотив

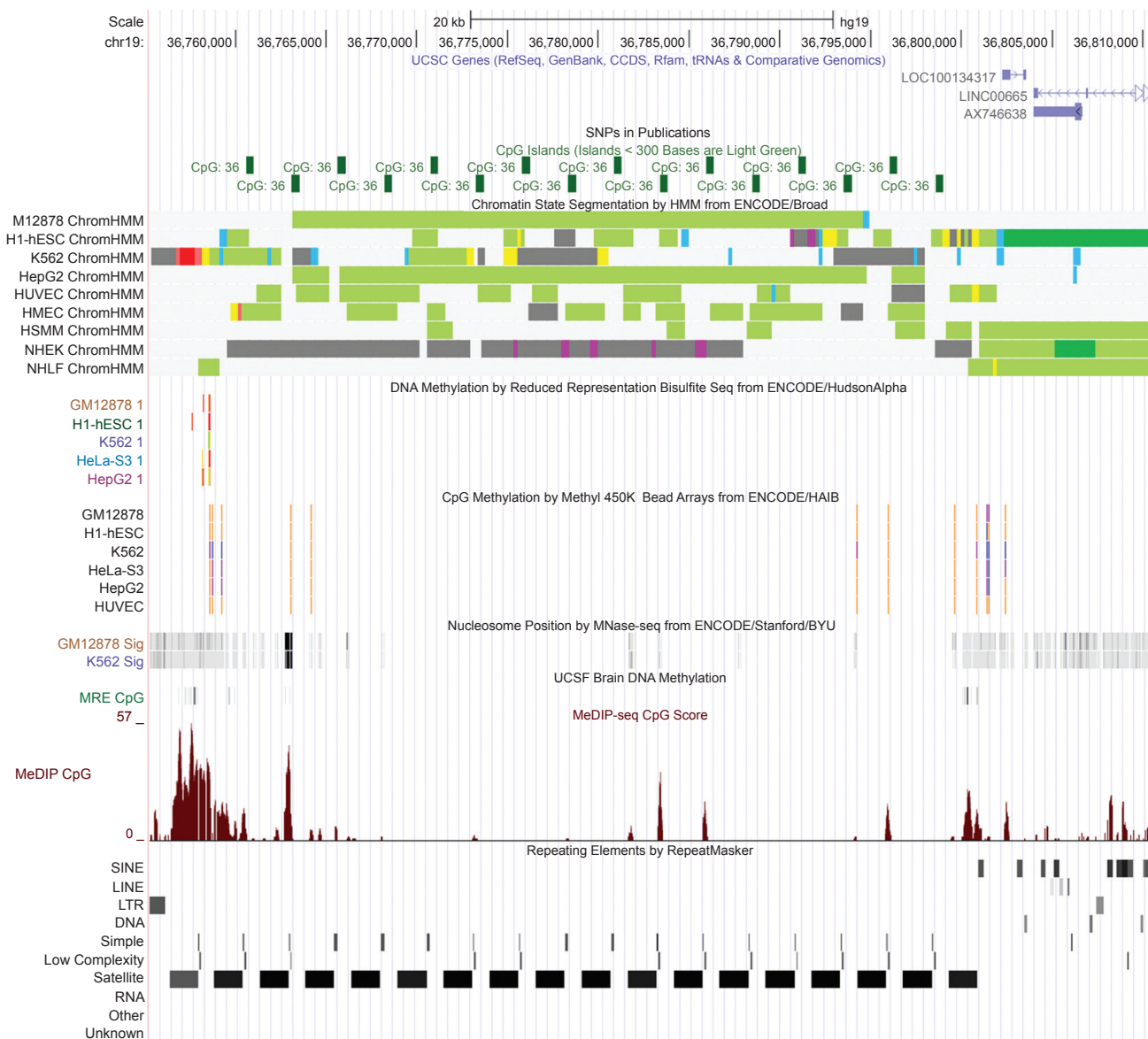


Fig. 1. Macrosatellite 19SST11 comprising monomers SST1_SR_CGI (chr19:36.755.278-36.810.578).

Three segments constituting the monomer are apparent from the RepeatMasker track (below) and green-colored CGI track (above): Satellite (SST1); Simple repeat (SR); CGI.

SST1, простой повтор (трек RepeatMasker) и CpG-остров (зеленый трек CpG-islands): SST1_SR_CGI (рис. 1). Длина данного мономера составляет 2.4 т. п. н.

Выявлено, что при достаточно неравномерной (не тандемной) общей кластеризации SST1 на хромосомах существуют три выраженных высокогомологических тандема этих повторов: два – на хромосоме 19 и один – на хромосоме 4. От остальных SST1 кластеров они отличаются также тем, что находятся далеко от центromеры (19q12.12, 4q28.4) и содержат в составе мономера CpG-остров. В большинстве своем острова неметилированы, и, возможно, в силу этого три рассмотренных тандема повторов экспрессируются во всех тканях (Tremblay et al., 2010).

Два кластера тандемных высокогомологических повторов со структурой мономера расположены в районе 19q13.12

(рис. 1 и 3). Мономер повтора состоял из SST1, спейсера (300 п. н.), содержащего в начале несколько простых коротких повторов, и CpG-острова длиной 400 п. н.: (SST1, spacer, CGI). Кластеры были ориентированы противонаправленно.

Важно отметить, что и SST1, и CpG-острова ведут себя аналогично по отношению к нуклеосомам, а именно, имеют ярко выраженную «деплецию» нуклеосом (Vinogradov, 2005). При чередовании этих двух повторов (см. рис. 1 и 2) вся область тандемов оказалась свободной от нуклеосом (трек Nucleosome position by MNase-seq).

На основе выравнивания мономеров мы вычислили внутритандемную и межатандемную гомологию для трех рассмотренных тандемов (табл. 5). Из приведенных в таблице данных видна высокая гомология сателлитов, что указывает на возможную эволюционную поддержку

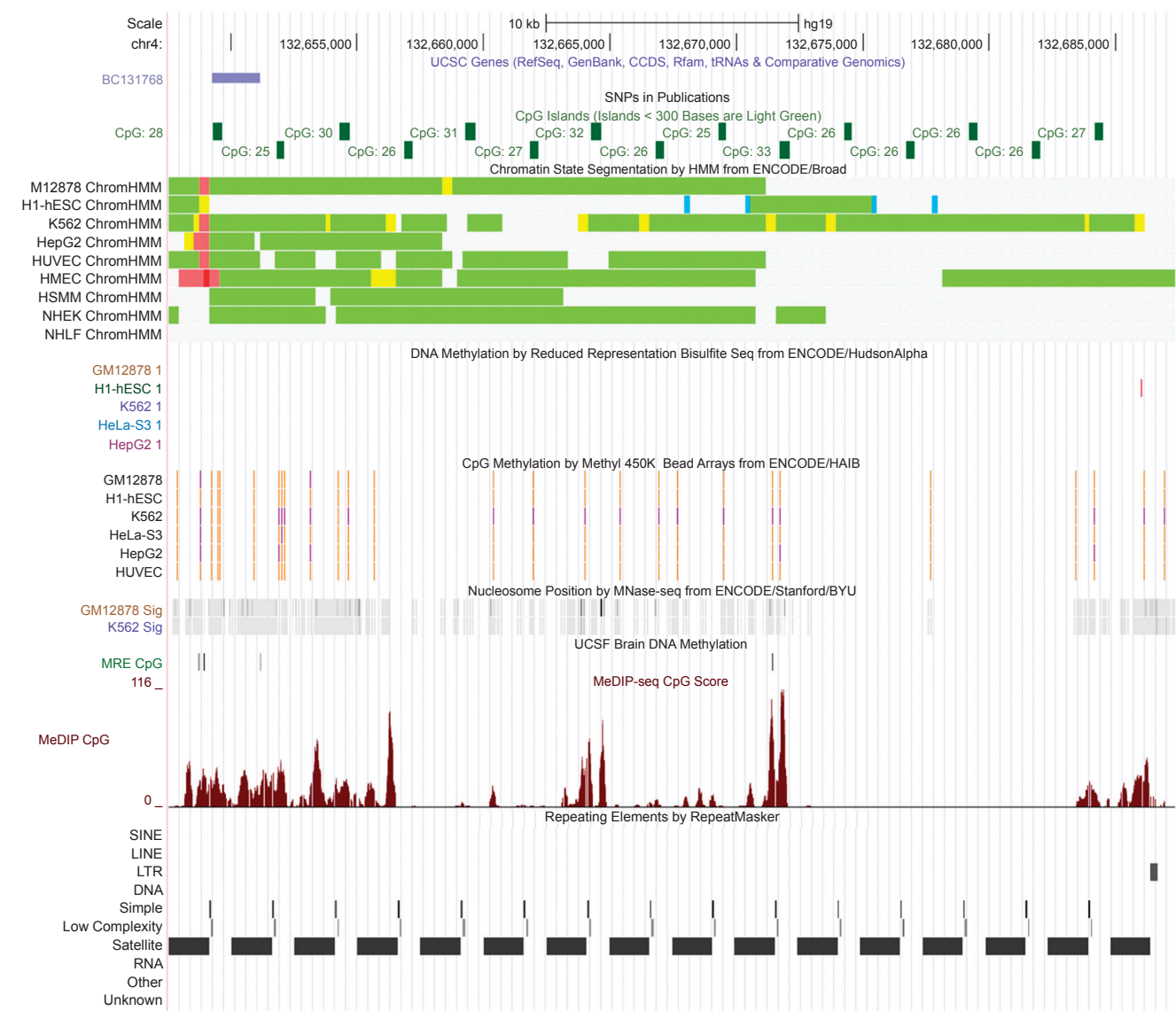


Fig. 2. Macrosatellite 4SST1 comprising 16 tandems of monomer SST1_CGI in the 4q28.3 region with high similarity of the monomers to those of macrosatellites 19SST11 and 19SST12 (see Table 4).

Table 5. Inter- and intra-tandem similarities calculated as the percentage of identity in blast pairwise alignment for three clusters: 19SST11, 19SST12, and 4SST1

Macrosatellite	19SST11	19SST12	4SST1
19SST11	99.8	97.1	93.4
19SST12		98.7	93.4
4SST1			94.6

консервативности этих сателлитов, наблюдающуюся в значительном числе тандемно повторяющихся треков (Warburton et al., 2008).

CpG промоторы 17 генов цинкового пальца (ZNF) между тандемами порождены SST1_CpGi мономерами. Схема района, насыщенная 56 SST1_CGI мономерами и состоящая из двух макросателлитов 19SST11 и 19SST12, а также кластера генов KRAB-ZNF, может быть пред-

ставлена в виде распределения длин соседних мономеров (рис. 3).

Сверхдлинные CGI. Рассмотрены 16 CpG-островов длиной более 10 т.п.н., выделенные в отдельный класс (см. также (Warburton et al., 2008)). Как видно из аннотации (табл. 6), сверхдлинные острова, за рядом исключений, порождены тандемными повторами и во многих случаях проаннотированы как макросателлиты, поскольку

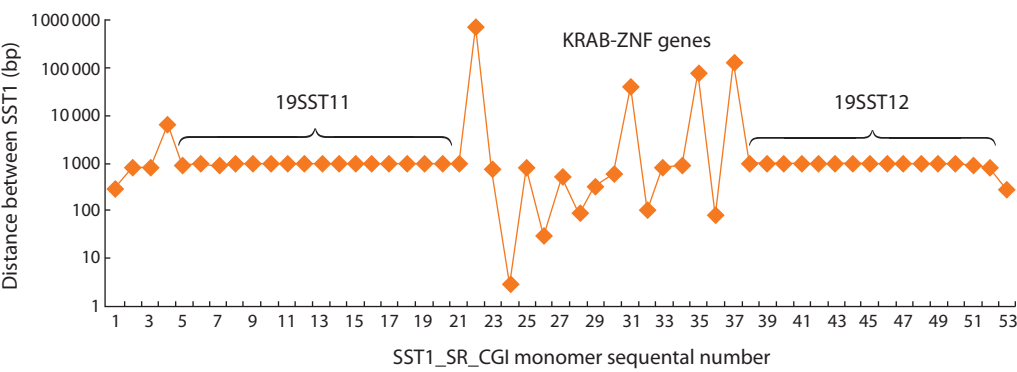


Fig. 3. Distribution of intermonomer distances between SST1 repeats in chromosome 19 macrosatellite doublet 19SST11-19SST12 (Chadwick, 2008).
The intersatellite region comprises 17 KRAB-ZNF genes whose promoters contain SST1_RS_CGI monomers (Table 7).

Table 6. Sixteen ultralong (> 10 kb) CGIs

#	Annotated repeats	Repeat name	Gene	Chromosome	Length (kb)	Hypometh
1	Simple	–	tRNA	chr1	32 361	0
2	Simple	–	miRNA	chr1	40 058	1
3	Simple	–	ZNF (antisense)	chr10	14 472	1
4	–	–	Promoter of EBF3	chr10	10 527	1
5	Telomere	LSAU	DUX cluster	chr10	20 640	0
6	–	–	piRNA cluster	chr15	12 438	0
7	–	–	TBX (AS)	chr17	10 206	1
8	–	–	NKX2 homeobox 2	chr20	10 782	1
9	–	–	–	chr4	13 414	0
10	Telomere	LSAU	DUX cluster (D4Z4)	chr4	27 227	0
11	–	–	FOXC1	chr6	11 260	1
12	–	–	UNC homeobox (UNCX), mRNA	chr7	13 699	1
13	–	–	Promoter of TNRC18	chr7	10 753	1
14	–	–	Plectin (PLEC)	chr8	11 865	0
15	–	–	Promoter of BCOR	chrX	15 813	1
16	DXZ4	–	–	chrX	45 712	0

–, no data.

картировать гены в них, в силу повторенности в начальной версии генома, не представлялось технически возможным (Treangen, Salzberg, 2012). Впоследствии гены были найдены, и макросателлиты были переаннотированы в генные кластеры – например, D4Z4 (DUX cluster). Они отличаются от разделенных кластеров CGI, рассмотренных выше, тем, что сливаются в один CpG-остров, хотя и имеют тандемную структуру.

После ручной аннотации (см. табл. 6) мы разделили 16 островов на следующие категории по типам генов: 1) кластеры некодирующих РНК (тРНК, миРНК, пиРНК – tRNA, miRNA, piRNA), найденные на хромосомах 1

(miRNA 40 т. п. н.; tRNA 32 т. п. н., см. табл. 6) и хромосоме 15 (piRNA, 12 т. п. н.). В последнем случае, несмотря на очевидную тандемную структуру, признаки наличия простых повторов отсутствуют, вероятно, вследствие высокой дивергенции;

2) два кластера гомеобокс-генов DUX (хр. 10, 20.6 т. п. н.) и DUX (D4Z4; хр. 4, 27.2 т. п. н.) в теломерах, перемежающиеся теломерными хромосом-специфичными сг-богатыми повторами LSAU в обоих случаях. Заметим, что по этому теломерному макросателлиту (D4Z4) наблюдалась внутривидовая транслокация между сегментами хромосом 4q → 10q (Zhang et al.,

- 2005; Thijssen et al., 2014), благодаря значительной гомологии мономеров тандемов, как и в случае SST1 повторов (см. табл. 5);
- 3) гены гомеобокса FOXC1 (хр. 6, 11.2 т. п. н.), NKX2 (хр. 20, 10.7 т. п. н.) и UNCX (хр. 7, 13.7 т. п. н.). В этих случаях CGI образованы уникальными последовательностями, их размеры лишь слегка превышают 10 т. п. н., а гены лежат внутри CGI. Вне эмбриональных клеток эти районы гетерохроматизированы PcG комплексом. Упомянутые выше кластеры DUX также являются генами гомеобоксов, с последующей гетерохроматизацией или метилированием;
 - 4) острова являются сверхдлинными промоторами генов EBF3 (хр. 10, 10.5 т. п. н.), TNRC18 (хр. 7, 10.7 т. п. н.), BCOR (хр. X, 15.8 т. п. н.) и в этом случае имеют структуру без повторов;
 - 5) случаи бинаправленных промоторов генов цинковых пальцев и их антисмысловых партнеров: ZNF (хр. 10, 14.5 т. п. н.), TBX (хр. 17, 10.2 т. п. н.); имеются признаки тандемной структуры CGI;
 - 6) оставшиеся неклассифицированные случаи включают вышеуказанный DXZ4 (хр. X, 45 т. п. н. в референсном геноме) и ген плектина PLEC (хр. 8, 11.8 т. п. н.), у которого протяженный остров находится внутри тела гена (см. табл. 6).

Достаточно часто наблюдалась нетандемная структура сверхдлинных CpG-островов, однако и в этих случаях не исключена гипотеза тандемного механизма возникновения и последующего вырождения: возможно, данные протяженные сг-богатые районы являются следствием древних дупликаций или резко повышенной скорости эволюции для приобретения новых свойств у паралогов.

В каждом случае, особенно для кластеров tRNA, miRNA, piRNA, каждый остров имеет свою историю и происхождение, а также отдельную функцию. Таким образом, несмотря на существующие закономерности тандемной структуры при возникновении сверхдлинных CGI, на этом этапе можно рассматривать данное явление как достаточно уникальный феномен.

В ряде случаев дупликация была, возможно, опосредована сателлитными повторами, как центромерными, так и теломерными, а также SINE Alu. Гены, входящие в сверхдлинные CGI, также являются гомеобоксными, или генами раннего развития. Их расположение в CGI островах обеспечивает возможности ткане-, стадио- и пол-специфичного выключения с помощью метилирования и последующего образования HP1 – гетерохроматина.

В контексте бинаправленных промоторов с антисмысловой lncRNA (см. табл. 6, строки 3 и 7) отметим, что при полногеномном сканировании в большом числе случаев наблюдается CGI промотор (Vavouri, Lehner, 2012) в силу его самокомплементарности, т. е. обогащения сг-состава одновременно в обеих нитях ДНК. Частое наличие инсультатора вблизи CGI промотора (Vavouri, Lehner, 2012) позволяет переключать транскрипцию с одной цепи на другую ткане- и стадиоспецифичным образом (Wood et al., 2013; Anderson, 2014). Следует упомянуть недавнюю статью об антиоксидантом ответе, где наблюдалась массовая экспрессия антисмысловых РНК (Giannakakis et al., 2015).

Обсуждение

Со времени открытия и формального определения CpG-островов (Gardiner-Garden, Frommer, 1987) выяснилось, что эти функциональные элементы играют крайне важную роль в функционировании геномов позвоночных (Deaton, Bird, 2011). В частности, большинство генов «домашнего хозяйства» имеют в промоторной области данный остров, распознаваемый множеством транскрипционных факторов (Sp1, CTCF и др.), и обеспечивают инициацию транскрипции соответствующих генов.

Позже выяснилось, что около 30 % CpG-островов не имеют отношения к промотору и расположены в меж- и внутригенных областях. Эволюционная консервативность данных элементов, показанная в ряде работ (Deaton, Bird, 2011), свидетельствует об их функциональной нагрузке, которая пока до конца не определена. Выдвигались гипотезы, что в целом CpG-острова связаны с глобальным ландшафтом метилирования, т. е. имеют эпигенетическое значение (Deaton, Bird, 2011). Не случайно CpG-острова часто встречаются в районах импринтированных генов (в частности, *PPA2C*, *DNMT1*, *TSHZ3*, *CHST8*, *ZNF225*⁺, *ZNF229*⁺, *DMWD*, *ZNF331*⁺, *LILRB4*, *NLRP2*, *ZIM2*⁺, *PEG3*⁺, *MIMT1*, *USP29*, *ZIM3*⁺, *ZNF264*⁺, *CHMP2A*, *MZF1*⁺), причем не только в промоторе, но и в их расширенном контексте (Deaton, Bird, 2011).

Режим эволюции CpG-островов путем дупликации не был рассмотрен ранее. Имеются факты, подтверждающие их адаптационное значение. В частности, на хромосоме 19 произошло достоверно большее число генных и сегментных дупликаций по сравнению с остальными хромосомами (Grimwood et al., 2004). При этом гены, содержащие активный CpG промотор, сохраняли его и при дупликациях. Тем не менее, кроме промоторных CpG, хромосома 19 содержит значительное число 3'UTR и внутригенных CpG-островов, которые в подавляющем числе случаев метилированы, что связано с усилением экспрессии.

Использование протяженных районов сильно смещенного нуклеотидного состава, каковыми являются CpG-острова, в процессе эволюции могло происходить по причинам: а) открытости района, практически полностью свободного от нуклеосом и, как правило, транскрибируемого в широком диапазоне тканей; б) использования сг-богатого состава для дупликации вследствие неравного кроссингвера; в) возможности выключения генов путем метилирования и последующей гетерохроматизации промотора или всего гена.

В нашей работе мы выявили тандемные дупликации CpG-островов в геноме человека. Про сг-богатые тандемные повторы в некодирующих областях сообщалось достаточно давно в контексте полиморфных сателлитов (Schaap et al., 2013). Дело в том, что эти макросателлиты полиморфны по длине, т. е. встречаются в геномах с переменным числом в силу повышенной митотической рекомбинации (Schaap et al., 2013). Это свойство позволяет использовать их как популяционно-специфический маркер. Наибольшая вариация сг-богатых тандемов отмечается у африканцев, наименьшая – у азиатов (Schaap et al., 2013). CpG-острова используются в популяционной генетике и имеют следующие названия (Schaap et al., 2013): RS447

Table 7. Genes and CGIs located between macrosatellites 19SST11–19SST12

Cpg_id	CpG island	Location	Gene
Macrosatellite 19SST11			
9361	CpG: 56	36 869 564	ZFP14
9362	CpG: 66	36 909 281	ZFP82
9363	CpG: 74	36 912 260	LOC644189
9364	CpG: 44	36 980 190	ZNF566
9365	CpG: 66	37 018 919	ZNF260
9366	CpG: 69	37 063 892	ZNF529
9367	CpG: 86	37 095 680	ZNF382
9367	CpG: 86	37 095 680	ZNF529
9368	CpG: 39	37 157 632	ZNF461
9370	CpG: 53	37 263 381	ZNF850
9372	CpG: 40	37 288 342	LOC284408
9373	CpG: 49	37 328 896	ZNF790
9374	CpG: 26	37 340 918	ZNF345
9374	CpG: 26	37 340 918	ZNF790
9375	CpG: 48	37 406 931	ZNF568
9375	CpG: 48	37 406 931	ZNF829
9377	CpG: 56	37 568 952	ZNF420
Macrosatellite 19SST12*			
9394	CpG: 45	37 825 101	HKR1
9395	CpG: 57	37 861 691	ZNF527
9397	CpG: 57	37 957 726	ZNF569
9398	CpG: 64	37 959 852	ZNF570
9399	CpG: 26	37 997 790	ZNF793
9400	CpG: 66	38 039 561	LOC100507433
9401	CpG: 56	38 085 148	ZNF540
9401	CpG: 56	38 085 148	ZNF571
9402	CpG: 119	38 145 826	ZFP30
9403	CpG: 47	38 182 793	ZNF781
9404	CpG: 50	38 210 107	ZNF607
9405	CpG: 26	38 270 79	ZNF573

* Immediately after the tandem pair following macrosatellite 19SST12. For continuation of imperfect tandem duplications, see Fig. 3.

(chromosome 4p non CpGi), MSR5p (5p), FLJ40296 (13q), RNU2 (17q) и D4Z4 (4q and 10q), а также X chromosomal DXZ4 и CT47.

Структурные свойства района макросателлитов 19SST11 и 19SST12 и генного кластера между ними. Наибольшими по размеру найденными tandemными кластерами CGI являются макросателлиты 19SST11 и 19SST12 на хромосоме 19 (см. табл. 4). Их длина в общей сложности составляет около 1 млн п. н. Между ними и сразу за ними находятся кластеры генов KRAB-ZNF (табл. 7, см. рис. 3), имеющих в качестве промотора мономер фланкирующих сателлитов. Существует гипотеза о возможности гомологичной неравной рекомбинации на

основе двух рассмотренных tandemных инвертированных кластеров 19SST11–19SST12 в районе 19q13.12, которая развернула сегмент 1 млн п. н. между повторами (Tremblay et al., 2010). Высокое сходство tandemов, расположенных в обратной ориентации, делает такую гипотезу правдоподобной. Это также подтверждается локальной мозаичной синтенией с позвоночными, описанной в работе (Grimwood et al., 2004).

Полногеномный анализ SST1 повтора показал, что он транскрипционно активен (в большинстве тканей) на всех хромосомах, кроме облигатно репрессированного сегмента хромосомы Y, где расположены короткие следы повтора (см. табл. 4). Области локусов, включающие в себя SST1,

в большинстве своем не содержат CpG-островов, за исключением рассмотренных трех tandemных кластеров на хромосоме 19 и 4, а также случаев SST1_CpG промоторов генов KRAB-ZNF, расположенных между tandemами на хромосоме 19 (см. рис. 3 и табл. 7).

Tandemные кластеры SST1_CGI хромосомы 19 находятся в районе гетерохроматина B4 (Rao et al., 2014), избыливающего кластерами KRAB-ZNF генов (см. табл. 7). Данные гены имеют отношение к метилированию ретровирусов в эмбриогенезе (Lukic et al., 2014) и реплицируются в ранней G-фазе (Rao et al., 2014). Соответственно, они репрессированы гетерохроматином HP1, который характерен для доменов B4 (Rao et al., 2014). Интродукция кластеров SST1_CGI «открывает» хроматин, освобождая его от нуклеосом. Об этом также свидетельствует базальная транскрипция (см. рис. 1 и 2, зеленые линии в треке HMM from ENCODE для клеточных линий), наблюдаемая в данных районах как на хромосоме 19, так и на хромосоме 4 во всех тканях (Tremblay et al., 2010). Тем не менее в большинстве зрелых тканей эти районы закрыты HP1 хроматином.

Наличие сайтов связывания CTCF в районах кластеров CGI. На рис. 1 и 2 видно также избирательное метилирование определенных CGI (коричневый трек MeDIP CpG) и наличие CTCF сайтов связывания (см. рис. 1, голубые полосы (Das, Chadwick, 2016)). В литературе имеются ссылки на связь tandemных повторов CGI, состояния хроматина, а также *ctcf* инсуляторов, которые в совокупности модулируют трехмерную конформацию хроматина, в частности на инактивированной X хромосоме (Rao et al., 2014). Избирательное выпетливание, опосредованное *ctcf* инсуляторами, приводит также к множественному альтернативному сплайсингу CGI-релевантных первых экзонов гена протокадерина PCDHA, рассмотренного выше. Таким образом, данная «структурная аномалия» на хромосоме 19, вполне возможно, связана с пространственной координацией хроматина (Rao et al., 2014; Das, Chadwick, 2016).

Кроме этого, инсуляторный функциональный потенциал, представленный многочисленными сайтами связывания *ctcf* фактора в CGI tandemах, используется и в случае макросателлита D4Z4 (Zhang et al., 2005; Ottaviani et al., 2010; Darrow, Chadwick, 2014; Das, Chadwick, 2016), а также в генах протокадерина для осуществления альтернативного сплайсинга (см. табл. 2) (Nichols, Corces, 2015), в кластере tPHK (см. табл. 7) (Darrow, Chadwick, 2014), кластере piPHK (piwiRNA) (см. табл. 7) (Williams et al., 2015), а также в сцепленных с болезнями кластерах двойных гомеобоксах DUX4 (см. табл. 7) (Das, Chadwick, 2016). Это подчеркивает тесную функционально-регуляторную связь CpG-островов и *ctcf* инсуляторов как в промоторах (Vavouri, Lehner, 2012; Kang et al., 2015), так и в некодирующих tandemных кластерах, таких как DXZ4 (Ottaviani et al., 2010; Horakova et al., 2012; Wang et al., 2012; Darrow, Chadwick, 2014; Rao et al., 2014). Как правило, все сверхдлинные CGI (см. табл. 7) содержат сайты связывания инсуляторов CTCF.

При оценке внутри- и межтандемной гомологии (см. табл. 3) выяснилось, что CpG-острова имеют крайне высокую гомологию, что говорит об их возможной функциональной значимости. В силу этих же причин tandem-

мы мономеров SST1_CGI могут выступать в качестве рестрикционных макросателлитных маркеров (Tremblay et al., 2010), хотя их использование затруднено наличием по крайней мере трех высокомолекулярных tandemных повторов на двух хромосомах, что снижает их специфичность.

DXZ4: феномен сверхдлинного CGI. Макросателлит DXZ4 (45–120 т.п.н., см. табл. 6, строка 16) составляет единый CpG-остров. Момеры CGI (DXZ4) содержат три структурных элемента, все из которых CG-богатые (Horakova et al., 2012). DXZ4 является консервативным среди приматов и, по всей видимости, кодирует длинную некодирующую РНК, хотя, возможно, происходит просто базальная транскрипция CpG богатого неметилированного сегмента. Его длина сильно варьирует среди популяций по числу tandemов: от 12 до 99, среднее число мономеров 54 (Horakova et al., 2012). Таким образом, DXZ4 – также популяционный маркер.

Tandemный массив DXZ4 (Chadwick, 2008; Horakova et al., 2012; Rao et al., 2014) разделяет инактивированную хромосому X на два пространственных супердомена: (0–115 млн п.н.) (DXZ4) и (115–153.3 млн п.н.) (Rao et al., 2014). Инициация эффекта «ламповых щеток» при инактивировании X хромосомы происходит эухроматизацией DXZ4 путем деметилирования, в отличие от гиперметилированного состояния на активной X хромосоме. На инактивированной X хромосоме с сегментом DXZ4 связывается ряд CTCF факторов, позволяя обеспечить до 27 больших суперпетель размером от 7 до 74 млн п.н. (Rao et al., 2014). Организация суперпетель идет подавляющим образом с помощью *ctcf* факторов: его содержат 23 из 24 районов основания петель (Rao et al., 2014). Четыре района основания суперпетель содержат длинные некодирующие РНК (lncRNA), а именно XIST, lnc550643, DXZ4 и FIRRE. Интересно, что два первых являются cg-богатыми и гиперметилированы в активной X хромосоме, что предотвращает связывание *ctcf* факторов (Wang et al., 2012).

Феномен инактивации X хромосомы, а также результаты исследования взаимодействия метилирования и сайтов связывания *ctcf* факторов (Wang et al., 2012; Kang et al., 2015) выявляют антагонистическое взаимодействие *ctcf* фактора и метилирования в cg-богатых районах, каковыми являются CGI.

Заключение

Сосредоточившись в основном на аннотации районов и выявлении общих характеристик, для их поиска мы использовали метод кластеризации CGI, что является грубым приближением при поиске tandemных дупликаций. Тем не менее в работе по выявлению длинных tandemных повторов с помощью специализированной программы Tandem Repeats Finder – TRF (Warburton et al., 2008) содержатся все основные макросателлиты/tandemные дупликации, которые мы нашли с учетом CGI-релевантности, за исключением несовершенных, как piPHK кластер.

Мы представили относительно немногочисленные, но выраженные случаи множественной дупликации CpG-островов, а также сверхдлинных CGI (см. табл. 2 и 6). Выяснилось, что основные аннотированные повторы,

опосредующие консервативную дупликацию, относятся к типам SINE (AluS на половых хромосомах) и центромерным повторам SST1 на хромосомах 4 и 19. Кроме этого, существуют неаннотированные «простые» сг-богатые повторы, которые создали уникальный якорный CGI остров DXZ4 на хромосоме X, участвующий в ее инактивации, а также D4Z4 (см. табл. 2 и 6).

В работе показано, что большинство плотных кластеров CGI, как отдельных (см. табл. 2), так и слитных (см. табл. 6), порождено тандемными дупликациями. Эти кластеры связаны с генами раннего развития, включая протяженный кластер piPHK (piwiRNA) (Williams et al., 2015), а также KRAB-ZNF гены, расположенные на хромосоме 19 в районе B4 хроматина.

Acknowledgments

This work was supported by the Russian Science Foundation, project 14-24-00123. The authors are grateful to the SB RAS supercomputing center (<http://www2.sccc.ru>) for access to computational resources.

Conflict of interest

The authors declare no conflict of interest.

References

- Anderson S.K. Probabilistic bidirectional promoter switches: noncoding RNA takes control. *Mol. Ther. Nucl. Acids*. 2014;3:e191.
- Babenko V.N., Kosarev P.S., Vishnevsky O.V., Levitsky V.G., Basin V.V., Frolov A.S. Investigating extended regulatory regions of genomic DNA sequences. *Bioinformatics*. 1999;15(7-8):644-653.
- Branciamore S., Chen Z.X., Riggs A.D., Rodin S.N. CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. *Proc. Natl. Acad. Sci. USA*. 2010;107(35):15485-15490.
- Chadwick B.P. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Gen. Res*. 2008;18:1259-1269. DOI 10.1101/gr.075713.107.
- Darrow E.M., Chadwick B.P. A novel tRNA variable number tandem repeat at human chromosome 1q23.3 is implicated as a boundary element based on conservation of a CTCF motif in mouse. *Nucl. Acids Res*. 2014;42(10):6421-6435.
- Das S., Chadwick B.P. Influence of repressive histone and DNA methylation upon D4Z4 transcription in non-myogenic cells. *PLoS One*. 2016;11(7):e0160022. DOI 10.1371/journal.pone.0160022.
- Deaton A.M., Bird A. CpG islands and the regulation of transcription. *Gen. Dev*. 2011;25(10):1010-1022.
- Epstein N.D., Karlsson S., O'Brien S., Modi W., Moulton A., Nienhuis A.W. A new moderately repetitive DNA sequence family of novel organization. *Nucl. Acids Res*. 1987;15:2327-2341.
- Fenouil R., Cauchy P., Koch F., Descostes N., Cabeza J.Z., Innocenti C., Ferrier P., Spicuglia S., Gut M., Gut I., Andrau J.C. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res*. 2012;22(12):2399-2408.
- Gardiner-Garden M., Frommer M. CpG islands in vertebrate genomes. *J. Mol. Biol*. 1987;196:261-282.
- Giacalone J., Friedes J., Francke U. A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat. Genet*. 1992;1(2):137-43.
- Giannakakis A., Zhang J., Jenjaroenpun P., Nama S., Zainolabidin N., Aau M.Y., Yarmishyn A.A., Vaz C., Ivshina A.V., Grinchuk O.V., Voorhoeve M., Vardy L.A., Sampath P., Kuznetsov V.A., Kurochkin I.V., Guccione E. Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. *Sci. Rep*. 2015;5:9737. DOI 10.1038/srep09737.
- Grandi F.C., Rosser J.M., Newkirk S.J., Yin J., Jiang X., Xing Z., Whitmore L., Bashir S., Ivics Z., Izsvák Z., Ye P., Yu Y.E., An W. Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. *Genome Res*. 2015;25(8):1135-1146.
- Grimwood J., Gordon L.A., Olsen A., Terry A., Schmutz J., Lamerdin J., ... Stubbs L., Rokhsar D.S., Myers R.M., Rubin E.M., Lucas S.M. The DNA sequence and biology of human chromosome 19. *Nature*. 2004;428(6982):529-535.
- Guo Y., Xu Q., Canzio D., Shou J., Li J., Gorkin D.U., Jung I., Wu H., Zhai Y., Tang Y., Lu Y., Wu Y., Jia Z., Li W., Zhang M.Q., Ren B., Krainer A.R., Maniatis T., Wu Q. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*. 2015;162(4):900-910. DOI 10.1016/j.cell.2015.07.038.
- Haerter J.O., Lövkqvist C., Dodd I.B., Sneppen K. Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucl. Acids Res*. 2014;42(4):2235-2244.
- Hewitt J.E., Lyle R., Clark L.N., Valleley E.M., Wright T.J., Wijmenga C., van Deutekom J.C.T., Francis F., Sharpe P.T., Hofker M., Frants R.R., Williamson R. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet*. 1994;3(8):1287-1295. DOI 10.1093/hmg/3.8.1287.
- Horakova A.H., Moseley S.C., McLaughlin C.R., Tremblay D.C., Chadwick B.P. The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum. Mol. Genet*. 2012;21(20):4367-4377.
- Illingworth R.S., Bird A.P. CpG islands – ‘a rough guide’. *FEBS Lett*. 2009;583(11):1713-20. DOI 10.1016/j.febslet.2009.04.012.
- Kang J.Y., Song S.H., Yun J., Jeon M.S., Kim H.P., Han S.W., Kim T.Y. Disruption of CTCF/cohesin-mediated high-order chromatin structures by DNA methylation downregulates PTGS2 expression. *Oncogene*. 2015;34:5677-5684. DOI 10.1038/nc.2015.17.
- Lukic S., Nicolas J.C., Levine A.J. The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ*. 2014;21(3):381-387. DOI 10.1038/cdd.2013.150.
- Nichols M.H., Corces V.G. A CTCF code for 3D genome architecture. *Cell*. 2015;162(4):703-705. DOI 10.1016/j.cell.2015.07.053.
- Ong C.T., Corces V.G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet*. 2014;15(4):234-246.
- Ottaviani A., Schluth-Bolard C., Gilson E., Magdinier F. D4Z4 as a prototype of CTCF and lamins-dependent insulator in human cells. *Nucleus*. 2010;1(1):30-36.
- Qu Y., Lennartsson A., Gaidzik V.I., Deneberg S., Karimi M., Bengtzen S., Höglund M., Bullinger L., Döhner K., Lehmann S. Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics*. 2014;9(8):1108-1119.
- Rao S.S., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., Aiden E.L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665-1680.
- Sandoval J., Heyn H., Moran S., Serra-Musach J., Pujana M.A., Bibikova M., Esteller M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*. 2011;6(6):692-702.
- Schaap M., Lemmers R.J., Maassen R., van der Vliet P.J., Hoogerheide L.F., van Dijk H.K., Baştürk N., de Knijff P., van der Maarel S.M. Genome-wide analysis of macrosatellite repeat copy number variation in worldwide populations: evidence for differences and commonalities in size distributions and size restrictions. *BMC Genomics*. 2013;14(14):143. DOI 10.1186/1471-2164-14-143.
- Smit A.F.A., Hubley R. RepeatModeler Open-1.0. 2008–2015. <http://www.repeatmasker.org>.

- Sohn B.H., Park I.Y., Lee J.J., Yang S.J., Jang Y.J., Park K.C., Kim D.J., Lee D.C., Sohn H.A., Kim T.W., Yoo H.S., Choi J.Y., Bae Y.S., Yeom Y.I. Functional switching of transforming growth factor-beta1 signaling in liver cancer via epigenetic modulation of a single CpG site in tristetraproline promoter. *Gastroenterol.* 2010;138:1898-1908.
- Thijssen P.E., Balog J., Yao Z., Pham T.P., Tawil R., Tapscott S.J., van der Maarel S.M. DUX4 promotes transcription of FRG2 by directly activating its promoter in facioscapulohumeral muscular dystrophy. *Skeletal Muscle.* 2014;4:19. DOI 10.1186/2044-5040-4-19.
- Treangen T.J., Salzberg S.L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 2012;13:36-46.
- Tremblay D.C., Alexander G., Jr., Moseley S., Chadwick B.P. Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. *BMC Genomics.* 2010;15(11):632. DOI 10.1186/1471-2164-11-632.
- Vavouri T., Lehner B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol.* 2012;13(11):R110.
- Vinogradov A.E. Noncoding DNA, isochores and gene expression: nucleosome formation potential. *Nucl. Acids Res.* 2005;33(2):559-563.
- Wang H., Maurano M.T., Qu H., Varley K.E., Gertz J., Pauli F., Lee K., Canfield T., Weaver M., Sandstrom R., Thurman R.E., Kaul R., Myers R.M., Stamatoyannopoulos J.A. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Gen. Res.* 2012;22(9):1680-1688.
- Warburton P.E., Hasson D., Guillem F., Lescale C., Jin X., Abrusan G. Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics.* 2008;9:533. DOI 10.1186/1471-2164-9-533.
- Williams Z., Morozov P., Mihailovic A., Lin C., Puvvula P.K., Juraneck S., Rosenwaks Z., Tuschl T. Discovery and characterization of piRNAs in the human fetal ovary. *Cell Rep.* 2015;13(4):854-863.
- Wood E.J., Chin-Inman K., Jia H., Lipovich L. Sense-antisense gene pairs: sequence, transcription, and structure are not conserved between human and mouse. *Front. Genet.* 2013;4:183.
- Zhang Y.Z., Sun S.C., Wu H.C., Fan Q.S., Song Y.J., Yu W., Jeanpierre M., Urtizberea J.A. Polymorphism of the D4Z4 locus associated with facioscapulohumeral muscular dystrophy 1A in Shanghai population. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi (=Chinese J. Med. Genetics).* 2005;22(4):380-382.

Комбинирование и оценка конгруэнтности филогенетических сигналов от нескольких генов с помощью геометрического подхода

В.М. Ефимов^{1, 2, 3}, В.Ю. Ковалева⁴, Ю.Н. Литвинов⁴

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет», Томск, Россия

⁴ Федеральное государственное бюджетное учреждение науки Институт систематики и экологии животных Сибирского отделения Российской академии наук, Новосибирск, Россия

Рассмотрены возможности недавно предложенного нами нового алгоритма (DJ-метода) для анализа конгруэнтности и комбинирования молекулярно-генетических данных на основе матриц евклидовых расстояний. Этот подход назван геометрическим, поскольку евклидовы дистанции удовлетворяют аксиомам метрики, что обеспечивает возможность помещения множества точек, представляющих последовательности, в некоторое геометрическое пространство без искажения взаимных расстояний и наделяния точек этого множества координатами в пространстве. Геометричность евклидовых расстояний позволяет применять к молекулярным данным весь арсенал методов многомерного анализа, что является актуальным для исследования соотношения внутри- и межвидовой изменчивости, вычисления центроидов таксонов и расстояний между ними, визуализации возможных направлений эволюции, комбинирования и оценки конгруэнтности филогенетических сигналов, относящихся к разным генам и даже к разным системам признаков. DJ-метод использован для оценки конгруэнтности и комбинирования филогенетических сигналов, получаемых от нескольких генов. Анализировались более 1500 нуклеотидных последовательностей двух ядерных (*apoB*, *brca1*) и двух митохондриальных (*co1*, *cytb*) генов 15 палеарктических и неарктических видов землероек-бурозубок рода *Sorex* (Soricidae, Eulipotyphla). Для каждого гена все его последовательности представлялись множеством точек в евклидовом пространстве. Для множества точек, относящихся к одному виду, вычислялся его центроид в этом же пространстве, а для каждого гена – матрица евклидовых расстояний между центроидами видов. Для оценки попарного сходства (конгруэнтности) матриц межвидовых расстояний применен тест Мантеля. Конгруэнтность генов яДНК составила 0.961, мтДНК – 0.748. Все матрицы межвидовых расстояний через взвешивание объединены в единую матрицу, по которой методом главных координат для всех видов построено единое пространство. В объединенном генетическом пространстве проявилось несколько направлений межвидовой изменчивости, отражающих разные по масштабу эволюционные события. По объединенной матрице межвидовых расстояний построено дерево, которое хорошо согласуется с принятой на сегодня зоологической систематикой. Это подтверждает работоспособность предложенного метода.

Ключевые слова: *Sorex*; мтДНК; яДНК; DJ-метод; филогенетика; евклидово пространство.

Combining and congruence evaluation of phylogenetic signals from different genes based on geometric approach

V.M. Efimov^{1, 2, 3}, V.Yu. Kovaleva⁴, Yu.N. Litvinov⁴

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Tomsk State University, Tomsk, Russia

⁴ Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia

A new Euclidean distance based algorithm is used for analysis of congruence and combining molecular genetic data. This approach is called geometric, since Euclidean distance satisfies all metric axioms and the points representing the sequences can be placed in a geometric space without distorting the mutual distances and can be endowed with the coordinates in this space. Geometricity of Euclidean distances allows methods of multivariate analysis to be applied to molecular data, which is relevant to intra- and interspecies variability investigating, visualization of possible directions of evolution, combining data and evaluation of the congruence of phylogenetic signals. The algorithm is used for the analysis of more than 1500 nucleotide sequences of two nuclear (*apoB*, *brca1*) and two mitochondrial (*co1*, *cytb*) genes of 15 Palaearctic and Nearctic shrew species of the genus *Sorex* (Soricidae, Eulipotyphla). All sequences of each gene are represented as a set of points in Euclidean space. Centroids of a set of points belonging to the same species are calculated. The matrix of Euclidean distances between the species centroids is calculated for each gene. Mantel test is applied to estimate pairwise similarity (congruence) of interspecies distances matrices relating to different genes. nDNA gene congruence is 0.961, that of mtDNA is 0.748. All matrices of the interspecies distances are combined into a joint matrix by weighing. Joint genetic space for all species is built by principal coordinate method from the joint matrix. Several variability directions reflecting evolutionary events of different scale are visualized in a joint genetic space. In addition, the joint matrix of interspecies distances is used for building a

phylogenetic tree, which is consistent with the zoological systematics accepted for today. This confirms the efficiency of our proposed method.

Key words: *Sorex*; mtDNA; nuclear DNA; DJ-method; phylogenetics; Euclidean space.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Ефимов В.М., Ковалева В.Ю., Литвинов Ю.Н. Комбинирование и оценка конгруэнтности филогенетических сигналов от нескольких генов с помощью геометрического подхода. Вавиловский журнал генетики и селекции. 2016;20(6):816-822. DOI 10.18699/VJ16.153

HOW TO CITE THIS ARTICLE:

Efimov V.M., Kovaleva V.Yu., Litvinov Yu.N. Combining and congruence evaluation of phylogenetic signals from different genes based on geometric approach. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):816-822. DOI 10.18699/VJ16.153

Одна из целей молекулярной систематики – изучение филогенетических взаимоотношений некоторого набора таксонов по нуклеотидным последовательностям, относящимся к разным генам. Для достижения этой цели необходимо решить две самостоятельные и важные задачи: проанализировать конгруэнтность филогенетических сигналов от разных генов (Гречко и др., 2006; Гречко, 2013) и объединить информацию, относящуюся к разным генам.

В настоящее время для решения обеих задач применяются два подхода: конкатенация (метод суперматрицы) и консенсус (метод супердерева) (Delsuc et al., 2005; Gadagkar et al., 2005; Philippe et al., 2005; Jeffroy et al., 2006; Planet, 2006). При использовании первого подхода выровненные последовательности из нескольких генов стыкуются в одну, и филогенетический анализ выполняется на объединенных в общую матрицу последовательностях (Nylander, 2004; Wortley, Scotland, 2006; de Queiroz, Gatesy, 2007). При использовании второго отдельное дерево строится для каждого гена, а общее дерево оценивается на основе консенсуса (Bininda-Emonds et al., 2002; Burleigh et al., 2011). Оба подхода ориентированы на деревья как форму представления окончательного результата. Это обусловлено тем, что на сегодняшний день основными подходами в филогенетике являются методы максимальной парсимонии, максимального правдоподобия и другие статистические подходы, опирающиеся на модели эволюции непосредственно нуклеотидных последовательностей (Лукашов, 2009) и неизбежно приводящие к деревьям.

Однако, вопреки широко распространенному в молекулярной филогенетике мнению, деревья – не единственный способ представления филогенетических взаимоотношений. Они могут быть дополнены отображением взаимного расположения последовательностей в многомерном пространстве (геометрический подход) или даже объединены с ним (Cavalli-Sforza, Edwards, 1967; Klingenberg, Ekau, 1996; Scippa et al., 2008; Klingenberg, Gidaszewski, 2010; Ковалева и др., 2012; Polly et al., 2013).

Некоторые исследователи выделяют в отдельный подход дистанционные методы, основанные на матрицах расстояний, как первичных, посчитанных непосредственно по последовательностям, так и вторичных, полученных из деревьев для каждого гена (Criscuolo et al., 2006; Criscuolo, Michel, 2009). Это не вызывало бы возражений, если бы используемые дистанции были дистанциями в строгом математическом смысле этого слова. Внимательное рассмотрение показывает, что это не совсем так. По сути

дела, дистанциями, расстояниями и даже метриками в молекулярной филогенетике называются произвольные меры различия между последовательностями. Для нуклеотидных и аминокислотных последовательностей уже предложено довольно много разнообразных генетических расстояний и постоянно предлагаются новые (Criscuolo, Michel, 2009), однако при этом не имеют в виду метрические свойства этих расстояний и, тем более, анализ взаимного расположения таксонов в каком-либо геометрическом пространстве. В частности, не являются евклидовыми такие известные генетические расстояния, как p -дистанция (см. Доп. материалы 1)¹, расстояние Нея (Beaumont et al., 1998), Джукса–Кантора, двупараметрическое расстояние Кимуры, LogDet (Ефимов и др., 2013) и, по-видимому, многие другие. Соответственно, анализ конгруэнтности и объединение расстояний производятся без учета их метрических свойств и достаточно произвольно. Объединение расстояний реализуется просто суммированием или усреднением (Criscuolo, Michel, 2009), анализ конгруэнтности – при помощи теста Мантеля (Mantel, 1967; Mantel, Valand, 1970). На практике дистанционные методы, используемые сегодня в молекулярной филогенетике, в том числе и как начальный этап при построении эволюционных моделей, ограничены тем, что по матрице объединенных расстояний между последовательностями тем или иным алгоритмом все равно строится филогенетическое дерево. Фактически большинство существующих дистанционных методов – это разновидности методов суперматрицы и супердерева.

Конечно, было бы гораздо удобнее, если бы дистанции между последовательностями сразу являлись метрическими расстояниями, т. е. удовлетворяли аксиомам метрики: неотрицательности, симметричности, неравенству треугольника. Еще лучше, если бы они были евклидовыми расстояниями. Выполнение аксиом метрики обеспечивает возможность помещения множества точек, представляющих последовательности, в некоторое геометрическое пространство без искажения взаимных расстояний и наделяния точек этого множества координатами в этом пространстве (Havel et al., 1983; Gower, Legendre, 1986).

Для целей настоящей статьи будем называть геометрическим любое молекулярно-генетическое расстояние, удовлетворяющее аксиомам метрики, поскольку используемые в настоящее время в молекулярной филогенетике термины «расстояние», «дистанция», «метрика» не име-

¹ Дополнительные материалы см. в Приложении 2 по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2016-20/appx2.pdf>

ют достаточно определенного геометрического смысла. К геометрическим расстояниям относится известное генетическое расстояние Кавалли-Сфорца – Эдвардса (Cavalli-Sforza, Edwards, 1967). Оно выведено при предположении, что генетическая разница возникает главным образом за счет генетического дрейфа. Это расстояние является евклидовым и обладает хорошими статистическими свойствами. Авторы с самого начала имели в виду отображение эволюции в виде пучка расходящихся траекторий в некотором многомерном евклидовом пространстве и прекрасно понимали геометрический смысл предлагаемого ими расстояния. В нашей недавней работе введены два новых геометрических расстояния, которые служат евклидовыми аналогами расстояний Джукса – Кантора и Кимуры (Ефимов и др., 2013). Если генетические расстояния являются геометрическими, то это позволяет применять весь арсенал методов многомерного анализа для исследования соотношения внутри- и межвидовой изменчивости, вычисления центроидов таксонов и расстояний между ними, визуализации возможных направлений эволюции, комбинирования и оценки конгруэнтности филогенетических сигналов, относящихся к разным генам и даже к разным системам признаков (Ковалева и др., 2012, 2013; Ефимов и др., 2013). Обсуждение этих возможностей на примере ядерных и митохондриальных генов является целью настоящей работы.

Материалы и методы

Проанализированы взятые из GenBank нуклеотидные последовательности двух ядерных (*apoB*, *brca1*) и двух митохондриальных (*col*, *cytb*) генов, относящихся к 15 палеарктическим и неарктическим видам землероек рода *Sorex* (Soricidae, Eulipotyphla): подрода *Sorex* – *Sorex araneus*, *S. bedfordiae*, *S. caecutiens*, *S. daphaenodon*, *S. isodon*, *S. minutissimus*, *S. minutus*, *S. roboratus*, *S. tundrensis* и подрода *Otisorrex* – *S. cinereus*, *S. fumeus*, *S. haydeni*, *S. monticolus*, *S. palustris*, *S. trowbridgii*. В общей сложности обработано 1588 последовательностей (см. Доп. материалы 2).

Особенностью молекулярной филогенетики, осложняющей дальнейший анализ, является то, что ее цель – это установление филогенетических взаимоотношений между таксонами, а исходные нуклеотидные последовательности, как правило, относятся к особям, и их число для разных таксонов на практике сильно различается. По этой причине построение эволюционного дерева выполняется для последовательностей, и только потом по нему каким-либо образом строится таксономическое дерево. Из этого следует, что таксономическая принадлежность особей должна быть определена заранее. Исходной информацией служит множество нуклеотидных последовательностей, классифицированных по таксонам и генам. Для каждого сочетания «таксон – ген» имеется хотя бы одна последовательность. Последовательности по разным генам не обязательно соответствуют одним и тем же особям.

В работе (Ковалева и др., 2012) нами предложен новый алгоритм анализа соответствия и комбинирования молекулярно-генетических и морфологических данных на основе матриц расстояний между видами (DJ-метод). Для того чтобы получить межвидовые расстояния из ну-

клеотидных последовательностей, относящихся к особям, нам пришлось дополнительно для каждого вида конструировать модальную (консенсусную) последовательность из нуклеотидов, с максимальной частотой встречающихся в каждой позиции. Это решение оказалось не совсем удачным, и в следующей работе (Ковалева и др., 2013) мы попытались более последовательно применить геометрический подход.

Усовершенствование алгоритма сводится к следующему. Сначала для каждого гена строится матрица евклидовых расстояний между всеми последовательностями всех таксонов. Полученная матрица расстояний переводится в матрицу объект – признак методом главных координат (Torgerson, 1952; Gower, 1966). Это означает, что каждой последовательности ставится в соответствие некоторая точка в евклидовом пространстве, причем размерность этого пространства будет всегда меньше числа последовательностей. Для множества точек, относящихся к одному таксону, вычисляется его центроид, который и представляет данный таксон. Между центроидами таксонов вычисляется матрица евклидовых расстояний, которая переводится в матрицу объект – признак методом главных координат. Это значит, что теперь каждому таксону ставится в соответствие некоторая точка в некотором новом евклидовом пространстве, причем размерность этого пространства будет всегда меньше числа таксонов.

Дополнительным преимуществом такого подхода является возможность контроля над систематическим искажением результатов, возникающим вследствие разного числа анализируемых последовательностей для каждого таксона. Более того, легко решается и сопутствующая проблема, появившаяся из-за накопления в базах данных таксономически неправильно определенных или искаженных при секвенировании или выравнивании последовательностей (Dubey et al., 2009; Гречко, 2013). Поскольку при отображении последовательностей в евклидово пространство четко видна структура каждого таксона и соотношение внутри- и межтаксонной изменчивости, то возможно отфильтровать сомнительные последовательности от основного ядра каждого таксона.

Геометричность евклидовых расстояний между видами позволяет легко и однозначно решить задачу комбинирования филогенетических сигналов, относящихся к разным генам. Так как каждую матрицу евклидовых расстояний можно методом главных координат без потери информации перевести в матрицу объект – признак, то конкатенация этих матриц для разных генов позволяет получить единую координатную матрицу, по которой однозначно вычисляется объединенная матрица евклидовых расстояний. Очевидно, что та же самая объединенная матрица получится, если мы поэлементно просуммируем квадраты межвидовых расстояний в исходных матрицах и извлечем квадратный корень. При желании исходные матрицы можно брать с некоторыми неотрицательными весами (Cavalli-Sforza, Edwards, 1967).

В настоящей работе в качестве весов выбраны координаты первого собственного вектора (Abeysondera et al., 2014). Объединенная матрица расстояний между видами обработана методами Крускала (Kruskal, 1964) и UPGMA (Unweighted pair-group method using arithmetic averages).

Результаты

Сначала по каждому гену между всеми его последовательностями с помощью пакетов Jascobi 4 и Excel вычислены E_{PQ} -дистанции (при $\alpha = 1$) (Ефимов и др., 2013), являющиеся евклидовыми аналогами двупараметрического расстояния Кимуры (Kimura, 1980). Методом главных координат все последовательности представлены точками в многомерном евклидовом пространстве. На рис. 1 приведена конфигурация полных последовательностей гена *cytb* для трех видов бурозубок группы «*araneus*» на плоскости первых двух главных координат. Отчетливо видны как основные ядра видов, так и отклоняющиеся последовательности, которые были исключены из дальнейшей обработки (см. рис. 1).

Далее для каждого гена вычислены центроиды видов и матрица евклидовых расстояний между ними. Для оценки попарного сходства между полученными матрицами межвидовых расстояний применен тест Мантеля (см. таблицу).

Принято считать, что филогенетический сигнал всегда отображается в виде дерева. Однако несколько не хуже отражает филогенетический сигнал и матрица межвидовых расстояний, вычисленная по молекулярно-генетическим последовательностям. Естественно полагать, что чем дальше на дереве отстоит гипотетический ближайший общий предок от данной пары видов, тем больше должно быть расстояние между этой парой в матрице межвидовых расстояний. А корреляция между полученными отдельно по каждому гену матрицами межвидовых расстояний (тест Мантеля) является оценкой сходства (конгруэнтности) филогенетических сигналов.

При оценке сходства филогенетических сигналов, полученных по каждому отдельному гену, оказалось, что ядерные гены *apoB* и *brca1* фактически дублируют друг друга, тогда как митохондриальные – *col* и *cytb* – несут филогенетические сигналы, несколько отличающиеся как от ядерных генов, так и между собой, хотя корреляция между ними достаточно высока и достоверна ($p < 10^{-6}$) (см. таблицу).

В нашей работе в качестве весов выбраны координаты первого собственного вектора. Объединенная матрица расстояний между видами обработана методами Крускала и UPGMA (рис. 2 и 3).

Результаты обработки методом Крускала показали, что максимальная доля дисперсии (93.3 %) приходится на первую ось шкалирования. Вдоль этого направления четко обозначились различия между палеарктическими и неарктическими видами бурозубок, т.е. между подро-

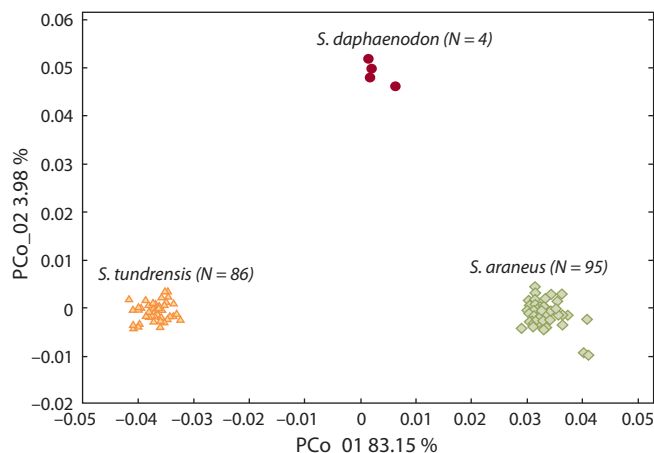


Fig. 1. Species clouds of the *araneus* group formed by nucleotide sequences of the *cytb* gene on the plane of the first two principal coordinates.

Three diverging *S. araneus* sequences are excluded from consideration.

дами *Sorex* и *Otisorex* (см. рис. 2). Кроме того, по второй оси шкалирования видна обособленность *S. bedfordiae* от палеарктических видов и *S. trowbridgii* от неарктических видов. При этом на долю второй и третьей осей шкалирования приходится 5.14 и 1.53 % дисперсии соответственно. Понятно, что на фоне таких соотношений долей дисперсии сложно анализировать внутривидовую структуру таксона, поэтому матрицы квадратов межвидовых расстояний обработаны с помощью SVD-метода (Singular value decomposition method) по отдельности для палеарктических и неарктических видов.

В подроде *Sorex* заметной структурированности не наблюдается (рис. 4). Более или менее уверенно можно говорить о группе «*araneus*»: *S. araneus*, *S. tundrensis*, *S. daphaenodon*. На дендрограмме она образует обособленную ветвь (см. рис. 3). Ранее эта группа видов была выделена на основании кариологических данных (Dannelid, 1991; Ivanitskaya, 1994). В настоящее время монофилию группы поддерживают данные секвенирования гена *cytb* мтДНК (Ohdachi et al., 1997, 2006; Fumagalli et al., 1999). Остальные виды располагаются в трехмерном пространстве во всех направлениях, не образуя скоплений.

В частности, не выделилась группа 42 хромосомных бурозубок «*caecutiens*» (*S. caecutiens*, *S. roboratus*, *S. isodon*). В работе (Fumagalli et al., 1999) на основании гена *cytb* мтДНК она выделена в несколько ином составе (*S. cae-*

Correlation coefficients between interspecies distance matrices of 15 *Sorex* species

Genes	<i>apoB</i>	<i>brca1</i>	<i>col</i>	<i>cytb</i>
<i>apoB</i>	1	0.961*	0.709*	0.759*
<i>brca1</i>	0.961*	1	0.684*	0.759*
<i>col</i>	0.709*	0.684*	1	0.748*
<i>cytb</i>	0.759*	0.759*	0.748*	1

Note. The matrices were compared by the Mantel test. The number of permutations $Np \geq 10^6$.

* $p < 10^{-6}$.

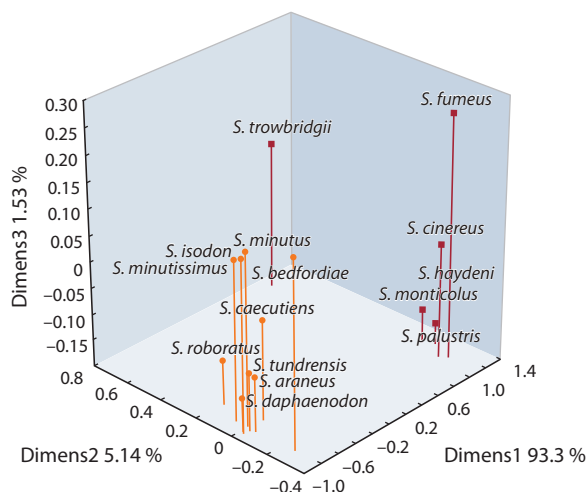


Fig. 2. Three-dimensional arrangement of shrew species by Kruskal's scaling of the joint matrix of genetic distances between centroids of 15 *Sorex* species with regard to two nuclear (*apoB* and *brca1*) and two mitochondrial (*co1* and *cytb*) genes.

Here and in Figs. 4 and 5 DIMENS1, 2, 3 are multidimensional scaling axes.

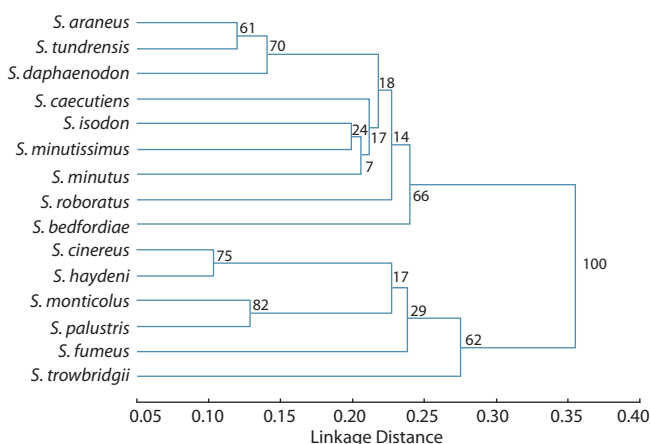


Fig. 3. Dendrogram constructed by the UPGMA algorithm from the joint matrix of genetic distances between centroids of 15 *Sorex* species.

Cophenetic correlation coefficient 0.994; numerals above branches indicate bootstrap values.

cutiens, *S. minutus*, *S. isodon*, *S. minutissimus*) с довольно слабой бутстреп-поддержкой.

В подроде *Otisorex* видны прежде всего высокая генетическая обособленность *S. trowbridgii* по первой оси шкалирования, *S. fumeus* – по второй, сестринские взаимоотношения *S. cinereus* и *S. haydeni* (группа «*cinereus*»), *S. monticolus* и *S. palustris* (группа «*vagrans*») (рис. 5).

Полученные данные согласуются как с результатами (Ohdachi et al., 2006), так и с хронологическими оценками времени дивергенции таксонов: расхождение *S. trowbridgii* с остальными видами – 8.91 Mya (млн лет назад), *S. fumeus* – 6.68 Mya, дивергенция группы «*vagrans*» от группы «*cinereus*» – 6.01 Mya, *S. cinereus* от *S. haydeni* – 3.71 Mya, *S. monticolus* от *S. palustris* – 1.96–2.41 Mya (Esteva et al., 2010).

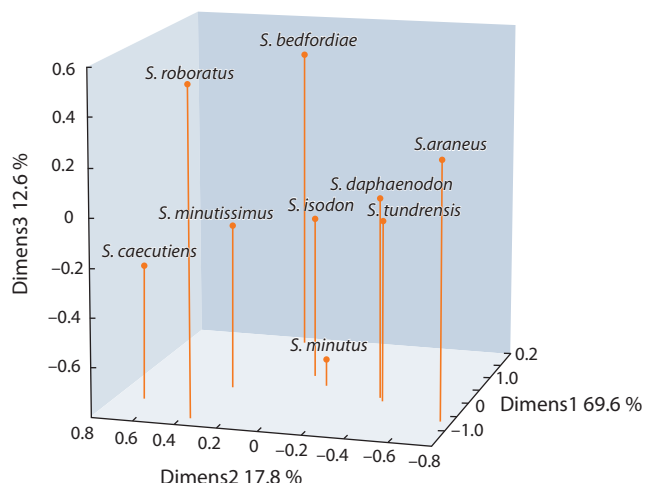


Fig. 4. Three-dimensional arrangement of shrew species by Kruskal's scaling of the joint matrix of genetic distances between centroids of 9 species of the *Sorex* subgenus with regard to two nuclear (*apoB* and *brca1*) and two mitochondrial (*co1* and *cytb*) genes.

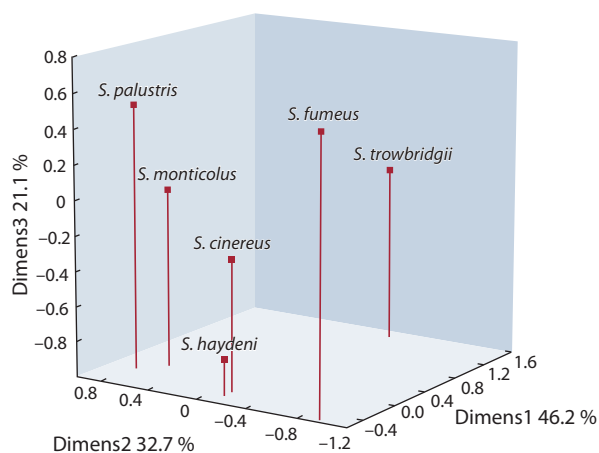


Fig. 5. Three-dimensional arrangement of shrew species by Kruskal's scaling of the joint matrix of genetic distances between centroids of 6 species of the *Otisorex* subgenus with regard to two nuclear (*apoB* and *brca1*) and two mitochondrial (*co1* and *cytb*) genes.

Обсуждение

В филогенетике сложилась традиция воспринимать генетические «дистанции» как настоящие расстояния, хотя, строго говоря, они не всегда таковыми являются (Ефимов и др., 2013). Дистанционные методы опираются на матрицы расстояний между видами. Однако матрицы содержат значительно больше информации, чем получаемые из них филогенетические деревья (Ковалева и др., 2012). По меткому выражению (Pershina et al., 2011), за филогенетическими деревьями мы не видим леса. Лес в данном случае – многомерная эволюция, различные варианты которой можно увидеть, только если учитывать все расстояния. Существует достаточно развитая технология отображения с минимальной потерей информации матриц сходства/различия в удобное метрическое

пространство – многомерное шкалирование (Дэйвисон, 1988). В результате этого отображения мы получаем так называемое «таксономическое или эволюционное пространство», полное представляющее «филогенетический сигнал» (Kitazoe et al., 2001, 2005; Huson, Bryant, 2006; Lee et al., 2006; Pershina et al., 2011; Дольник и др., 2012).

Из наших результатов следует, что достаточно конгруэнтные друг другу филогенетические сигналы выявились для всех четырех генов, как ядерных, так и митохондриальных (см. таблицу). В результате их объединения с помощью геометрического подхода получено дерево, практически совпадающее с принятой на сегодня зоологической систематикой (см. рис. 3) (Wilson, Reeder, 2005). Это подтверждает работоспособность предлагаемого нами метода. Кроме того, в объединенном генетическом пространстве выявляется несколько направлений изменчивости, отражающих эволюционные события, разные по масштабу и, по-видимому, относящиеся к разным эволюционным временам.

Следует отметить, что в последнее время наблюдается некоторое возрождение дистанционных методов. Вместо противопоставления филогенетических деревьев и многомерных геометрических представлений все чаще предпринимаются попытки построить филогенетическое дерево прямо в многомерном пространстве. С этой точки зрения особый интерес представляет евклидова задача Штейнера в n -мерном пространстве: для заданного множества точек найти составленную из отрезков кратчайшую сеть соединяющих путей (Lee et al., 2006; Brazil et al., 2009; Fonseca et al., 2014). Эта задача, известная не менее двух веков, имеет сейчас многочисленные применения в различных областях знания (Brazil et al., 2014).

Использование деревьев Штейнера как самых подходящих на роль филогенетических деревьев предложено еще в работе (Cavalli-Sforza, Edwards, 1967) и затем повторено в (Brazil et al., 2009). Прогресс в развитии вычислительных средств и алгоритмов делает подобную перспективу весьма актуальной. В настоящее время в Институте цитологии и генетики СО РАН разрабатывается пакет прикладных программ Jacobi 4, нацеленный на реализацию геометрического подхода для единообразного решения задач из различных областей биологии (Полунин и др., 2014). Задачу Штейнера предполагается включить в него в качестве одного из модулей.

Acknowledgments

This work was supported by State Budgeted Project 0324-2015-0003 and the Russian Foundation for Basic Research, project 14-04-00121a.

Conflict of interest

The authors declare no conflict of interest.

References

Abeyesundera M., Kenney T., Field C., Gu H. Combining distance matrices on identical taxon sets for multi-gene analysis with singular value decomposition. *PLoS One*. 2014;9(4):e94279. DOI 10.1371/journal.pone.0094279.

Beaumont M.A., Ibrahim K.M., Boursot P., Bruford M.W. Measuring genetic distance. *Molecular tools for screening biodiversity: Plants and*

Animals (Eds. A. Karp, D.S. Ingram, P.G. Isaac). London: Chapman & Hall, 1998;315-325. DOI 10.1007/978-94-009-0019-6_58.

Bininda-Emonds O.R.P., Gittleman J.L., Steel M.A. The (super) tree of life: procedures, problems, and prospects. *Annu. Rev. Ecol. Syst.* 2002;33:265-289. DOI 10.1146/annurev.ecolsys.33.010802.150511.

Brazil M., Graham R.L., Thomas D.A., Zachariasen M. On the history of the euclidean Steiner tree problem. *Archive History Exact Sciences*. 2014;68:327-354. DOI 10.1007/s00407-013-0127-z.

Brazil M., Thomas D.A., Nielsen B.K., Winter P., Wulff-Nilsen C., Zachariasen M. A novel approach to phylogenetic trees: d-dimensional geometric Steiner trees. *Networks*. 2009;53(2):104-111.

Burleigh J.G., Bansal M.S., Eulenstein O., Hartmann S., Wehe A., Vision T.J. Genome-scale phylogenetics: inferring the plant tree of life from 18,896 gene trees. *Syst. Biol.* 2011;60(2):117-125. DOI 10.1093/sysbio/syq072.

Cavalli-Sforza L.L., Edwards A.W. Phylogenetic analysis. Models and estimation procedures. *Am. J. Human Genet.* 1967;19(3):233-257.

Criscuolo A., Berry V., Douzery E.J., Gascuel O. SDM: a fast distance & based approach for (super) tree building in phylogenomics. *Syst. Biol.* 2006;55(5):740-755. DOI 10.1080/10635150600969872.

Criscuolo A., Michel C.J. Phylogenetic inference with weighted codon evolutionary distances. *J. Mol. Evol.* 2009;68(4):377-392. DOI 10.1007/s00239-009-9212-y.

Dannelid E. The genus *Sorex* (Mammalia, Soricidae) – distribution and evolutionary aspects of Eurasian species. *Mammal Rev.* 1991;21(1):1-20. DOI 10.1111/j.1365-2907.1991.tb00284.x.

de Queiroz A., Gatesy J. The supermatrix approach to systematic. *Trends Ecol. Evol.* 2007;22(1):34-41. DOI 10.1016/j.tree.2006.10.002.

Delsuc F., Brinkmann H., Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nature Rev. Genet.* 2005;6(5):361-375. DOI 10.1038/nrg1603.

Deyvison M. Mnogomernoe shkalirovanie [Multidimensional scaling]. Moscow, Finansy i statistika Publ., 1988. (in Russian)

Dol'nik A.S., Tamazyay G.S., Pershina E.V., Vyatkina K.V., Porozov Yu.B., Pinaev A.G., Andronov E.E. The evolutionary space of bacterial 16S rRNA gene v. 1.0. *Selskokhozyaystvennaya Biologiya = Agricultural Biology*. 2012;5:111-120. (in Russian)

Dubey S., Michaux J., Brünner H., Hutterer R., Vogel P. False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene. *Mol. Phylogen. Evol.* 2009;50(3):633-641. DOI 10.1016/j.ympev.2008.12.008.

Efimov V.M., Melchakova M.A., Kovaleva V.Yu. Geometric properties of evolutionary distances. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/1):714-723. (in Russian)

Esteva M., Cervantes F.A., Brant S.V., Cook J.A. Molecular phylogeny of long-tailed shrews (genus *Sorex*) from Mexico and Guatemala. *Zootaxa*. 2010;2615:47-65.

Fonseca R., Brazil M., Winter P., Zachariasen M. Faster exact algorithms for computing Steiner trees in higher dimensional euclidean spaces. 11th DIMACS Implementation challenge on Steiner tree problems. Providence, Rhode Island: Brown Univ., 2014.

Fumagalli L., Taberlet P., Stewart D.T., Gielly L., Haussier J., Vogel P. Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. *Mol. Phylogen. Evol.* 1999;11(2):222-235. DOI 10.1006/mpev.1998.0568.

Gadagkar S.R., Rosenberg M.S., Kumar S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J. Experimental Zoology. Pt. B: Molecular and Developmental Evolution*. 2005;304B(1):64-74. DOI 10.1002/jez.b.21026.

Gower J.C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*. 1966;53(3/4):325-338. DOI 10.2307/2333639.

Gower J.C., Legendre P. Metric and Euclidean properties of dissimilarity coefficients. *J. Classification*. 1986;3(1):5-48. DOI 10.1007/bf01896809.

Grechko V.V. The problems of molecular phylogenetics with the example of squamate reptiles: mitochondrial DNA markers. *Molekulyar-*

- naya biologiya = Molecular Biology (Moscow). 2013;47(1):61-82. (in Russian)
- Grechko V.V., Fedorova L.V., Ryabinin D.M., Ryabinina N.L., Chobanu D.G., Kosushkin S.A., Darevskiy I.S. The use of nuclear DNA molecular markers for studying speciation and systematics as exemplified by the "*Lacerta agilis* complex" (Sauria: Lacertidae). *Molekulyarnaya biologiya* = Molecular Biology (Moscow). 2006; 40(1):61-73. (in Russian)
- Havel T.F., Kuntz I.D., Crippen G.M. The theory and practice of distance geometry. *Bull. Mathem. Biol.* 1983;45(5):665-720. DOI 10.1007/bf02460044.
- Huson D.H., Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 2006;23(2):254-267. DOI 10.1093/molbev/msj030.
- Ivanitskaya E.Y. Comparative cytogenetics and systematics of *Sorex*: a cladistic approach. *Advances in the biology of shrews* (Eds. J.F. Merritt, G.L. Kirkland (Jr.), R.K. Rose). Pittsburgh: Carnegie Museum Nat. History, Spec. Publ. 1994;313-323.
- Jeffroy O., Brinkmann H., Delsuc F., Philippe H. Phylogenomics: the beginning of incongruence? *Trends Gen.* 2006;22(4):225-231. DOI 10.1016/j.tig.2006.02.003.
- Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 1980;16:111-120. DOI 10.1007/BF01731581.
- Kitazoe Y., Kurihara Y., Narita Y., Okuhara Y., Tominaga A., Suzuki T. A new theory of phylogeny inference through construction of multidimensional vector space. *Mol. Biol. Evol.* 2001;18(5):812-828.
- Kitazoe Y., Kishino H., Okabayashi T., Watabe T., Nakajima N., Okuhara Y., Kurihara Y. Multidimensional vector space representation for convergent evolution and molecular phylogeny. *Mol. Biol. Evol.* 2005;22(3):704-715. DOI 10.1093/molbev/msi051.
- Klingenberg C.P., Ekau W. A combined morphometric and phylogenetic analysis of an ecomorphological trend: pelagization in Antarctic fishes (Perciformes: Nototheniidae). *Biol. J. Linnean Soc.* 1996;59(2):143-177. DOI 10.1111/j.1095-8312.1996.tb01459.x.
- Klingenberg C.P., Gidaszewski N.A. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *System. Biol.* 2010;59(3):245-261. DOI 10.1093/sysbio/syp106.
- Kovaleva V.Yu., Abramov S.A., Dupal T.A., Efimov V.M., Litvinov Yu.N. Correspondence analysis and integration of molecular and morphological data in zoological classification. *Izvestiya RAN. Seriya biologicheskaya* = Bulletin of the RAS. Biological series. 2012;4:404-414. (in Russian)
- Kovaleva V.Yu., Litvinov Yu.N., Efimov V.M. Shrews (Soricidae, Eulipotyphla) from the Russian Far East and Siberia: Combination and search for congruence of molecular genetic and morphological data. *Zoologicheskii zhurnal* = Zoological Journal (Moscow). 2013;92(11): 1383-1398. (in Russian)
- Kruskal J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika.* 1964;29(1):1-27. DOI 10.1007/BF02289565.
- Lee S.H., Hwang K.S., Lee H.R., Kim S.S., Lee K.M., Lee C.H., Lee D. Embedding operational taxonomic units in three-dimensional space for evolutionary distance relationship in phylogenetic analysis. *Proc. 5th WSEAS Intern. Conf. on Circuits, Systems, Electronics, Control and Signal Processing.* USA. 2006;192-196.
- Lukashov V.V. *Molekulyarnaya evolyutsiya i filogeneticheskiy analiz* [Molecular evolution and phylogenetic analysis]. Moscow, Binom. Laboratoriya znaniy Publ., 2009. (in Russian)
- Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Research.* 1967;27:209-220.
- Mantel N., Valand R.S. A technique of nonparametric multivariate analysis. *Biometrics.* 1970;26:547-558. DOI 10.2307/2529108.
- Nylander J.A.A. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, 2. 2004.
- Ohdachi S.D., Hasegawa M., Iwasa M.A., Vogel P., Oshida T., Lin L.-K., Abe H. Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome *b* gene sequences: with special reference to the Soricinae. *J. Zool.* 2006;270(1):177-191. DOI 10.1111/j.1469-7998.2006.00125.x.
- Ohdachi S., Masuda R., Abe H., Adachi J., Dokuchaev N.E., Haukislam V., Yoshida M.C. Phylogeny of Eurasian soricine shrews (Insectivora, Mammalia) inferred from the mitochondrial cytochrome *b* gene sequences. *Zool. Science.* 1997;14(3):527-532.
- Pershina E.V., Dolnik A.S., Tamazyan G., Ikonnikova E.V., Vyatkina K.V., Pinaev A.G., Andronov E.E. An evolutionary space for microbial evolution and community structure analysis. Department of Bioengineering and Bioinformatics of M.V. Lomonosov Moscow State Univ. 2011;54(3):40.
- Philippe H., Delsuc F., Brinkmann H., Lartillot N. Phylogenomics. *Ann. Rev. Ecol., Evol., System.* 2005;541-562. DOI 10.1146/annurev.ecolsys.35.112202.130205.
- Planet P.J. Tree disagreement: measuring and testing incongruence in phylogenies. *J. Biomed. Inform.* 2006;39(1):86-102. DOI 10.1016/j.jbi.2005.08.008.
- Polly P.D., Lawing A.M., Fabre A.C., Goswami A. Phylogenetic principal components analysis and geometric morphometrics. *Hystrix, Italian J. Mammal.* 2013;24(1):33-41. DOI 10.4404/hystrix-24.1-6383.
- Polunin D.A., Shtayger I.A., Efimov V.M. Development of the JACOBI 4 package for multidimensional analysis of microarray data. *Vestnik NGU. Ser. Informatsionnye tekhnologii* = Novosibirsk State University Journal of Information Technologies. 2014;12(2):90-98. (in Russian)
- Scippa G.S., Trupiano D., Rocco M., Viscosi V., Di Michele M., D'Andrea A., Chiatante D. An integrated approach to the characterization of two autochthonous lentil (*Lens culinaris*) landraces of Molise (south-central Italy). *Heredity.* 2008;101(2):136-144. DOI 10.1038/hdy.2008.39.
- Torgerson W.S. Multidimensional scaling: I. Theory and method. *Psychometrika.* 1952;17(4):401-419. DOI 10.1007/BF02288916.
- Wilson D.E., Reeder D.A.M. (Ed.). *Mammal species of the world: a taxonomic and geographic reference.* Baltimore: JHU Press, 2005;12:2142 p.
- Wortley A.H., Scotland R.W. The effect of combining molecular and morphological data in published phylogenetic analyses. *Syst. Biol.* 2006;55(4):677-685. DOI 10.1080/10635150600899798.



Структурно-функциональная дивергенция гомеологичных генов в аллополиплоидном геноме растений

А.Ю. Глаголева^{1, 2}, О.Ю. Шоева¹, Е.К. Хлесткина^{1, 2}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

При гибридизации близких видов растений, имеющих сходные геномы, могут образовываться аллополиплоидные формы. Известно, что в ходе эволюции через аллополиплоидизацию прошли многие виды растений, что сыграло значительную роль в формировании огромного разнообразия растений, а также их высокого адаптивного потенциала. Сейчас, благодаря полному секвенированию широкого спектра видов покрытосеменных растений и сравнительному анализу структуры геномов, восстановлена цепь событий, в результате которых появились геномы современных растительных таксонов. Эти исследования показали, что многие диплоидные виды, прежде чем стать таковыми, прошли не один цикл полиплоидизации и дальнейшей диплоидизации. Цель обзора – на основе известных данных определить долю генов растительного генома, подверженных изменениям в случае аллополиплоидизации, и проиллюстрировать разнообразие механизмов, лежащих в основе функциональной дивергенции гомеологичных копий генов, т.е. генов-ортологов в субгеномах аллополиплоидного вида. Изменения отдельных копий могут быть связаны с эпигенетическими особенностями организации гена (статус метилирования промоторной области или наличие копий-специфичных малых интерферирующих РНК) или затрагивать первичную структуру гена в его кодирующей части или регуляторных районах. Исследования, проведенные на искусственно созданных аллополиплоидных формах растений, показали широкое распространение у них так называемого транскрипционного доминирования и изменение уровня транскрипции по сравнению с генами диплоидных родительских форм. Изучение транскрипции отдельных гомеологичных копий генов позволило оценить, насколько распространено полное подавление транскрипции отдельных гомеологов у вновь созданных синтетических (0.4–5.0 % генов) и естественных (около 30 % генов) аллополиплоидов. У пшеницы полное подавление вместе с частичными изменениями экспрессии затрагивает в сумме до 49 % генов.

Ключевые слова: пшеница; аллополиплоид; гомеологичные гены; транскрипционное доминирование; метилирование промотора; малые интерферирующие РНК.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Глаголева А.Ю., Шоева О.Ю., Хлесткина Е.К. Структурно-функциональная дивергенция гомеологичных генов в аллополиплоидном геноме растений. Вавиловский журнал генетики и селекции. 2016;20(6):823-831. DOI 10.18699/VJ16.204

HOW TO CITE THIS ARTICLE:

Glagoleva A.Y., Shoeva O.Y., Khlestkina E.K. Structural and functional divergence of homoeologous genes in allopolyploid plant genomes. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):823-831. DOI 10.18699/VJ16.204

REVIEW

Received 30.10.2016

Accepted for publication 12.12.2016

© AUTHORS, 2016

Structural and functional divergence of homoeologous genes in allopolyploid plant genomes

A.Y. Glagoleva^{1, 2}, O.Y. Shoeva¹,
E.K. Khlestkina^{1, 2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

Allopolyploid organisms can be formed by hybridization between closely related plant species with similar genomes. It is believed that many plant species have passed through allopolyploidization, which played a significant role in the formation of a huge diversity of plants, as well as their high adaptive capacity. Thanks to the whole genome sequencing of a wide range of angiosperm species and comparative analysis of genome structure, the sequence of events that formed the genomes of modern plant taxa was restored. These studies have shown that many diploid species have passed through more than one cycle of polyploidization-diploidization. The purpose of this review is to summarize the estimates of what proportion of genes is undergoing changes due to allopolyploidization and to illustrate the variety of mechanisms underlying the functional divergence of homoeologous copies (orthologous genes in allopolyploid sub-genomes). Changes of individual copies can be associated with epigenetic features of the gene organization (the methylation status of the promoter region or the presence of copy-specific small interfering RNA) or can affect structure of the coding or regulatory regions of the gene. Studies on artificial allopolyploid plants showed widespread transcriptional dominance and change of the transcription level as compared with the genes of diploid parental forms. The study of the transcription of certain homoeologous gene copies allowed estimating the extent of the complete suppression of certain homoeologous genes in newly synthesized (0.4–5.0 %) and natural (30 %) allopolyploids. One the whole, full or partial suppression affects up to 49% of the wheat genes.

Key words: wheat; allopolyploid; homoeologous genes; transcriptional dominance; promoter methylation; small interfering RNA.

Полиплоидия возникает в результате умножения полного набора хромосом. В случае умножения одного и того же набора хромосом формируются автополиплоиды, а при отдаленной гибридизации могут образовываться аллополиплоидные формы. Данные цитогенетических исследований позволили утверждать, что через аллополиплоидизацию и последующую диплоидизацию прошли многие виды растений (Masterson, 1994; Wolfe, 2001). В последние годы, благодаря появлению и сравнению результатов секвенирования геномов разных видов одно- и двудольных растений (рис. 1), стала известной эволюционная история происхождения геномов современных видов, многие из которых – результат нескольких циклов полиплоидизации – диплоидизации¹ (Salse et al., 2009).

Генетический материал разных видов и родов при аллополиплоидизации претерпевает значительные изменения, приводящие к стабилизации или дестабилизации полиплоидного генома. В случае стабилизации и успешной коадаптации разных геномов вновь созданная аллополиплоидная форма способна нормально развиваться и давать потомство. Было показано, что в основе коадаптации разных геномов в одном ядре лежат значительные структурные изменения. У аллополиплоидов часто наблюдается элиминация, а в более редких случаях, наоборот, появление новых фрагментов генома (Song et al., 1995; Pestsova et al., 1998; Comai, 2000; Ozkan et al., 2001; Kashkush et al., 2002; Salina et al., 2004; Ma, Gustafson, 2008; Щербань, 2013). Крупные структурные перестройки (элиминации блоков повторяющихся последовательностей, транслокации и т. п.) приводят к тому, что сходные хромосомы из разных субгеномов (гомеологичные хромосомы) не могут образовывать нормальные биваленты в ходе деления клеток аллополиплоидного организма. При этом имеет место частичная деконденсация хромосом, которая при наличии таких перестроек и обуславливает запрет на спаривание гомеологичных (но не гомологичных) хромосом. За деконденсацию у аллополиплоидной пшеницы отвечает локус *Ph1* (*pairing homoeologous 1*), который содержит ряд генов, кодирующих циклинзависимые киназы (*Cdk*). У стабильных аллополиплоидных форм пшеницы активность генов *Cdk* в локусе *Ph1* снижена, что способствует деконденсации и невозможности спаривания гомеологичных хромосом. Благодаря этому механизму в ходе клеточного деления у аллополиплоидной пшеницы имеет место спаривание только гомологичных хромосом и образование нормальных бивалентов. Аналогичные механизмы выявлены и у других аллополиплоидных видов (Griffiths et al., 2006; Al-Kaff et al., 2008).

Помимо структурных изменений, у аллополиплоидов наблюдается изменение экспрессии генов вплоть до полной супрессии отдельных гомеологов (Comai et al., 2000; Kashkush et al., 2002; Bottley et al., 2006; Pont et al., 2011). Изменение экспрессии генов у аллополиплоидных форм имеет важное значение для стабилизации работы генома. Цель обзора – на основе известных данных оценить долю

генов растительного генома, подверженных изменениям в случае аллополиплоидизации, и проиллюстрировать разнообразие механизмов, лежащих в основе функциональной дивергенции гомеологичных копий.

Полиплоидизация – ключевое событие в эволюции геномов растений

Полиплоидизация является одним из ключевых событий в эволюции геномов растений. Наличие «дублирующих» геномов в одном ядре способствует сохранению функционального состояния генома даже при крупных перестройках, в частности при потере целых хромосом. Примером стабильных форм с недостающими хромосомами могут служить анеуплоидные линии аллогексаплоидной пшеницы (Sears, 1954).

На основе данных полногеномного секвенирования (см. рис. 1) и сравнительного анализа (Salse et al., 2009) геномов риса, кукурузы, сорго и ячменя была создана модель предкового генома однодольных, порядок Злакоцветные (Poales), состоящего из пяти протохромосом (рис. 2). На этой модели показано, что дивергенция геномов современных Злакоцветных может быть описана четырьмя основными эволюционными событиями: α , β , δ и γ . Перед дивергенцией имела место полная дупликация предкового генома, состоящего из пяти протохромосом (δ -событие). После тетраплоидизации ($n = 10$) произошли транслокации и слияния между двумя хромосомами, и таким образом сформировался геном с базовым набором хромосом $n = 12$ (γ -событие). Далее в ходе дивергенции образовались предковые геномы различных подсемейств. Согласно этой модели, у Ehrhartoideae (подсемейство Рисовые) сохранилось количество хромосом предкового генома ($n = 12$) и геном риса претерпел только линейные перестройки с дупликациями хромосом *r11* и *r12* (α -события). Предковый геном Panicoideae (подсемейство Просовые) сформировался после слияния двух хромосом ($n = 10$) (β -событие). Кукуруза и сорго, принадлежащие этому подклассу, эволюционировали от общего предка поразному. В то время как структура современного генома сорго схожа с предковым геномом, геном кукурузы был полностью дублирован ($n = 20$), а затем в результате транслокаций и слияния хромосом (α -событие) сформировался геном современной кукурузы с базовым набором хромосом $n = 10$ (см. рис. 2). Наконец, при формировании генома Pooideae (подсемейство Мятликовые) из промежуточного предкового генома, состоящего из 12 хромосом, произошло пять хромосомных слияний. В результате образовался предковый геном трибы Triticeae (к ней относятся, в частности, пшеница, ячмень и рожь); он состоит из 7 хромосом (β -событие). Впоследствии геном общего предка трибы Triticeae дивергировал, а у отдельных ее представителей подвергся ряду геномных и хромосомных перестроек – полиплоидизации, дупликациям отдельных сегментов генома, транслокациям (Salse et al., 2009; Abrouk et al., 2010; Pont et al., 2011).

Сравнение данных полногеномного секвенирования представителей двудольных растений: арабидопсиса, винограда и тополя – предсказало две возможные модели формирования их геномов в ходе эволюции (см. рис. 2). В обеих моделях в качестве общего предка выступает

¹ В данном случае диплоидизация (син. диспloidизация) – процесс дивергенции субгеномов и возникновения крупных перестроек, нередко имеющих редуциционный характер, в результате которых геном перестает быть аллополиплоидным.

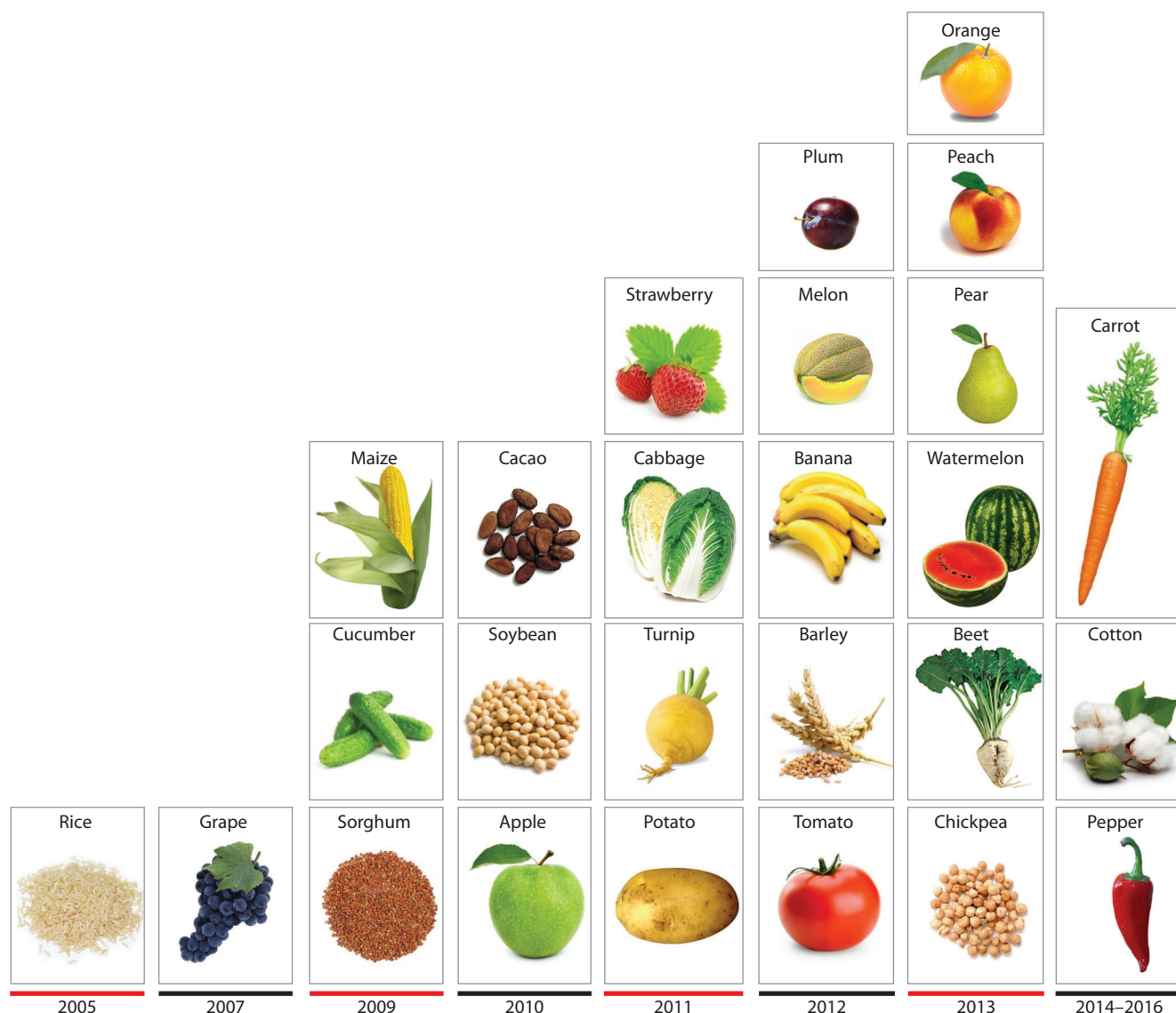


Fig. 1. Major food crops with sequenced genomes: rice (Goff et al., 2002), grape (Jaillon et al., 2007), maize (Schnable et al., 2009), cucumber (Huang et al., 2009), sorghum (Paterson et al., 2009), cacao (Argout et al., 2010), soybean (Huang et al., 2010), apple (Velasco et al., 2010), cabbage (Wang et al., 2011a), strawberry (Shulaev et al., 2011), turnip (Wang et al., 2011b), potato (Xu et al., 2011), melon (Garcia-Mas et al., 2012), banana (D'Hont et al., 2012), tomato (The Tomato Genome Consortium, 2012), barley (Mayer et al., 2012), plum (Zhang et al., 2012), pear (Wu et al., 2013), watermelon (Guo et al., 2013), orange (Xu et al., 2013), peach (The International Peach Genome Initiative et al., 2013), beet (Juliane et al., 2013), chickpea (Varshney et al., 2013), cotton (Li et al., 2014), pepper (Kim et al., 2014), carrot (Iorizzo et al., 2016).

форма с базовым набором, состоящим из семи протохромосом. Согласно первой модели, геном общего диплоидного предка сначала подвергся гексаплоидизации ($n = 21$; γ -событие), затем у предка современного арабидопсиса был еще дважды дублирован (α - и β -события), у предка тополя – однократно дублирован (β -событие), а у предка винограда дубликциям более не подвергался. При этом современные арабидопсис, тополь и виноград полиплоидами уже не являются, так как в ходе дальнейшей эволюции в геномах их предков происходили многочисленные слияния и транслокации хромосом. Согласно второй модели, геномы перечисленных двудольных растений произошли от тетраплоидного предкового генома ($n = 14$; γ -событие), который в ходе дальнейшей эволюции

подвергался дубликциям (α - и β -события), образуя предковые формы арабидопсиса, тополя и винограда, и, как и в первом сценарии, в геномах их предков происходили дальнейшие многочисленные слияния и транслокации хромосом, что в итоге привело к образованию современных диплоидных геномов арабидопсиса, тополя и винограда (Salse et al., 2009).

Таким образом, на основе сравнительных исследований данных полногеномного секвенирования стало возможным восстановить порядок событий, в результате которых образовались геномы современных одно- и двудольных (Salse et al., 2009; Abrouk et al., 2010). Дубликация генома дает ряд преимуществ, таких как увеличение дозы гена (что в отдельных случаях может иметь адаптивное значе-

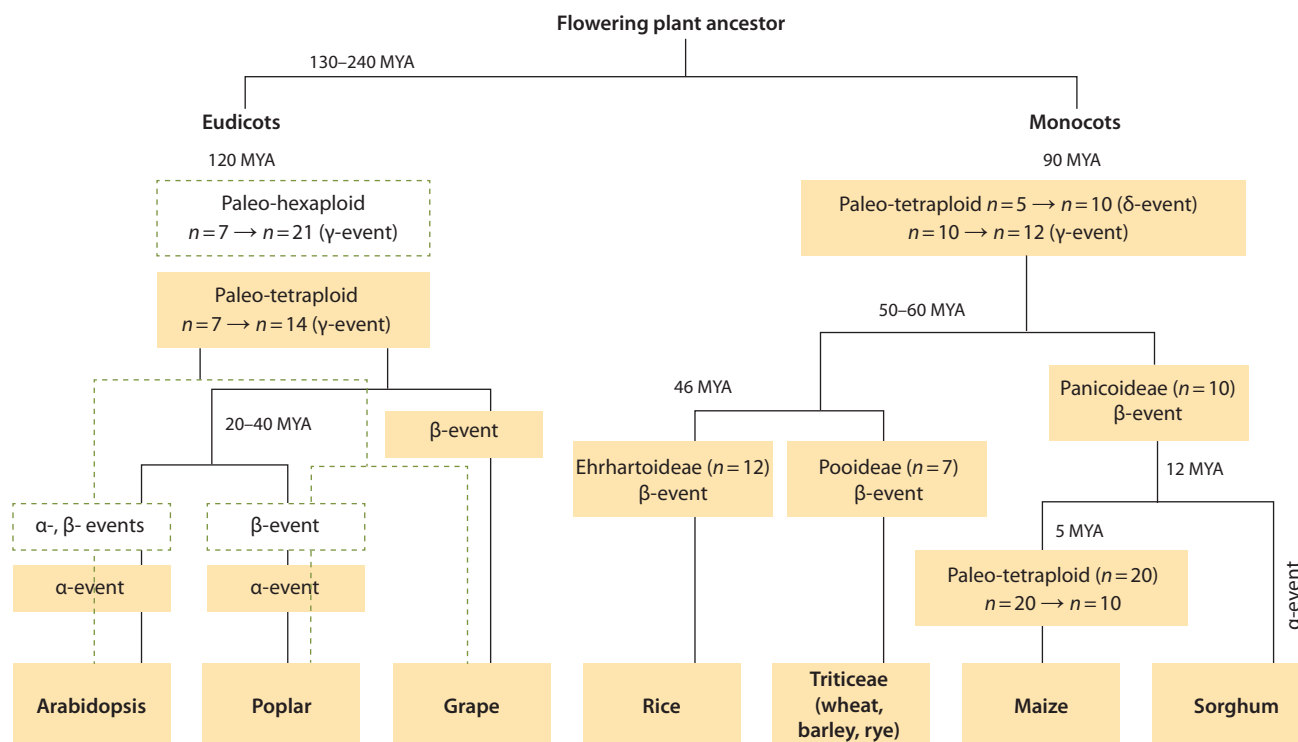


Fig. 2. Hypothetical order of events (by Salse et al., 2009) resulting in the formation of modern monocot and dicot species genomes, reconstructed by comparing the whole genome sequences of *Arabidopsis*, poplar, grape, rice, maize, sorghum and barley.

ние), «межгеномный гетерозис» (за счет сочетания разных гомеоаллелей в полиплоидном ядре), а также приобретение генами новых функций (Jiao et al., 2014). Известно, что полиплоидные формы чаще встречаются в районах с неблагоприятным климатом (Першина, 2009; Ekimova et al., 2012). Полиплоидизация сыграла значительную роль в формировании огромного разнообразия растений, а также их высокого адаптивного потенциала.

Транскрипция гомеологичных генов в аллополиплоидных геномах

Экспрессия генов у аллополиплоидных видов может претерпевать изменения по сравнению с экспрессией их у диплоидных доноров в связи с необходимостью нормализации уровня экспрессии. Высокий уровень экспрессии вследствие возрастания дозы гена при полиплоидизации может в некоторых случаях приводить к нарушениям в развитии растений (Chen, 2007). Однако изменения в экспрессии генов у аллополиплоидов по сравнению с диплоидными родителями необязательно связаны с умножением числа копий генов и необходимостью нормализации уровня экспрессии. Например, было выявлено, что у межвидовых гибридов (Reeder, 1985; Chen, Pikaard, 1997) транскрипция генов рРНК одного родительского вида доминирует над другим видом (феномен ядрышкового доминирования). В основе ядрышкового доминирования лежат конкуренция между генами рРНК, имеющими неравное число энхансерных элементов, неспособность ключевого фактора транскрипции узнавать промоторы генов другого вида (Reeder, 1985), а также эпигенетические механизмы

(Preuss et al., 2008). Таким образом, в основе изменения экспрессии генов у аллополиплоидов могут лежать иные причины, чем необходимость нормализации уровня экспрессии. Это могут быть и дисбаланс энхансеров, и видоспецифичность транскрипционных факторов, а также дисбаланс малой РНК (Ha et al., 2009).

Широкомасштабные исследования экспрессии генов у искусственно созданных аллотетраплоидов (Wang et al., 2006; Rapp et al., 2009) и аллогексаплоидов (Pumphrey et al., 2009) показали, что транскрипционное доминирование (явление, при котором транскрипция генов одного родительского вида доминирует над генами другого вида при объединении различных геномов в одном ядре) может иметь место как при наблюдаемой неаддитивной² (Wang et al., 2006; Pumphrey et al., 2009), так и при аддитивной экспрессии генов у аллополиплоидов (Rapp et al., 2009). При изучении тетраплоидного арабидопсиса установлено, что аллополиплоидизация сопровождается неслучайными изменениями паттернов метилирования генов, приводящими к изменению уровня транскрипции по сравнению с генами диплоидных родительских форм вплоть до полного подавления экспрессии генов (Madlung et al., 2002).

Исследование транскрипции отдельных гомеологичных копий генов позволило оценить, насколько распространено полное подавление транскрипции отдельных гомеологов у вновь созданных синтетических: около 0.4 % генов у тетраплоидов арабидопсиса (Comai et al., 2000), 1–5 % у

² При неаддитивной экспрессии суммарный уровень у гибрида не равен среднему значению уровней экспрессии у родителей.

тетраплоидов пшеницы (Kashkush et al., 2002) – и у естественных: около 30 % генов у мягкой пшеницы (Bottley et al., 2006) аллополиплоидов. Полное или частичное подавление экспрессии затрагивает до 49 % генов пшеницы (Pont et al., 2011).

Несмотря на то что для некоторых генов в аллополиплоидных геномах наблюдается супрессия отдельных гомеологических копий, для большинства генов (70–99 %) характерна коэкспрессия всех гомеологов (Comai et al., 2000; Kashkush et al., 2002; Bottley et al., 2006). Это свидетельствует о том, что при дубликации генетических систем в одном геноме в результате аллополиплоидизации возможна их согласованная работа, которая не приводит к инактивации одной из копий или к изменению паттернов экспрессии гомеологических генов, причем одни гены могут быть случайным образом инактивированы в одном из субгеномов, другие – в другом.

Одинаковая экспрессия гомеологических копий наблюдалась для целого ряда генов гексаплоидной пшеницы. Так, R. Morimoto и сотрудники (2005) изучили три гомеологических локуса гена *Wknox-1* пшеницы, который является гомологом гомеобоксного гена *knotted1*, функционирующего в апикальной меристеме побегов арабидопсиса. Наблюдалась одинаковая экспрессия всех трех гомеологов *Wknox-1*. Также выявлено, что в проростках пшеницы одинаково экспрессируются три гомеологические копии гена *F3h-1*, кодирующего ключевой фермент биосинтеза флавоноидных пигментов (Khlestkina et al., 2008). В условиях солевого стресса паттерны экспрессии изменялись по сравнению с контролем, при этом для трех копий наблюдались одинаковые изменения (Shoeva, Khlestkina, 2015).

N. Shitsukawa и коллеги (2007) изучили экспрессию генов *WSEP* и *WLHS-1* пшеницы из класса генов E MADS box, контролирующих развитие цветка у растений. Для трех гомеологов *WSEP* выявлены сходные уровни экспрессии, тогда как экспрессия гомеологов *WLHS-1* различалась. Копия *WLHS-1* из генома В транскрибировалась на очень низком уровне по сравнению с копиями из геномов А и D. Причины, лежащие в основе функциональной дивергенции трех гомеологов, различны. В-геномная копия *WLHS-1* проявляет слабую активность вследствие подавления транскрипции, связанного с метилированием цитозинового остатков (эпигенетический механизм). В то же время А-геномная копия, хотя и транскрибируется, функционально активного белкового продукта не производит из-за инсерции в районе, кодирующем функциональный домен (генетический механизм). Таким образом, *WLHS-1* в геноме D является, по сути, единственной функционально активной копией данного гена у пшеницы (Shitsukawa et al., 2007).

У пшеницы были описаны и другие случаи, при которых наблюдался различный уровень транскрипции коэкспрессирующихся гомеологических генов. При изучении экспрессии структурного гена биосинтеза гиббереллинов *GA20ox-1* в развивающемся зерне пшеницы N.E. Appleford с сотрудниками (2006) наблюдали пониженный уровень транскрипции В-геномной копии этого гена по сравнению с *GA20ox-1* из геномов А и D. T. Nomura и коллеги (2005) исследовали экспрессию структурного гена биосинтеза

бензоксазинонов *TaBx* в корнях и побегах пшеницы. Три гомеолога *TaBx* экспрессировались на разном уровне в различных тканях. R. Kikuchi и сотрудники (2008) изучили экспрессию трех гомеологических локусов *WFT* пшеницы, которые являются гомологами последовательности гена *FT* арабидопсиса, влияющего на сроки цветения, и показали, что ген *WFT* в геноме В транскрибируется более интенсивно по сравнению с двумя другими гомеологами. Наличие дополнительных копий одного и того же гена у полиплоидов создает дополнительные адаптивные возможности. Например, влияющие на развитие пшеницы гены *VRN-1* (vernalization sensitivity – чувствительность к яровизации) демонстрируют согласованный аддитивный характер экспрессии различных гомеологов при доминировании одного из них, при этом суммарный уровень транскрипции варьирует в зависимости от того, какой гомеолог доминирует (Loukoianov et al., 2005; Shcherban et al., 2013). Эти механизмы лежат в основе широкой изменчивости пшеницы по такому важному адаптивному признаку, как срок колошения.

Три гомеологические копии гена *Chi*, кодирующего один из ключевых ферментов биосинтеза флавоноидных соединений, по-разному экспрессировались в корнях пшеницы в ответ на солевой стресс: экспрессия *Chi-D1* усиливалась, также наблюдалось небольшое увеличение мРНК *Chi-A1*, тогда как транскрипта *Chi-B1* выявлено не было (Shoeva et al., 2014). D. Zhang с коллегами (2015) изучили экспрессию трех копий гена *GASR7*, связанного с определением длины зерна пшеницы. Основной вклад в общий уровень экспрессии вносила В-геномная копия этого гена. Сниженный уровень транскрипции А-геномной копии объясняется наличием малых интерферирующих РНК, специфичных для *pTaGASR7A*, D-геномной копии – повышенным уровнем метилирования ее промотора (Zhang et al., 2015).

Таким образом, исследования экспрессии генов у аллополиплоидной пшеницы показали, что в некоторых случаях уровни транскрипции коэкспрессирующихся гомеологических генов могут быть сходными (Morimoto et al., 2005; Shitsukawa et al., 2007; Khlestkina et al., 2008; Shoeva, Khlestkina, 2015), а в других случаях могут отличаться (Nomura et al., 2005; Appleford et al., 2006; Shitsukawa et al., 2007; Shoeva et al., 2014; Zhang et al., 2015). Причем различия в функциональности гомеологических генов могут быть не только следствием полиплоидизации (Shitsukawa et al., 2007). Функциональная дивергенция может возникнуть еще на уровне диплоидных предков (Nomura et al., 2005). Так, N. Shitsukawa с коллегами (2007) при сравнении генов *WLHS-A1* у ди-, тетра- и гексаплоидной пшеницы сделали вывод о том, что ген потерял свою функциональность в результате аллополиплоидизации. Между тем уровни экспрессии гомеологических генов *TaBx* у тетраплоидной пшеницы и диплоидных сородичей отличались друг от друга в том же соотношении, что и у гексаплоидной пшеницы. Это свидетельствует о том, что неравная экспрессия гомеологов *TaBx* закрепилась еще на уровне диплоидных предков и в таком виде сохранилась и после аллополиплоидизации (Nomura et al., 2005). При изучении дивергенции гена *Lr34* у пшеницы и ее диплоидных предков было показано, что этот важный

ген сформировался на полиплоидном уровне в D-геноме гексаплоидной пшеницы в результате всего двух мутаций в кодирующей части гена (Krattinger et al., 2011; Shcherban et al., 2016). При этом A-геномный гомеолог был инактивирован инсерцией транспозона еще у диплоидного предка этого генома, а B-геномная копия сохранила функцию, отличную от функции гена в геноме D.

Таким образом, в результате аллополиплоидизации геном растений претерпевает значительные генетические и эпигенетические изменения, которые способствуют коадаптации разных геномов в одном ядре. Однако различия, наблюдаемые между гомеологичными генами, локализованными в разных субгеномах, могут быть следствием как аллополиплоидизации, так и дивергенции, предшествующей объединению геномов диплоидных предков. При этом изменения отдельных копий могут не только затрагивать структуру кодирующей части гена или регуляторных областей, но и касаться эпигенетической регуляции экспрессии гена (статус метилирования промоторной области или наличие копий-специфичных малых интерферирующих РНК).

Acknowledgments

This work was supported by the Russian Foundation for Basic Research, project 16-04-00912, and the Siberian Branch of the Russian Academy of Sciences, Integrated Project II.2P «Integration and Development», program «Molecular and Cell Biology», project 0324-2015-0016.

Conflict of interest

The authors declare no conflict of interest.

References

- Abrouk M., Murat F., Pont C., Messing J., Jackson S., Faraut T., Tannier E., Plomion C., Cooke R., Feuillet C., Salse J. Palaeogenomics of plants: synteny-based modelling of extinct ancestors. *Trends Plant Sci.* 2010;15:479-487.
- Al-Kaff N., Knight E., Bertin I., Foote T., Hart N., Griffiths S., Moore G. Detailed dissection of the chromosomal region containing the Ph1 locus in wheat *Triticum aestivum*: with deletion mutants and expression profiling. *Ann. Bot.* 2008;101:863-872.
- Appleford N.E., Evans D.J., Lenton J.R., Gaskin P., Croker S.J., Devos K.M., Phillips A.L., Hedden P. Function and transcript analysis of gibberellin-biosynthetic enzymes in wheat. *Planta.* 2006;223:568-582.
- Argout X., Salse J., Aury J.M., Guiltinan M.J., Droc G., Gouzy J., Allegre M., Chaparro C., Legavre T., Maximova S.N., Abrouk M., Murat F., Fouet O., Poulain J., Ruiz M., Roguet Y., Rodier-Goud M., Barbosa-Neto J.F., Sabot F., Kudrna D., Ammiraju J.S., Schuster S.C., Carlson J.E., Sallet E., Schiex T., Dievart A., Kramer M., Gelley L., Shi Z., Bérard A., Viot C., Boccara M., Risterucci A.M., Guignon V., Sabau X., Axtell M.J., Ma Z., Zhang Y., Brown S., Bourge M., Golser W., Song X., Clement D., Rivallan R., Tahi M., Akaza J.M., Pitollat B., Gramacho K., D'Hont A., Brunel D., Infante D., Kebe I., Costet P., Wing R., McCombie W.R., Guiderdoni E., Quetier F., Panaud O., Wincker P., Bocs S., Lanaud C. The genome of theobroma cacao. *Nat. Gen.* 2010;43(2):101-108. DOI 10.1038/ng.736.
- Bottley A., Xia G.M., Koebner R.M. Homoeologous gene silencing in hexaploid wheat. *Plant J.* 2006;47:897-906.
- Chen Z.J. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Ann. Rev. Plant Biol.* 2007;58:377-406.
- Chen Z.J., Pikaard C.S. Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in Brassica. *Proc. Natl Acad. Sci. USA.* 1997;94:3442-3447.
- Comai L. Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* 2000;43:387-399.
- D'Hont A., Denoeud F., Aury J.M., Baurens F.C., Carreel F., Garsmeur O., Noel B., Bocs S., Droc G., Rouard M., Da Silva C., Jabbari K., Cardi C., Poulain J., Souquet M., Labadie K., Jourda C., Lengellé J., Rodier-Goud M., Alberti A., Bernard M., Correa M., Ayyampalayam S., Mckain M.R., Leebens-Mack J., Burgess D., Freeling M., Mbéguié-A-Mbéguié D., Chabannes M., Wicker T., Panaud O., Barbosa J., Hribova E., Heslop-Harrison P., Habas R., Rivallan R., Francois P., Poirion C., Kilian A., Burthia D., Jenny C., Bakry F., Brown S., Guignon V., Kema G., Dita M., Waalwijk C., Joseph S., Dievart A., Jaillon O., Leclercq J., Argout X., Lyons E., Almeida A., Jeridi M., Dolezel J., Roux N., Risterucci A.M., Weissenbach J., Ruiz M., Glaszmann J.C., Quétier F., Yahiaoui N., Wincker P. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature.* 2012;488(7410):213-217. DOI 10.1038/nature11241.
- Dohm J.C., Minoche A.E., Holtgräwe D., Capella-Gutiérrez S., Zakrzewski F., Tafer H., Rupp O., Sörensen T.R., Stracke R., Reinhardt R., Goesmann A., Kraft T., Schulz B., Stadler P.F., Schmidt T., Gabaldón T., Leirach H., Weisshaar B., Himmelbauer H. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature.* 2013;505(7484):546-549. DOI 10.1038/nature12817.
- Ekimova N.V., Muratova E.N., Silkin P.P. The role of polyploidy in adaptation and settling of steppe shrubs in Central Asia. *Russian Journal Genetics: Applied Research.* 2012;2:105-109.
- Garcia-Mas J., Benjak A., Sanseverino W., Bourgeois M., Mir G., González V.M., Hénaff E., Câmara F., Cozzuto L., Lowy E., Alioto T., Capella-Gutiérrez S., Blanca J., Cañizares J., Ziaresolo P., Gonzalez-Ibeas D., Rodríguez-Moreno L., Droege M., Du L., Alvarez-Tejado M., Lorente-Galdos B., Melé M., Yang L., Weng Y., Navarro A., Marques-Bonet T., Aranda M.A., Nuez F., Picó B., Gabaldón T., Roma G., Guigó R., Casacuberta J.M., Arús P., Puigdomènech P. The genome of melon (*Cucumis melo* L.). *PNAS.* 2012;109(29):11872-11877. DOI 10.1073/pnas.1205415109.
- Goff S.A., Ricke D., Lan T.H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., Hadley D., Hutchison D., Martin C., Katagiri F., Lange B.M., Moughamer T., Xia Y., Budworth P., Zhong J., Miguel T., Paszkowski U., Zhang S., Colbert M., Sun W.L., Chen L., Cooper B., Park S., Wood T.C., Mao L., Quail P., Wing R., Dean R., Yu Y., Zharkikh A., Shen R., Sahasrabudhe S., Thomas A., Cannings R., Gutin A., Pruss D., Reid J., Tavtigian S., Mitchell J., Eldredge G., Scholl T., Miller R.M., Bhatnagar S., Adey N., Rubano T., Tusneem N., Robinson R., Feldhaus J., Macalma T., Oliphant A., Briggs S. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science (New York).* 2002; 296(5565): 92-100.
- Griffiths S., Sharp R., Foote T.N., Bertin I., Wanous M., Reader S., Colas I., Moore G. Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature.* 2006;439:749-752.
- Guo S., Zhang J., Sun H., Salse J., Lucas W.J., Zhang H., Zheng Y., Mao L., Ren Y., Wang Z., Min J., Guo X., Murat F., Ham B.K., Zhang Z., Gao S., Huang M., Xu Y., Zhong S., Bombarely A., Mueller L.A., Zhao H., He H., Zhang Y., Zhang Z., Huang S., Tan T., Pang E., Lin K., Hu Q., Kuang H., Ni P., Wang B., Liu J., Kou Q., Hou W., Zou X., Jiang J., Gong G., Klee K., Schoof H., Huang Y., Hu X., Dong S., Liang D., Wang J., Wu K., Xia Y., Zhao X., Zheng Z., Xing M., Liang X., Huang B., Lv T., Wang J., Yin Y., Yi H., Li R., Wu M., Levi A., Zhang X., Giovannoni J.J., Wang J., Li Y., Fei Z., Xu Y. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Gen.* 2013;45:51-58. DOI 10.1038/ng.2470.
- Ha M., Lu J., Tian L., Ramachandran V., Kasschau K.D., Chapman E.J., Carrington J.C., Chen X., Wang X.J., Chen Z.J. Small RNAs serve

- as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proc. Natl Acad. Sci. USA*. 2009; 106(42):17835-17840. DOI 10.1073/pnas.0907003106.
- Huang S., Li R., Zhang Z., Li L., Gu X., Fan W., Lucas W.J., Wang X., Xie B., Ni P., Ren Y., Zhu H., Li J., Lin K., Jin W., Fei Z., Li G., Staub J., Kilian A., van der Vossen E.A., Wu Y., Guo J., He J., Jia Z., Ren Y., Tian G., Lu Y., Ruan J., Qian W., Wang M., Huang Q., Li B., Xuan Z., Cao J., Asan, Wu Z., Zhang J., Cai Q., Bai Y., Zhao B., Han Y., Li Y., Li X., Wang S., Shi Q., Liu S., Cho W.K., Kim J.Y., Xu Y., Heller-Uszynska K., Miao H., Cheng Z., Zhang S., Wu J., Yang Y., Kang H., Li M., Liang H., Ren X., Shi Z., Wen M., Jian M., Yang H., Zhang G., Yang Z., Chen R., Liu S., Li J., Ma L., Liu H., Zhou Y., Zhao J., Fang X., Li G., Fang L., Li Y., Liu D., Zheng H., Zhang Y., Qin N., Li Z., Yang G., Yang S., Bolund L., Kristiansen K., Zheng H., Li S., Zhang X., Yang H., Wang J., Sun R., Zhang B., Jiang S., Wang J., Du Y., Li S. The genome of the cucumber, *Cucumis sativus* L. *Nature Gen.* 2009;41(12):1275-1281. DOI 10.1038/ng.475.
- Iorizzo M., Ellison S., Senalik D., Zeng P., Satapoomin P., Huang J., Bowman M., Iovene M., Sanseverino W., Cavagnaro P., Yildiz M., Macko-Podgórní A., Moranska E., Grzebelus E., Grzebelus D., Ashrafi H., Zheng Z., Cheng S., Spooner D., Van Deynze A., Simon P. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nature Gen.* 2016;48(6):657-666. DOI 10.1038/ng.3565.
- Jailion O., Aury J.M., Noel B., Policriti A., Clepet C., Casagrande A., Choise N., Aubourg S., Vitulo N., Jubin C., Vezzi A., Legeai F., Huguency P., Dasilva C., Horner D., Mica E., Jublot D., Poulain J., Bruyère C., Billault A., Segurens B., Gouyvenoux M., Ugarte E., Cattonaro F., Anthouard V., Vico V., Del Fabbro C., Alaux M., Di Gaspero G., Dumas V., Felice N., Paillard S., Juman I., Moroldo M., Scalabrin S., Canaguier A., Le Clainche I., Malacrida G., Durand E., Pesole G., Laucou V., Chatelet P., Merdinoglu D., Delledonne M., Pezzotti M., Lecharny A., Scarpelli C., Artiguenave F., Pè M.E., Valle G., Morgante M., Caboche M., Adam-Blondon A.F., Weissenbach J., Quétier F., Wincker P. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 2007;449(7161):463-467.
- Jiao Y., Paterson A.H. Polyploidy-associated genome modifications during land plant evolution. *Philosophical Transactions of the Royal Society B: Biol. Sci.* 2014;369(1648):20130355.
- Kashkush K., Feldman M., Levy A.A. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*. 2002; 160:1651-1659.
- Khlestkina E.K., Röder M.S., Salina E.A. Relationship between homoeologous regulatory and structural genes in allopolyploid genome - a case study in bread wheat. *BMC Plant Biol.* 2008;8:88.
- Kikuchi R., Kawahigashi H., Handa H. The floral integrator WFT in wheat: expression profiles of three homoeologous genes. *Proc. 11th Intern. Wheat Gen. Symp. Brisbane*. 2008:271.
- Kim S., Park M., Yeom S.I., Kim Y.M., Lee J.M., Lee H.A., Seo E., Choi J., Cheong K., Kim K.T., Jung K., Lee G.W., Oh S.K., Bae C., Kim S.B., Lee H.Y., Kim S.Y., Kim M.S., Kang B.C., Jo Y.D., Yang H.B., Jeong H.J., Kang W.H., Kwon J.K., Shin C., Lim J.Y., Park J.H., Huh J.H., Kim J.S., Kim B.D., Cohen O., Paran I., Suh M.C., Lee S.B., Kim Y.K., Shin Y., Noh S.J., Park J., Seo Y.S., Kwon S.Y., Kim H.A., Park J.M., Kim H.J., Choi S.B., Bosland P.W., Reeves G., Jo S.H., Lee B.W., Cho H.T., Choi H.S., Lee M.S., Yu Y., Do Choi Y., Park B.S., van Deynze A., Ashrafi H., Hill T., Kim W.T., Pai H.S., Ahn H.K., Yeom I., Giovannoni J.J., Rose J.K., Sørensen I., Lee S.J., Kim R.W., Choi I.Y., Choi B.S., Lim J.S., Lee Y.H., Choi D. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nature Gen.* 2014;46(3):270-278. DOI 10.1038/ng.2877.
- Krattinger S.G., Lagudah E.S., Wicker T., Risk J.M., Ashton A.R., Selter L.L., Matsumoto T., Keller B. *Lr34* multi-pathogen resistance ABC transporter: molecular analysis of homoeologous and orthologous genes in hexaploid wheat and other grass species. *Plant J.* 2011; 65:392-403.
- Li F., Fan G., Wang K., Sun F., Yuan Y., Song G., Li Q., Ma Z., Lu C., Zou C., Chen W., Liang X., Shang H., Liu W., Shi C., Xiao G., Gou C., Ye W., Xu X., Zhang X., Wei H., Li Z., Zhang G., Wang J., Liu K., Kohel R.J., Percy R.G., Yu J.Z., Zhu Y.X., Wang J., Yu S. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Gen.* 2014;46(6):567-572. DOI 10.1038/ng.2987.
- Loukoianov A., Yan L., Blechi A., Sanchez A., Dubcovsky J. Regulation of *VRN-1* vernalization genes in normal and transgenic polyploid wheat. *Plant Physiol.* 2005;138:2364-2373.
- Ma X.F., Gustafson J.P. Allopolyploidization-accommodated genomic sequence changes in triticale. *Ann. Bot. (Lond.)*. 2008;101:825-832.
- Madlung A., Masuelli R.W., Watson B., Reynolds S.H., Davison J., Comai L. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic Arabidopsis allotetraploids. *Plant Physiol.* 2002;129:733-746.
- Masterson J. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science*. 1994;264:421-424.
- Mayer K.F., Waugh R., Brown J.W., Schulman A., Langridge P., Platzer M., Fincher G.B., Muehlbauer G.J., Sato K., Close T.J., Wise R.P., Stein N. A physical, genetic and functional sequence assembly of the barley genome. *Nature*. 2012;491(7426):711-716.
- Morimoto R., Kosugi T., Nakamura C., Takumi S. Intragenic diversity and functional conservation of the three homoeologous loci of the KN1-type homeobox gene *Wknox1* in common wheat. *Plant Mol. Biol.* 2005;57:907-924.
- Nomura T., Ishihara A., Yanagita R.C., Endo T.R., Iwamura H. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proc. Natl Acad. Sci. USA*. 2005;102: 16490-16495.
- Ozkan H., Levy A.A., Feldman M. Allopolyploidy-induced rapid genome evolution in the wheat *Aegilops-Triticum* group. *Plant Cell*. 2001;13:1735-1747.
- Paterson A., Bowers J., Bruggmann R., Dubchak I., Grimwood J., Gundlach H., Haberer G., Hellsten U., Mitros T., Poliakov A., Schmutz J., Spannagl M., Tang H., Wang X., Wicker T., Bharti A.K., Chapman J., Feltus F.A., Gowik U., Grigoriev I.V., Lyons E., Maher C.A., Martis M., Narechania A., Otillar R.P., Penning B.W., Salamov A.A., Wang Y., Zhang L., Carpita N.C., Freeling M., Gingle A.R., Hash C.T., Keller B., Klein P., Kresovich S., McCann M.C., Ming R., Peterson D.G., Mehboob-ur-Rahman, Ware D., Westhoff P., Mayer K.F., Messing J., Rokhsar D.S. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*. 2009;457(7229):551-556. DOI 10.1038/nature07723.
- Pershina L.A. On the role of wide hybridization and polyploidy in plants' evolution. *Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeding Scientists*. 2009;13:336-344. (in Russian)
- Pestsova E.G., Goncharov N.P., Salina E.A. Elimination of a tandem repeat of telomeric heterochromatin during the evolution of wheat. *Theor. Appl. Genet.* 1998;97:1380-1386.
- Pont C., Murat F., Confolent C., Balzergue S., Salse J. RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum* L.). RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum* L.). *Gen. Biol.* 2011;12:R119. DOI 10.1186/gb-2011-12-12-r119.
- Preuss S.B., Costa-Nunes P., Tucker S., Pontes O., Lawrence R.J., Mosher R., Kasschau K.D., Carrington J.C., Baulcombe D.C., Viegas W., Pikaard C.S. Multi-megabase silencing in nucleolar dominance results from siRNA-directed de novo DNA methylation recognized by specific methylcytosine binding proteins. *Mol. Cell*. 2008;32(5):673-684. DOI 10.1016/j.molcel.2008.11.009.
- Pumphrey M., Bai J., Laudencia-Chingcuanco D., Anderson O., Gill B.S. Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. *Genetics*. 2009; 181:1147-1157.
- Rapp R.A., Udall J.A., Wendel J.F. Genomic expression dominance in allopolyploids. *BMC Biol.* 2009;7:18.
- Reeder R.H. Mechanisms of nucleolar dominance in animals and plants. *J. Cell Biol.* 1985;101:2013-2016.

- Salina E.A., Numerova O.M., Ozkan H., Feldman M. Alterations in subtelomeric tandem repeats during early stages of allopolyploidy in wheat. *Genome*. 2004;47:860-867.
- Salse J., Abrouk M., Bolot S., Guilhot N., Courcelle E., Faraut T., Waugh R., Close T.J., Messing J., Feuillet C. Reconstruction of monotelodoneous proto-chromosomes reveals faster evolution in plants than in animals. *Proc. Natl Acad. Sci. USA*. 2009;106(35):14908-14913.
- Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian X., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang Z.C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A. Genome sequence of the palaeopolyploid soybean. *Nature*. 2010;463(12):178-183. DOI 10.1038/nature08670.
- Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S., Liang C., Zhang J., Fulton L., Graves T.A., Minx P., Reily A.D., Courtney L., Kruchowski S.S., Tomlinson C., Strong C., Delhaunty K., Fronick C., Courtney B., Rock S.M., Belter E., Du F., Kim K., Abbott R.M., Cotton M., Levy A., Marchetto P., Ochoa K., Jackson S.M., Gillam B., Chen W., Yan L., Higinbotham J., Cardenas M., Waligorski J., Applebaum E., Phelps L., Falcone J., Kanchi K., Thane T., Scimone A., Thane N., Henke J., Wang T., Ruppert J., Shah N., Rotter K., Hodges J., Ingenthron E., Cordes M., Kohlberg S., Sgro J., Delgado B., Mead K., Chinwalla A., Leonard S., Crouse K., Collura K., Kudrna D., Currie J., He R., Angelova A., Rajasekar S., Mueller T., Lomeli R., Scara G., Ko A., Delaney K., Wissotski M., Lopez G., Campos D., Braidotti M., Ashley E., Golsen W., Kim H., Lee S., Lin J., Dujmic Z., Kim W., Talag J., Zuccolo A., Fan C., Sebastian A., Kramer M., Spiegel L., Nascimento L., Zutavern T., Miller B., Ambrose C., Muller S., Spooner W., Narechania A., Ren L., Wei S., Kumari S., Faga B., Levy M.J., McMahon L., Van Buren P., Vaughn M.W., Ying K., Yeh C.T., Emrich S.J., Jia Y., Kalyanaraman A., Hsia A.P., Barbazuk W.B., Baucom R.S., Brutnell T.P., Carpita N.C., Chaparro C., Chia J.M., Deragon J.M., Estill J.C., Fu Y., Jeddellah J.A., Han Y., Lee H., Li P., Lisch D.R., Liu S., Liu Z., Nagel D.H., McCann M.C., SanMiguel P., Myers A.M., Nettleton D., Nguyen J., Penning B.W., Ponnala L., Schneider K.L., Schwartz D.C., Sharma A., Soderlund C., Springer N.M., Sun Q., Wang H., Waterman M., Westerman R., Wolfgruber T.K., Yang L., Yu Y., Zhang L., Zhou S., Zhu Q., Bennetzen J.L., Dawe R.K., Jiang J., Jiang N., Presting G.G., Wessler S.R., Aluru S., Martienssen R.A., Clifton S.W., McCombie W.R., Wing R.A., Wilson R.K. The B73 maize genome: complexity, diversity, and dynamics. *Science (N.Y.)*. 2009;326(5956):1112-1115.
- Sears E.R. The aneuploids of common wheat. *Missouri Agr. Expt. Sta. Res. Bull.* 1954;572:1-59.
- Shcherban A.B. The reorganization of plant genomes during allopolyploidisation. *Vavilovskii Zhurnal Genetiki i Selektzii = Vavilov Journal of Genetics and Breeding*. 2013;17(2):277-285. (in Russian)
- Shcherban A.B., Khlestkina E.K., Efremova T.T., Salina E.A. The effect of two differentially expressed wheat *VRN-B1* alleles on the heading time is associated with structural variation in the first intron. *Genetica*. 2013;141:133-141. DOI 10.1007/s10709-013-9712-y.
- Shcherban A.B., Kochieva E.Z., Salina E.A. Diversification of the homoeologous *Lr34* sequences in polyploid wheat species and their diploid progenitors. *J. Mol. Evol.* 2016;82(6):291-302. DOI 10.1007/s00239-016-9748-6.
- Shitsukawa N., Tahira C., Kassai K., Hirabayashi C., Shimizu T., Takumi S., Mochida K., Kawaura K., Oghihara Y., Murai K. Genetic and epigenetic alteration among three homoeologous genes of a class E MADS box gene in hexaploid wheat. *Plant Cell*. 2007;19:1723-1737.
- Shoeva O.Yu., Khlestkina E.K. Differently expressed 'early' flavonoid biosynthesis genes become to be co-regulated under salinity stress. *Cereal. Res. Commun.* 2015;43(4):537-543. DOI http://dx.doi.org/10.1556/0806.43.2015.025.
- Shoeva O.Yu., Khlestkina E.K., Berges H., Salina E.A. The homoeologous genes encoding chalcone-flavanone isomerase in *Triticum aestivum* L.: structural characterization and expression in different parts of wheat plant. *Gene*. 2014;538:334-341.
- Shulaev V., Sargent D.J., Crowhurst R.N., Mockler T.C., Folkerts O., Delcher A.L., Jaiswal P., Mockaitis K., Liston A., Mane S.P., Burns P., Davis T.M., Slovin J.P., Bassil N., Hellens R.P., Evans C., Harkins T., Kodira C., Desany B., Crasta O.R., Jensen R.V., Allan A.C., Michael T.P., Setubal J.C., Celton J.M., Rees D.J., Williams K.P., Holt S.H., Ruiz Rojas J.J., Chatterjee M., Liu B., Silva H., Meisel L., Adato A., Filichkin S.A., Troggio M., Viola R., Ashman T.L., Wang H., Dharmawardhana P., Elser J., Raja R., Priest H.D., Bryant D.W. Jr, Fox S.E., Givan S.A., Wilhelm L.J., Naithani S., Christoffels A., Salama D.Y., Carter J., Lopez Girona E., Zdepski A., Wang W., Kerstetter R.A., Schwab W., Korban S.S., Davik J., Monfort A., Denoyes-Rothan B., Arus P., Mittler R., Flinn B., Aharoni A., Bennetzen J.L., Salzberg S.L., Dickerman A.W., Velasco R., Borodovsky M., Veilleux R.E., Foltá K.M. The genome of woodland strawberry (*Fragaria vesca*). *Nature Gen.* 2011;43(2):109-116. DOI 10.1038/ng.740.
- Song K.M., Lu P., Tang K.L., Osborn T.C. Rapid genome changes in synthetic polyploids of Brassica and its implications for polyploidy evolution. *Proc. Natl Acad. Sci. USA*. 1995;92:7719-7723.
- The international Peach Genome Initiative, Verde I., Abbott A.G., Scalabrini S., Jung S., Shu S., Marroni F., Zhebentyayeva T., Detlori M.T., Grimwood J., Cattonaro F., Zuccolo A., Rossini L., Jenkins J., Vendramin E., Meisel L.A., Decroocq V., Sosinski B., Prochnik S., Mitros T., Policriti A., Cipriani G., Dondini L., Ficklin S., Goodstein D.M., Xuan P., Del Fabbro C., Aramini V., Copetti D., Gonzalez S., Horner D.S., Falchi R., Lucas S., Mica E., Maldonado J., Lazzari B., Bielenberg D., Pirona R., Miculan M., Barakat A., Testolin R., Stella A., Tartarini S., Tonutti P., Arús P., Orellana A., Wells C., Main D., Vizzotto G., Silva H., Salamini F., Schmutz J., Morgante M., Rokhsar D.S. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Gen.* 2013;45(5):487-494. DOI 10.1038/ng.2586.
- The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012;485(7400):635-641. DOI 10.1038/nature11119.
- Varshney R.K., Song C., Saxena R.K., Azam S., Yu S., Sharpe A.G., Cannon S., Baek J., Rosen B.D., Tar'an B., Millan T., Zhang X., Ramsay L.D., Iwata A., Wang Y., Nelson W., Farmer A.D., Gaur P.M., Soderlund C., Penmetsa R.V., Xu C., Bharti A.K., He W., Winter P., Zhao S., Hane J.K., Carrasquilla-Garcia N., Condie J.A., Upadhyaya H.D., Luo M.C., Thudi M., Gowda C.L., Singh N.P., Lichtenzveig J., Gali K.K., Rubio J., Nadarajan N., Dolezel J., Bansal K.C., Xu X., Edwards D., Zhang G., Kahl G., Gil J., Singh K.B., Datta S.K., Jackson S.A., Wang J., Cook D.R. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnol.* 2013;31:240-246. DOI 10.1038/nbt.2491.
- Velasco R., Zharkikh A., Affourtit J., Dhingra A., Cestaro A., Kalyanaraman A., Fontana P., Bhatnagar S., Troggio M., Pruss D., Salvi S., Pindo M., Baldi P., Castelletti S., Cavaiuolo M., Coppola G., Costa F., Cova V., Dal Ri A., Goremykin V., Komjanc M., Longhi S., Magnago P., Malacarne G., Malnoy M., Micheletti D., Moretto M., Perazzolli M., Si-Ammour A., Vezzulli S. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nature Gen.* 2010;42(10):833-839. DOI 10.1038/ng.654.
- Wang J., Tian L., Lee H.S., Wei N.E., Jiang H., Watson B., Madlung A., Osborn T.C., Doerge R.W., Comai L., Chen Z.J. Genomewide non-additive gene regulation in Arabidopsis allotetraploids. *Genetics*. 2006;172:507-517.
- Wang X., Wang H., Wang J., Sun R., Wu J., Liu S., Bai Y., Mun J.H., Bancroft I., Cheng F., Huang S., Li X., Hua W., Wang J., Wang X., Freeling M., Pires J.C., Paterson A.H., Chalhoub B., Wang B., Hay-

- ward A., Sharpe A.G., Park B.S., Weisshaar B., Liu B., Li B., Liu B., Tong C., Song C., Duran C., Peng C., Geng C., Koh C., Lin C., Edwards D., Mu D., Shen D., Soumpourou E., Li F., Fraser F., Conant G., Lassalle G., King G.J., Bonnema G., Tang H., Wang H., Belcram H., Zhou H., Hirakawa H., Abe H., Guo H., Wang H., Jin H., Parkin I.A., Batley J., Kim J.S., Just J., Li J., Xu J., Deng J., Kim J.A., Li J., Yu J., Meng J., Wang J., Min J., Poulain J., Wang J., Hatakeyama K., Wu K., Wang L., Fang L., Trick M., Links M.G., Zhao M., Jin M., Ramchiary N., Drou N., Berkman P.J., Cai Q., Huang Q., Li R., Tabata S., Cheng S., Zhang S., Zhang S., Huang S., Sato S., Sun S., Kwon S.J., Choi S.R., Lee T.H., Fan W., Zhao X., Tan X., Xu X., Wang Y., Qiu Y., Yin Y., Li Y., Du Y., Liao Y., Lim Y., Narusaka Y., Wang Y., Wang Z., Li Z., Wang Z., Xiong Z., Zhang Z., Brassica rapa Genome Sequencing Project Consortium. The genome of the mesopolyploid crop species *Brassica rapa*. Nature Gen. 2011a; 43:1035-1039. DOI 10.1038/ng.919.
- Wang X., Wang H., Wang J., Sun R., Wu J., Liu S., Bai Y., Mun J.H., Bancroft I., Cheng F., Huang S., Li X., Hua W., Wang J., Wang X., Freeling M., Pires J.C., Paterson A.H., Chalhou B., Wang B., Hayward A., Sharpe A.G., Park B.S., Weisshaar B., Liu B., Li B., Liu B., Tong C., Song C., Duran C., Peng C., Geng C., Koh C., Lin C., Edwards D., Mu D., Shen D., Soumpourou E., Li F., Fraser F., Conant G., Lassalle G., King G.J., Bonnema G., Tang H., Wang H., Belcram H., Zhou H., Hirakawa H., Abe H., Guo H., Wang H., Jin H., Parkin I.A., Batley J., Kim J.S., Just J., Li J., Xu J., Deng J., Kim J.A., Li J., Yu J., Meng J., Wang J., Min J., Poulain J., Wang J., Hatakeyama K., Wu K., Wang L., Fang L., Trick M., Links M.G., Zhao M., Jin M., Ramchiary N., Drou N., Berkman P.J., Cai Q., Huang Q., Li R., Tabata S., Cheng S., Zhang S., Zhang S., Huang S., Sato S., Sun S., Kwon S.J., Choi S.R., Lee T.H., Fan W., Zhao X., Tan X., Xu X., Wang Y., Qiu Y., Yin Y., Li Y., Du Y., Liao Y., Lim Y., Narusaka Y., Wang Y., Wang Z., Li Z., Wang Z., Xiong Z., Zhang Z., Brassica rapa Genome Sequencing Project Consortium. The genome of the mesopolyploid crop species *Brassica rapa*. Nature Gen. 2011b;43:1035-1039. DOI 10.1038/ng.919.
- Wolfe K.H. Yesterday's polyploidization and the mystery of diploidization. Nat. Rev. Genet. 2001;2:233-241.
- Wu J., Wang Z., Shi Z., Zhang S., Ming R., Zhu S., Khan M.A., Tao S., Korban S.S., Wang H., Chen N.J., Nishio T., Xu X., Cong L., Qi K., Huang X., Wang Y., Zhao X., Wu J., Deng C., Gou C., Zhou W., Yin H., Qin G., Sha Y., Tao Y., Chen H., Yang Y., Song Y., Zhan D., Wang J., Li L., Dai M., Gu C., Wang Y., Shi D., Wang X., Zhang H., Zeng L., Zheng D., Wang C., Chen M., Wang G., Xie L., Sovero V., Sha S., Huang W., Zhang S., Zhang M., Sun J., Xu L., Li Y., Liu X., Li Q., Shen J., Wang J., Paull R.E., Bennetzen J.L., Wang J., Zhang S. The genome of the pear (*Pyrus bretschneideri* Rehd.). Gen. Res. 2013;23:396-408. DOI 10.1101/gr.144311.112.
- Xu Q., Chen L.L., Ruan X., Chen D., Zhu A., Chen C., Bertrand D., Jiao W.B., Hao B.H., Lyon M.P., Chen J., Gao S., Xing F., Lan H., Chang J.W., Ge X., Lei Y., Hu Q., Miao Y., Wang L., Xiao S., Biswas M.K., Zeng W., Guo F., Cao H., Yang X., Xu X.W., Cheng Y.J., Xu J., Liu J.H., Luo O.J., Tang Z., Guo W.W., Kuang H., Zhang H.Y., Roose M.L., Nagarajan N., Deng X.X., Ruan Y. The draft genome of sweet orange (*Citrus sinensis*). Nature Gen. 2013; 45(1):59-66. DOI 10.1038/ng.2472.
- Xu X., Xu S., Pan S., Cheng B., Zhang D., Mu P., Ni G., Zhang S., Yang R., Li J., Wang G., Orjeda F., Guzman M., Torres R., Lozano O., Ponce D., Martinez G.N., De La Cruz S.K., Chakrabarti V.U., Patil K.G., Skryabin B.B., Kuznetsov N.V., Ravin T.V., Kolganova A.V., Beletsky A.V., Mardanov A., Di Genova D.M., Bolser D.M.A., Martin G., Li Y. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475(7355):189-195. DOI 10.1038/nature10158.
- Zhang D., Wang B., Zhao J., Zhao X., Zhang L., Liu D., Dong L., Wang D., Mao L., Li L. Divergence in homoeolog expression of the grain length-associated gene GASR7 during wheat allohexaploidization. Crop J. 2015;3(1):1-9. DOI http://dx.doi.org/10.1016/j.cj.2014.08.005.
- Zhang Q., Chen W., Sun L., Zhao F., Huang B., Yang W., Tao Y., Wang J., Yuan Z., Fan G., Xing Z., Han C., Pan H., Zhong X., Shi W., Liang X., Du D., Sun F., Xu Z., Hao R., Lv T., Lv Y., Zheng Z., Sun M., Luo L., Cai M., Gao Y., Wang J., Yin Y., Xu X., Cheng T., Wang J. The genome of *Prunus mume*. 2012;3:1318. DOI 10.1038/ncomms2290.

Анализ генной сети Neurogenesis:Prepattern, контролирующей первый этап формирования щетиночного узора у *Drosophila melanogaster*

Д.П. Фурман^{1, 2}, Т.А. Бухарина¹ 

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Хитиновый покров насекомых не способен реагировать на раздражители, и приемниками сигналов из окружающей среды для них служат специализированные рецепторы. Дрозофила воспринимает тактильные стимулы посредством внешних сенсорных органов – микро- и макрохет, расположенных на голове и спинке (нотуме). Микрохеты многочисленны и образуют правильные ряды, ориентированные вдоль тела. Число макрохет относительно невелико, и их расположение на голове и нотуме строго определено. Макрохеты выполняют функцию механорецепторов, обеспечивающих мухе сохранение равновесия в полете. Набор макрохет называют щетиночным узором. Полноценный щетиночный узор взрослой *Drosophila melanogaster* складывается в результате многоступенчатого процесса. Основопологающая его стадия состоит в создании прообраза расположения будущих макрохет – препаттерна (предструктуры), представленного proneуральными кластерами. Пронеуральные кластеры обособляются из массы клеток крыловых имажинальных дисков на стадии личинки третьего возраста и ранней предкуколки под действием факторов предструктуры, отождествляемых с транскрипционными факторами, направленными на регулирование экспрессию своих генов-мишеней в соответствующих районах. В статье впервые приведены результаты реконструкции и анализа генной сети, обеспечивающей процесс становления предструктуры, рассмотрены принципы ее организации и функционирования. Сеть насчитывает 80 объектов, связанных 109 регуляторными взаимодействиями. Ключевыми объектами сети, показывающими наибольшую связность с другими ее компонентами, являются proneуральные белки ASC, кодируемые генами *achaete* и *scute*, а также белки Decapentaplegic (Dpp) и Wingless (Wg). Структура сети характеризуется иерархической организацией и имеет по крайней мере три уровня управления. Функционирование сети как ансамбля генов в целом достигается координированной работой регуляторных контуров, осуществляющих как внутри-, так и межуровневый контроль активности генов. Результирующий эффект действия сети состоит в активации proneуральных генов комплекса AS-C, экспрессия которых отличает клетки proneурального кластера от окружающих клеток эктодермы.

Ключевые слова: *Drosophila melanogaster*; макрохеты; щетиночный узор; предструктура; *achaete-scute* комплекс; генная сеть; регуляторные контуры.

Analysis of the Neurogenesis: Prepattern gene network controlling the first stage in *Drosophila melanogaster* bristle pattern development

D.P. Furman^{1, 2}, T.A. Bukharina¹ 

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

The external insect chitinous skeleton is unable to respond to stimuli; the external signals are received by specialized receptors. *Drosophila* perceives the tactile stimuli by its external sensory organs, the microchaetes and macrochaetes residing on the head and back (notum). The microchaetes (hairs) are numerous and arranged in perfect rows along the body. The macrochaetes (bristles) are rather few and are strictly positioned on the head and notum, being referred to as bristle pattern. Bristles act as mechanoreceptors, providing balance for flying *drosophila*. The proper bristle pattern of an adult fly develops through several stages. Its basic stage is formation of prepattern for the future bristles, represented by proneural clusters. The proneural clusters separate from the ectodermal cells in imaginal discs in the third instar larvae and early prepupae. They are induced by prepattern factors, identified with the transcription factors driving expression of their target genes in certain disc regions. Reconstruction of the gene network controlling prepattern development and its analysis are for the first time described as well as the principles underlying arrangement and function of this network. The hierarchical structure of the network, its key components, and regulatory circuits are identified. The network comprises 80 entities interconnected via 109 regulatory interactions. The key objects of the network, displaying the greatest connectivity with its other components, are the ASC proneural proteins encoded by the *achaete* and *scute* genes, and the proteins Decapentaplegic (Dpp) and Wingless (Wg). The structure of the network is hierarchical and has at least three control levels. The network acts as a gene ensemble owing to coordinated functioning of the regulatory circuits

controlling activities of the corresponding genes both within and between the levels. The resulting effect of the network operation consists in activation of the *AS-C*, proneural genes, the expression of which distinguishes the cells of proneural cluster from the surrounding ectodermal cells.

Key words: *Drosophila melanogaster*; bristles; bristle pattern; prepatter; *achaete-scute* complex; gene network; regulatory circuits.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Фурман Д.П., Бухарина Т.А. Анализ геной сети Neurogenesis:Prepattern, контролирующей первый этап формирования щетиночного узора у *Drosophila melanogaster*. Вавиловский журнал генетики и селекции. 2016;20(6):832-839. DOI 10.18699/VJ16.199

HOW TO CITE THIS ARTICLE:

Furman D.P., Bukharina T.A. Analysis of the Neurogenesis:Prepattern gene network controlling the first stage in *Drosophila melanogaster* bristle pattern development. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):832-839. DOI 10.18699/VJ16.199

Под щетиночным узором подразумевается набор макрохет (больших щетинок), расположенных на голове и нотуме дрозофилы. Макрохеты – это внешние сенсорные органы, механорецепторы, воспринимающие тактильные раздражители. Они необходимы для сохранения равновесия мухи в полете. Характеристические признаки щетиночного узора (число и позиции макрохет) строго фиксированы, что позволяет использовать его для классификации видов дрозофилид.

У *Drosophila melanogaster* щетиночный узор нотума образован одиннадцатью парами макрохет, имеющими собственные названия в соответствии с локализацией. Каждая из макрохет входит как один из компонентов в состав механорецептора, состоящего из собственно щетинки, окружающего ее основание цоколя, биполярного нейрона и покрывающей нейрон оболочки. Щетинка и цоколь видны на поверхности тела, иннервирующий нейрон лежит в полости тела под механорецептором. Компоненты механорецептора формируются четырьмя специализированными клетками: щетинковой (трихогеном), гнездовой (тормогеном), нейральной и клеткой оболочки нерва (текогеном). Все эти клетки происходят путем последовательных делений единственной родительской клетки, которая обособляется из пронеуральных кластеров – групп из 20–30 клеток крыловых имажинальных дисков (Lai, Orgogozo, 2004).

Именно позиционирование пронеуральных кластеров однозначно задает топографию щетиночного узора. Закладка кластеров с последующим выделением родительских клеток происходит на стадии третьего личиночного и раннего кукольного возраста. Кластеры для каждого механорецептора закладываются независимо, так что мутации генов *AS-C* фенотипически проявляются в утрате отдельных щетинок или их комбинаций (Серебровский, Дубинин, 1929).

Клетки кластеров отличаются от окружения наличием белков *Achaete* и *Scute* (Romani et al., 1989; Cubas et al., 1991; Skeath, Carroll, 1991; Gómez-Skarmeta et al., 2003). Кодрующие эти белки гены *achaete* и *scute* (*ac*, *sc*) входят в одноименный комплекс *AS-C*, и им отводится центральное место в формировании узора. Механизмы, обеспечивающие локальную активность комплекса в кластерах, долгое время оставались предметом дискуссий. В 1954 г. К. Штерном (Stern, 1954) была предложена гипотеза, согласно которой экспрессия генов *AS-C* начинается

в ответ на действие неких факторов, пространственно распределенных в эктодерме имажинальных дисков. В результате клетки в районах присутствия этих факторов предетерминируются к развитию по нейральному пути, формируя пронеуральные кластеры и создавая тем самым предструктуру – прообраз щетиночного узора (Reeves, Posakony, 2005). Таким образом, постулировалось, что щетиночный узор возникает вследствие взаимодействия двух компонентов: предструктуры и системы ответа на предструктуру – генов комплекса *achaete-scute*.

Дальнейшее изучение закономерностей становления щетиночного узора позволило конкретизировать понятие факторов предструктуры. В настоящее время с ними отождествляются транскрипционные факторы, регулирующие активность *AS-C*: U-shaped (*Ush*), *Pannier* (*Pnr*), белки генов комплекса *iroquois* (*Iro-C*) (*Araucan*, *Scapolican* и *Mirror*) и др. Экспрессия генов, кодирующих эти транскрипционные факторы, определяется своим набором регуляторных белков, в том числе *Decapentaplegic* (*Dpp*), *Hedgehog* (*Hh*), *Engrailed* (*En*), *Wingless* (*Wg*) (Ruiz-Gomez, Modolell, 1987; Gómez-Skarmeta et al., 1995, 1996, 2003; Cubadda et al., 1997; Haenlin et al., 1997; Calleja et al., 2000, 2002; Romain et al., 2000; Aldaz et al., 2003; Joshi et al., 2006; Ikmi et al., 2008).

Основные морфогенетические события, обуславливающие формирование щетиночного узора на нотуме дрозофилы, связаны с парой крыловых имажинальных дисков, каждый из которых дает полуторакс взрослой мухи.

В ходе многолетнего исследования молекулярно-генетических механизмов, детерминирующих формирование щетиночного узора, накоплен значительный объем экспериментальных данных. На основе их систематизации и анализа нами реконструирована геновая сеть, обеспечивающая процесс становления предструктуры. Выявлена иерархическая организация и ключевые объекты сети, показано наличие регуляторных контуров, управляющих ее функционированием.

Материалы и методы

Имеющиеся данные по генетическому контролю развития макрохет сведены в базу данных Neurogenesis (Furman, Bukharina, 2008; Бухарина, Фурман, 2009). Ресурс создан путем ручного аннотирования более 350 литературных первоисточников, позволяющего наиболее полно экстрагировать содержащиеся в них экспериментальные данные.

В настоящее время в базе сосредоточена информация о 110 генах, 158 продуктах этих генов (147 белках и 11 РНК) и 13 типах клеток.

На основе этой базы данных с помощью технологии GeneNet (Колчанов и др., 2000) реконструирована генная сеть Neurogenesis:Prepattern, осуществляющая управление процессом становления предструктуры как набора пространственно распределенных транскрипционных факторов в клетках крыловых имагинальных дисков, которая и определит возможность формирования макрохет в строго определенных районах нотума дрозофилы.

Структура сети анализировалась на предмет поиска ключевых объектов и замкнутых регуляторных контуров – основных характеристик генных сетей.

Результаты и обсуждение

Реконструированная генная сеть Neurogenesis:Prepattern, роль которой состоит в позиционировании районов расположения будущих макрохет, сформирована 80 объектами, в число которых входят 44 белка, 34 гена, 2 РНК. Объекты объединены 109 связями. Сеть охватывает события, происходящие на стадии третьего личиночного и раннего кукольного возраста дрозофилы.

В структуре сети выявляются три ключевых объекта, показывающих наибольшую связность с другими ее компонентами – это белки Decapentaplegic (Dpp) и Wingless (Wg) с 11 и 8 связями соответственно, а также *AS-C*, на долю которого приходится 23 связи (рис. 1, а). Первые два выступают в основном как регулирующие факторы: Dpp направляет 9, а Wg – 7 процессов. В отличие от них *AS-C* фокусирует на себе действие большинства компонентов сети и выступает объектом 18 прямых и 6 опосредованных регуляторных воздействий, осуществляющих контроль его экспрессии.

Структура сети характеризуется иерархической организацией и имеет по меньшей мере три уровня управления, осуществляющих каскадную регуляцию комплекса *AS-C* как финальной мишени координированного действия ее элементов (см. рис. 1, б).

Основные фигуранты **первого уровня управления** представлены двумя из трех ключевых объектов сети – морфогенами Dpp и Wg, участниками одноименных сигнальных путей, определяющих дифференцировку клеток, рост, компартментализацию и метаморфоз имагинальных дисков начиная с самых ранних этапов их развития – времени установления дорзо-вентральной и антерио-постериорной осей диска. Кроме них, в число элементов первого уровня входят белки Engrailed (En) и Hedgehog (Hh) – регуляторы активности гена *dpp* с антагонистичным эффектом на его экспрессию (Serrano, O'Farrell, 1997; Affolter, Basler, 2007; Akiyama, Gibson, 2015).

Присутствие в клетке Hh инициирует экспрессию *dpp*, тогда как присутствие En ее запрещает. При наличии обоих белков экспрессии *dpp* также не происходит (Zessa et al., 1995; Layalle et al., 2011). Таким образом, En служит одним из лимитирующих факторов при функционировании первого уровня каскадной регуляции комплекса *achaete-scute*.

Содержание белков Wg и Dpp подчинено отношениям взаимной регуляции, основанным на перекрестном по-

давлении экспрессии генов друг друга. Посредниками в этом процессе выступают протеинкиназа GSK-3 (glycogen synthase kinase-3), кодируемая геном *shaggy* (*sgg*), и белок Dishevelled (Dsh), внутриклеточный передатчик сигнала в Wnt сигнальном пути (Komiya, Habas, 2008).

Схема взаимодействий, регламентирующих содержание морфогенов Wg и Dpp, представлена на рис. 2 (Penton, Hoffmann, 1996; Heslip et al., 1997).

Районы имагинального диска, в которых экспрессируется ген *dpp*, потенциально способны дать начало как нотуму, так и крыловым структурам, а дальнейшая их судьба зависит от уровня экспрессии *wg*. Зоны диска, где *wg* экспрессируется слабо или не экспрессируется вовсе, дадут начало имагинальному нотуму и шарниру крыла, тогда как высокий уровень экспрессии жестко детерминирует развитие зачатка дефинитивного крыла.

Компоненты первого уровня системы – белки Dpp и Wg – осуществляют контроль функционирования соподчиненных генов *brinker* (*brk*), *pannier* (*pnr*), *u-shaped* (*ush*), *spalt* (*sal*) и *spalt-related* (*salr*), кодирующих белки, принадлежащие **второму уровню управления** регуляцией комплекса *AS-C*. Этот уровень определяет рост диска и дальнейшую компартментализацию презумптивного нотума, в том числе места закладки пронеуральных кластеров.

Известно, что большая часть пронеуральных кластеров сосредоточена в тех частях диска, где в клетках под контролем морфогена Dpp произошла активация генов *pannier* и *u-shaped* с одновременным подавлением экспрессии *brinker* (Cubadda et al., 1997; Haenlin et al., 1997; Campbell, Tomlinson, 1999; Martín et al., 2004).

При содействии участия сигнальных белков Dpp, Wg и Hh осуществляется регуляция экспрессии генов комплекса *sal/salr* – еще двух компонентов второго уровня регуляции, конфигурирующих локализацию большинства будущих макрохет нотума. Показано, что Dpp и Hh активируют, а Wg подавляет их экспрессию (de Celis et al., 1999).

Таким образом, основную часть элементов второго уровня контролирует Dpp. На долю Wg приходится лишь регуляция активности белка Pnr. Известно, что Wg инициирует экспрессию гена *sgg*, кодирующего киназу, которая фосфорилирует Pnr. Модифицированная форма Pnr теряет способность связываться с регуляторными районами генов-мишеней и выполнять функцию транскрипционного фактора (Yang et al., 2012).

Наряду с вертикалью регуляции активности генов этого уровня за счет вышестоящих в иерархии Dpp и Wg, существует и горизонтальный внутриуровневый контроль их активности за счет регуляторного контура, устанавливающего баланс белков Pnr и Ush (рис. 3), от которого, в свою очередь, зависит состояние активации – репрессии подконтрольных контуру генов *AS-C*.

Белок Pnr кодируется одним геном, но существует в двух изоформах – Pnr α и Pnr β , экспрессия которых контролируется двумя разными районами промотора. Зоны локализации изоформ в имагинальном диске частично перекрываются, но в дорзальной части презумптивного нотума преобладает Pnr β , тогда как в латеральной – Pnr α . Образовав комплекс с Pnr, Ush превращает его из активатора в репрессор транскрипции: Pnr β /Ush подавляет наработку изоформы Pnr α , комплексы Pnr α /Ush и Pnr β /Ush репресси-

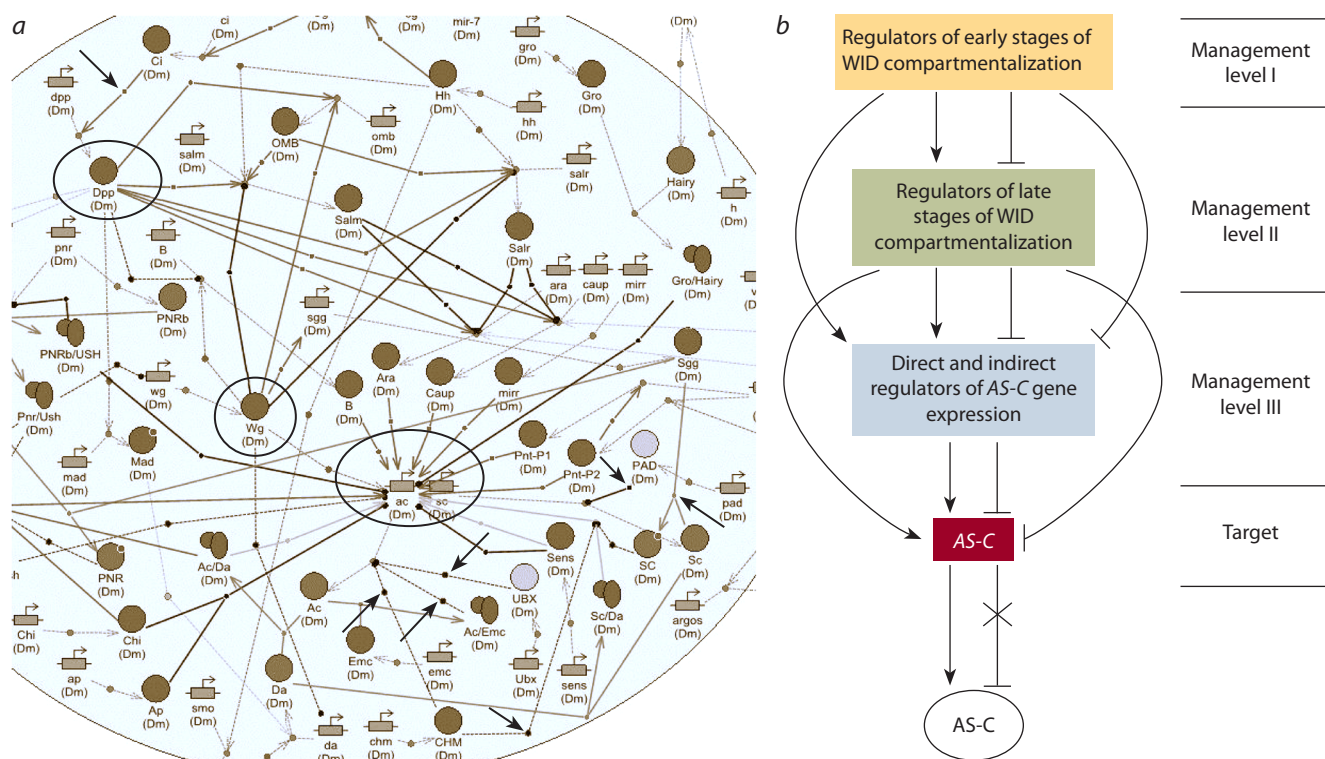


Fig. 1. A fragment and schematic organization of the Neurogenesis:Prepattern gene network.

(a) Key components of the network: proteins Dpp, Wg, and the *achaete-scute* gene complex (enclosed in ovals). Black arrows indicate the points of application of indirect regulatory impacts on the complex expression. (b) Hierarchical organization of the Neurogenesis:Prepattern gene network and the scheme of interactions between management levels. WID, wing imaginal disk. Here and in Figs. 2 and 3 sharp-ended arrows → indicate activation, blunt-ended ⊥, repression.

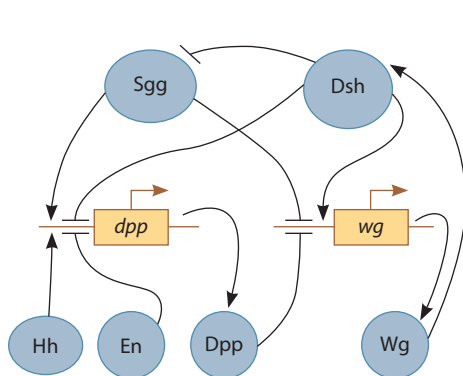


Fig. 2. Regulation of the *dpp* and *wg* genes.

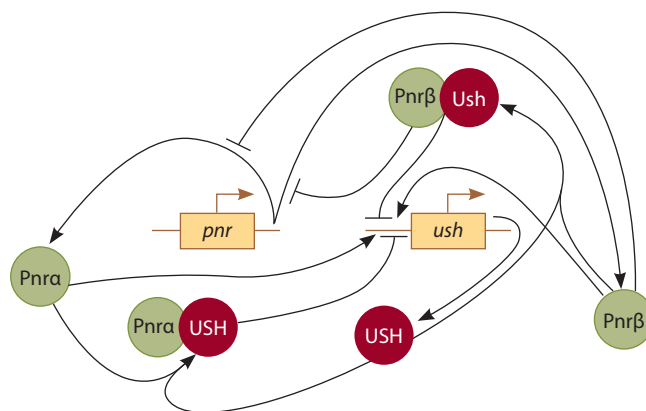


Fig. 3. Regulatory circuit determining the balance between proteins Pnr and Ush.

руют транскрипцию *u-shaped* (Hannon et al., 1991; Tsai et al., 1991; Ramain et al., 1993; Cubadda et al., 1997; Haenlin et al., 1997; Fromental-Ramain et al., 2008, 2010).

На третьем уровне управления сосредоточены элементы контроля за функционированием основного ключевого объекта сети с максимальной связностью – пронейральных генов комплекса *achaete-scute*.

Активность объектов этого уровня регулируется элементами предыдущего. Так, *spalt/spalt-related* участвуют в каскадной регуляции, кодируя транскрипционные факторы, репрессирующие гены *Iro-C* (de Celis et al., 1999). Pannier активирует, а комплекс Pnr/Ush репрессирует

экспрессию *Iro-C* (Calleja et al., 2000). Ограничение зон экспрессии *Iro-C* – один из обязательных факторов, обеспечивающий правильное позиционирование пронейральных кластеров. Транскрипционный фактор Brinker репрессирует ген *tailup* (*tup*), а одноименный белок, в свою очередь, репрессирует активность *AS-C* в одном районе нотума (будущем щитке), но активирует в другом (дорзотомальном нотуме) (de Navascués, Modolell, 2007, 2010).

Основные объекты третьего уровня представлены регуляторами активности *AS-C*. Прежде всего это транскрип-

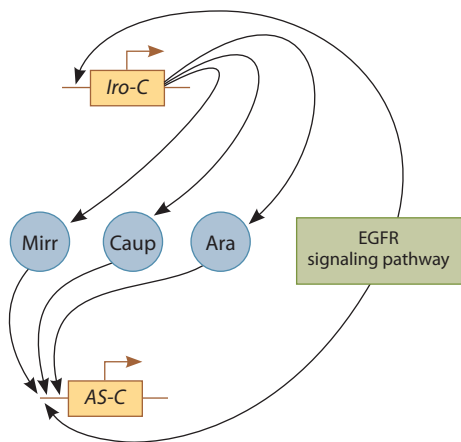


Fig. 4. EGFR signaling pathway is an activator of gene complexes *achaete-scute* и *iroquois*.

ционные факторы – активаторы/репрессоры экспрессии, места посадки которых в энхансерных и промоторных зонах комплекса создают «рисунок», определяющий функциональное состояние его генов. В качестве регуляторов могут выступать не только самостоятельные транскрипционные факторы, имеющие сайты связывания в регуляторных областях генов *achaete-scute*, но и гетерогенные белковые комплексы с ДНК-связывающей активностью, сформированные из белков, принадлежащих к семействам bHLH и Zinc-finger. Функцию адаптера в этих случаях выполняет белок Chip (Romain et al., 2000; Matthews, Visvader, 2003; de Navascués, Modolell, 2007; Bronstein et al., 2010).

Транскрипция *AS-C* инициируется транскрипционными факторами семейства NF-kB/Rel (Ayyar et al., 2007, 2010), белками комплексов *iroquois* (Araucan, Caupolican и Mirror) (Leyns et al., 1996) и Bar (BarH1 и BarH2) (Sato et al., 1999), а также дуэтом Chip/Pnr в составе гетеромера (de Navascués, Modolell, 2007; Bronstein et al., 2010).

Репрессию транскрипции *AS-C* осуществляет транскрипционный фактор Hairy (Gómez-Skarmeta et al., 1995) и гексамер 2Chip/2AP/2SSDP, в состав которого наряду с Chip и Ap входит SSDP (Sequence-specific single-stranded DNA-binding protein) (Bronstein et al., 2010).

Особняком в ряду регуляторов стоит Tailup – он может оказывать как активирующее, так и репрессирующее влияние на экспрессию *AS-C*. В составе сложных гексамеров, содержащих также белки Chip и SSDP, Tur выполняет функцию активатора экспрессии *AS-C* в дорзоцентральной области презумптивного нотума (de Navascués, Modolell, 2007, 2010). Однако в составе других белковых комплексов Tur подавляет развитие щетинок в области презумптивного щитка. Точный механизм этого эффекта до сих пор не выяснен (de Navascués, Modolell, 2007, 2010).

Регуляция функциональной активности *AS-C* может происходить за счет прямых связей ДНК-транскрипционного фактора в регуляторном районе комплекса, но может обеспечиваться и другими механизмами. Одним из них является изменение архитектуры регуляторных зон комплекса и взаимодействий «энхансер–промотор» в результате ремоделирования хроматина. На этом механизме

базируется активаторный эффект белка Toutatis (Tou) (Heitzler et al., 2003) и репрессорный эффект белка Osa (Vanolst et al., 2005).

Другой механизм основан на модификационном изменении транскрипционных факторов, контролирующих активность *AS-C*. Так, фосфорилирование Pnr киназой Sgg приводит к утрате способности Pnr инициировать транскрипцию генов комплекса (Yang et al., 2012). Ацетилтрансфераза Chameau (CHM), модифицирующая Pnr, Ush и Chip (Romain et al., 2000), обладает функцией репрессора активности *AS-C*, но фокус ее действия неизвестен (Hainaut et al., 2012; Nakagawa et al., 2015).

Наряду с рассмотренными отношениями контроля элементами вышележащих уровней элементов соседних нижележащих уровней, формирующих вертикальный вектор управления, в системе существуют **регуляторные взаимодействия иного характера**.

В частности, возможны взаимодействия, нарушающие строгую иерархичность управления. Отмечены случаи прямой связи между объектами первого и третьего уровней, когда Dpp и Wg активируют соответственно *tup*, *bar*, *Iro-C* и *sgg*. Другим примером такого рода служит управление транскрипцией ключевого элемента генной сети (*AS-C*) объектами второго уровня – транскрипционными факторами Pnr и Ush.

Нарушение принципа регуляции «сверху вниз» может происходить и благодаря участникам сети, которые присутствуют более чем на одном уровне управления генами *AS-C* и играют вспомогательную роль в функционировании сети. К ним относятся киназа Sgg и адаптер Chip.

Белки сигнальных путей занимают особое место в функционировании генных сетей, поскольку осуществляют межклеточные взаимодействия, необходимые для инициации и регуляции процессов, определяющих морфофункциональное состояние клеток. В становлении предструктуры щетиночного узора заметную роль играет EGFR-сигнальный путь. Конечной мишенью передачи сигнала являются гены *Iro-C* или *AS-C*. Активация генов этих комплексов зависит от разных лигандов-трансдукторов. Для генов *Iro-C* это Vein, а для генов *AS-C* – Spitz (рис. 4).

Таким образом EGFR-сигнальный путь является активатором как объекта третьего уровня управления *Iro-C*, так и подчиненного белкам этого комплекса *AS-C* (Gómez-Skarmeta et al., 2003; Shilo, 2003).

Уровни регуляции могут взаимодействовать по принципу не только прямой, но и обратной связи, когда элементы нижележащих уровней определяют состояние объектов, расположенных на вышележащих уровнях. В рассматриваемой сети обнаружено два таких случая. Первый – это влияние изоформ Pnr и Ush на экспрессию *wg*, а второй – влияние Tur на экспрессию *pnr* и *ush* (рис. 5).

Наличие регуляторных контуров – одна из характеристик генных сетей. Как отмечено выше, в сети Neurogenesis:Prepattern контуры выявляются не только в пределах одного уровня. По крайней мере два контура связывают элементы разных уровней. Первый контур представлен белками Dpp, Brk, Tur и генами *brk*, *pnr*, *ush*, *salm/salr* и *tup* (см. рис. 5, а). В рамках контура Dpp репрессирует ген *brk* и активирует гены *pnr* и *tup*, а белок Tur, в свою

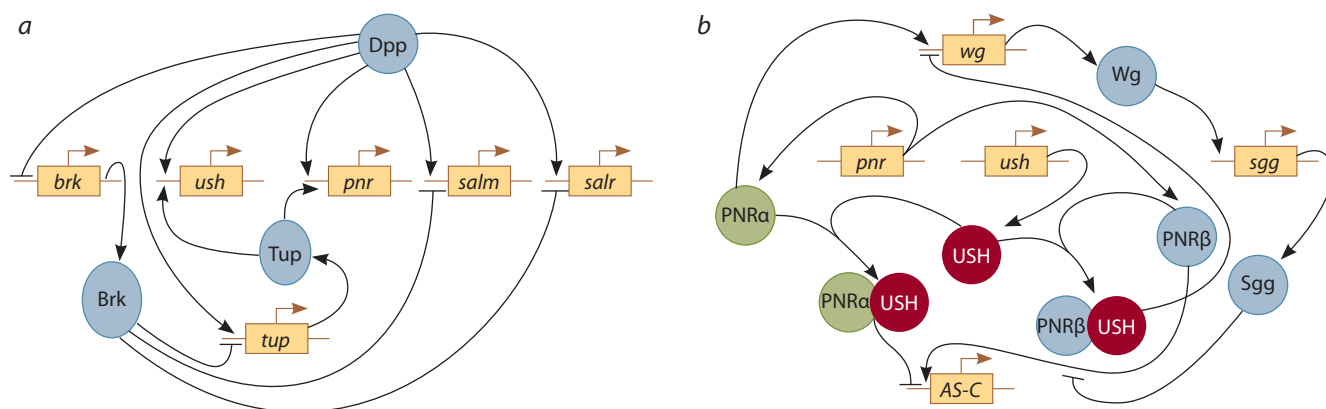


Fig. 5. Interactions between regulators at different management levels: (a) Interactions between the regulators of the first (Dpp), second (*brk*, *pnr*, *ush*, *salm*, and *salr*) and third (Tup) management levels. (b) Interactions between regulators of the first (Wg) and second (*sgg*, *pnr*, and *ush*) management levels and AS-C.

очередь, усиливает экспрессию *pnr*. Функционирование контура обеспечивает достаточное содержание в клетках транскрипционных факторов Pnr, Ush, Tup при одновременном отсутствии Brk и завершает морфофункциональную компартментализацию имагинального диска установлением района презумптивного нотума.

Во второй контур входят белки Wg, обе изоформы Pnr, Ush, Sgg и гены *AS-C*, *sgg* (см. рис. 5, б). В зависимости от преобладания в клетках той или иной изоформы Pnr происходит активация *AS-C* (при преобладании Pnrβ) или *wg* (при преобладании Pnrα). Одновременно образуются гетеромеры изоформ с Ush, негативно влияющие на транскрипцию *AS-C* (комплекс Pnrα/Ush) или *wg* (комплекс Pnrβ/Ush). Нарботка Wg приводит к активации *sgg*, а киназа Sgg фосфорилирует Pnrβ, преобразуя его в неактивную форму. Таким образом, превышение содержания Pnrα над содержанием Pnrβ в клетках имагинального диска закрывает для них нейтральный путь дифференцировки, полностью исключая возможность активации генов *AS-C*. Этот контур функционирует в пределах презумптивного нотума и обеспечивает уточнение взаимного расположения пронеуральных кластеров (Sato, Saigo, 2000; Fromental-Ramain et al., 2008, 2010; Yang et al., 2012).

Закключение

Постоянство щетиночного рисунка имаго дрозофилы достигается точностью числа и локализации в имагинальных дисках пронеуральных кластеров – районов расположения будущих макрохет. Топографию кластеров определяет молекулярно-генетическая система, контролирующая процесс их закладки с самых ранних стадий развития диска путем распределения специфичных транскрипционных факторов, отвечающих за экспрессию генов комплекса *achaete-scute*. Функционирование этой системы может быть описано в терминах генной сети Neurogenesis:Prepattern. Фокусом ее действия является *AS-C* – объект с максимальной связностью.

Рассмотрение совокупности элементов сети и их взаимодействий позволяет сделать вывод, что в рамках парадигмы «предструктура – система ответа на предструктуру» понятие предструктуры, видимо, должно быть расширено

и должно включать не только набор транскрипционных факторов, непосредственно контролирующих экспрессию *AS-C* как системы ответа на предструктуру по Штерну (белки Iro-C, Bar, Tup, Hair, Pnr и Pnr/Ush, NF-kB/Rel, Ap, Pointed – белок EGFR-пути).

Сюда же следует отнести белки, определяющие или модифицирующие как функциональное состояние этих транскрипционных факторов (Sgg, Chip, Shm, SSDP), так и архитектуру регуляторных районов *AS-C* (Osa, Tou) (рис. 6), а также набор регуляторов экспрессии самих факторов предструктуры в существующем на сегодня понимании этого термина (Dpp, Wg, Hh, En, Brk, Salm/Salr, белки EGFR-пути).

Neurogenesis:Prepattern, как и другие генные сети морфогенеза, демонстрирует каскадную регуляцию процессов и сопряженную с этим иерархическую организацию. Главный вектор действия сети ориентирован на наработку и поддержание уровня пронеуральных белков AS-C в клетках кластера, обеспечивающих формирование макрохет как органов периферической нервной системы дрозофилы. Механизм каскадной регуляции комплекса *AS-C* включает не менее трех уровней управления, каждый из которых содержит регуляторные контуры, детерминирующие состояние активации–репрессии его генов. Именно от правильного функционирования комплекса зависит приобретение клетками потенциалов к нейральному пути развития.

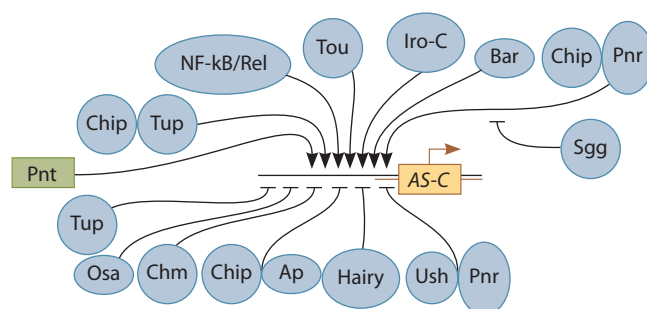


Fig. 6. Regulators of the AS-C complex.

Acknowledgments

Database provisioning was supported by State Budgeted Project 0324-2015-0003. The gene network reconstruction was supported by the Russian Science Foundation, project 14-24-00123.

Conflict of interest

The authors declare no conflict of interest.

References

- Affolter M., Basler K. The Decapentaplegic morphogen gradient: from pattern formation to growth regulation. *Nat. Rev. Genet.* 2007;8(9): 663-674.
- Akiyama T., Gibson M.C. Decapentaplegic and growth control in the developing *Drosophila* wing. *Nature*. 2015;527(7578):375-8. DOI 10.1038/nature15730.
- Aldaz S., Morata G., Azpiazu N. The Pax-homeobox gene *eyegone* is involved in the subdivision of the thorax of *Drosophila*. *Development*. 2003;130:4473-4482.
- Ayyar S., Negre B., Simpson P., Stollewerk A. An arthropod cis-regulatory element functioning in sensory organ precursor development dates back to the Cambria. *BMC Biol.* 2010;24:127. DOI 10.1186/1741-7007-8-127.
- Ayyar S., Pistillo D., Calleja M., Brookfield A., Gittins K., Goldstone C., Simpson P. NF- κ B/Rel-mediated regulation of the neural fate in *Drosophila*. *PLoS One*. 2007;2:e1178.
- Bronstein R., Levkovitz L., Yosef N., Yanku M., Ruppin E., Sharan R., Westphal H., Oliver B., Segal D. Transcriptional regulation by CHIP/LDB complexes. *PLoS Genetics*. 2010;6:e1001063. DOI 10.1371/journal.pgen.1001063.
- Bukharina T.A., Furman D.P. Genetical control of mechanoreceptors formation in *Drosophila melanogaster* – the description in database “Neurogenesis”. *Informatsionnyy vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders*. 2009;13(1):186-193 (in Russian).
- Calleja M., Herranz H., Estella C., Casal J., Lawrence P., Simpson P., Morata G. Generation of medial and lateral dorsal body domains by the *pannier* gene of *Drosophila*. *Development*. 2000;127:3971-3980.
- Calleja M., Renaud O., Usui K., Pistillo D., Morata G., Simpson P. How to pattern an epithelium: lessons from *achaete-scute* regulation on the notum of *Drosophila*. *Gene*. 2002;292(1-2):1-12.
- Campbell G., Tomlinson A. Transducing the Dpp morphogen gradient in the wing of *Drosophila*: regulation of Dpp targets by *brinker*. *Cell*. 1999;96(4):553-562.
- Cubadda Y., Heitzler P., Ray R.P., Bourouis M., Romain P., Gelbart W., Simpson P., Haenlin M. *u-shaped* encodes a zinc finger protein that regulates the proneural genes *achaete* and *scute* during the formation of bristles in *Drosophila*. *Genes Dev.* 1997;11:3085-3095.
- Cubas P., de Celis J.F., Campuzano S., Modolell J. Proneural clusters of *achaete-scute* expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. *Genes Dev.* 1991;5:996-1008.
- de Celis J.F., Barrio R., Kafatos F.C. Regulation of the *spalt/spalt*-related gene complex and its function during sensory organ development in the *Drosophila* thorax. *Development*. 1999;126:2653-2662.
- de Navascués J., Modolell J. *tailup*, a LIM-HD gene, and Iro-C cooperate in *Drosophila* dorsal mesothorax specification. *Development*. 2007;134:1779-1788.
- de Navascués J., Modolell J. The pronotum LIM-HD gene *tailup* is both a positive and a negative regulator of the proneural genes *achaete* and *scute* of *Drosophila*. *Mech. Development*. 2010;127:393-406. DOI 10.1016/j.mod.2010.05.001.
- Fromental-Romain C., Taquet N., Romain P. Transcriptional interactions between the *pannier* isoforms and the cofactor U-shaped during neural development in *Drosophila*. *Mech. Development*. 2010;127:442-457. DOI 10.1016/j.mod.2010.08.002.
- Fromental-Romain C., Vanolst L., Delaporte C., Romain P. *pannier* encodes two structurally related isoforms that are differentially expressed during *Drosophila* development and display distinct functions during thorax patterning. *Mech. Development*. 2008;125:43-57.
- Furman D.P., Bukharina T.A. Genetic control of macrochaetae development in *Drosophila melanogaster*. *Rus. J. Dev. Biol.* 2008;39(4): 195-206.
- Gómez-Skarmeta J.L., Campuzano S., Modolell J. Half a century of neural prepattern: the story of a few bristles and many genes. *Nat. Rev. Neurosci.* 2003;4:587-598.
- Gómez-Skarmeta J.L., Díez del Corral R., de la Calle-Mustienes E., Ferrés-Marcó D., Modolell J. *araucan* and *caupolican*, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell*. 1996;85:95-105.
- Gómez-Skarmeta J.L., Rodríguez I., Martínez C., Culi J., Ferrés-Marcó D., Beamonte D., Modolell J. Cis-regulation of *achaete* and *scute*: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev.* 1995;9:2598-2608.
- Haenlin M., Cubadda Y., Blondeau F., Heitzler P., Lutz Y., Simpson P., Romain P. Transcriptional activity of *pannier* is regulated negatively by heterodimerization of the GATA DNA-binding domain with a cofactor encoded by the *u-shaped* gene of *Drosophila*. *Genes Dev.* 1997;11:3096-3108.
- Hainaut M., Sagnier T., Berenger H., Pradel J., Graba Y., Miotto B. The MYST-containing protein Chameau is required for proper sensory organ specification during *Drosophila* thorax morphogenesis. *PLoS One*. 2012;7(3):e32882. DOI 10.1371/journal.pone.0032882.
- Hannon R., Evans T., Felsenfeld G., Gould H. Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA*. 1991;88:3004-3008.
- Heitzler P., Vanolst L., Biryukova I., Romain P. Enhancer-promoter communication mediated by Chip during *Pannier*-driven proneural patterning is regulated by Osa. *Genes Dev.* 2003;17:591-596.
- Heslin T.R., Theisen H., Walker H., Marsh J.L. SHAGGY and DISHEVELLED exert opposite effects on *wingless* and *decapentaplegic* expression and on positional identity in imaginal discs. *Development*. 1997;124:1069-1078.
- Ikmi A., Netter S., Coen D. Prepatterning the *Drosophila* notum: the three genes of the *iroquois* complex play intrinsically distinct roles. *Dev. Biol.* 2008;317:634-648.
- Joshi M., Buchanan K.T., Shroff S., Orenic T.V. Delta and Hairy establish a periodic prepattern that positions sensory bristles in *Drosophila* legs. *Dev. Biol.* 2006;293:64-76.
- Kolchanov N.A., Ananko E.A., Kolpakov F.A., Podkolodnaya O.A., Ignatieva E.V., Goryachkovskaya T.N., Stepanenko I.L. *Gene Networks. Molekulyarnaya biologiya = Molecular Biology (Moscow)*. 2000;34(4):533-544 (in Russian).
- Komiya Y., Habas R. Wnt signal transduction pathways. *Organogenesis*. 2008;4(2):68-75.
- Lai E.C., Orgogozo V. A hidden program in *Drosophila* peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity. *Dev. Biol.* 2004;269(1):1-17.
- Layalle S., Volovitch M., Mugat B., Bonneaud N., Parmentier M.L., Prochiantz A., Joliot A., Maschat F. Engrailed homeoprotein acts as a signaling molecule in the developing fly. *Development*. 2011;138: 2315-2323. DOI 10.1242/dev.057059.
- Leyns L., Gómez-Skarmeta J.L., Dambly-Chaudière C. *iroquois*: a pre-pattern gene that controls the formation of bristles on the thorax of *Drosophila*. *Mech. Development*. 1996;59(1):63-72.
- Martin F.A., Pérez-Garijo A., Moreno E., Morata G. The *brinker* gradient controls wing growth in *Drosophila*. *Development*. 2004; 131(20):4921-4930.
- Matthews J.M., Visvader J.E. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep.* 2003;4:1132-1137.
- Nakagawa T., Ikehara T., Doiguchi M., Imamura Y., Higashi M., Yoneda M., Ito T. Enhancer of Acetyltransferase Chameau (EACHM) is a novel transcriptional co-activator. *PLoS One*. 2015;10(11):e0142305. DOI 10.1371/journal.pone.0142305.

- Penton A., Hoffmann F.M. Decapentaplegic restricts the domain of wingless during *Drosophila* limb patterning. *Nature*. 1996;382:162-165.
- Ramain P., Heitzler P., Haenlin M., Simpson P. *pannier*, a negative regulator of *achaete* and *scute* in *Drosophila*, encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. *Development*. 1993;119:1277-1291.
- Ramain P., Khechumian R., Khechumian K., Arbogast N., Ackermann C., Heitzler P. Interactions between Chip and the Achaete/Scute-Daughterless heterodimers are required for pannier-driven proneural patterning. *Mol. Cell*. 2000;6:781-790.
- Reeves N., Posakony J.W. Genetic programs activated by proneural proteins in the developing *Drosophila* PNS. *Dev. Cell*. 2005;8:413-425.
- Romani S., Campuzano S., Macagno E.R., Modolell J. Expression of *achaete* and *scute* genes in *Drosophila* imaginal discs and their function in sensory organ development. *Genes Dev*. 1989;3: 997-1007.
- Ruiz-Gómez M., Modolell J. Deletion analysis of the *achaete-scute* locus of *Drosophila melanogaster*. *Genes Dev*. 1987;1:1238-1246.
- Sato M., Kojima T., Michiue T., Saigo K. *Bar* homeobox genes are latitudinal prepattern genes in the developing *Drosophila* notum whose expression is regulated by the concerted functions of *decapentaplegic* and *wingless*. *Development*. 1999;26(7):1457-1466.
- Sato M., Saigo K. Involvement of *pannier* and *u-shaped* in regulation of Decapentaplegic-dependent *wingless* expression in developing *Drosophila* notum. *Mech. Development*. 2000;93(1-2):127-138.
- Serebrovsky A.S., Dubinin N.P., The artificial production of mutations and the gene problem. *Uspekhi eksperimentalnoy biologii* = Adventures of Experimental Biology. 1929;4:235-247 (in Russian).
- Serrano N., O'Farrell P.H. Limb morphogenesis: connections between patterning and growth. *Curr. Biol*. 1997;7(3):R186-195.
- Shilo B.Z. Signaling by the *Drosophila* epidermal growth factor receptor pathway during development. *Exp. Cell Res*. 2003;284:140-149.
- Skeath J.B., Carroll S.B. Regulation of *achaete-scute* gene expression and sensory organ pattern formation in the *Drosophila* wing. *Genes Dev*. 1991;5:984-995.
- Stern C. Two or three bristles. *Am. Sci*. 1954;42:213-247.
- Tsai S.F., Strauss E., Orkin S.H. Functional analysis and *in vivo* footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes Dev*. 1991;5:919-931.
- Vanolst L., Fromental-Ramain C., Ramain P. Toutatis, a TIP5-related protein, positively regulates Pannier function during *Drosophila* neural development. *Development*. 2005;132:4327-4338.
- Yang M., Hatton-Ellis E., Simpson P. The kinase Sgg modulates temporal development of macrochaetes in *Drosophila* by phosphorylation of Scute and Pannier. *Development*. 2012;139:325-334. DOI 10.1242/dev.074260.
- Zecca M., Basler K., Struhl G. Sequential organizing activities of *engrailed*, *hedgehog* and *decapentaplegic* in the *Drosophila* wing. *Development*. 1995;121:2265-2278.

Молекулярно-генетические механизмы взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открытоугольной глаукоме

О.В. Сайк¹✉, Н.А. Коновалова², П.С. Деменков¹, Н.В. Иванисенко^{1,3}, Т.В. Иванисенко¹, Д.Е. Иванощук¹,
О.С. Коновалова², О.А. Подколотная¹, И.Н. Лаврик^{1,4}, Н.А. Колчанов¹, В.А. Иванисенко¹

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное бюджетное образовательное учреждение высшего образования «Тюменский государственный медицинский университет» Министерства здравоохранения Российской Федерации, Тюмень, Россия

³ Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

⁴ Магдебургский университет имени Отто фон Гуерике, Магдебург, Германия

Глаукома – хроническое, прогрессирующее заболевание, которым страдают более 60 млн человек в мире. Первичная открытоугольная глаукома (ПОУГ) занимает одно из первых мест по распространенности среди различных форм глаукомы. Например, в 2011 г. эта форма заболевания наблюдалась более чем у 2.7 млн человек в США. В настоящее время ПОУГ является основной причиной необратимой потери зрения. У больных с открытоугольной глаукомой риск слепоты достигает 27 %, даже несмотря на проводимое лечение. Известно, что гибель клеток зрительного нерва может быть спровоцирована механическим стрессом, вызванным повышенным внутриглазным давлением, наблюдающимся при ПОУГ, индуцирующим нейрональный апоптоз. В настоящее время существует огромное количество научных публикаций, описывающих белки и гены, которые участвуют в патогенезе ПОУГ, в том числе в нейрональном апоптозе и клеточном ответе на механический стресс. Тем не менее молекулярно-генетические механизмы, лежащие в основе патофизиологии ПОУГ, до сих пор плохо изучены. Реконструкция ассоциативных генных сетей, описывающих функциональные взаимодействия между этими генами/белками, включая биохимические реакции, регуляторные взаимодействия, события транспорта и т. д., требует использования методов автоматизированного извлечения знаний из текстов научных публикаций. В работе проанализированы ассоциативные сети, описывающие молекулярно-генетические взаимодействия между белками и генами, вовлеченными в ответ клетки на механический стресс (ОКМС), нейрональный апоптоз и патогенез ПОУГ. Показано, что гены, ассоциированные с ПОУГ, статистически значимо чаще, чем ожидалось по случайным причинам, представлены среди генов, участвующих во взаимодействии между ОКМС и нейрональным апоптозом, что может быть важным фактором, влияющим на гибель ганглиозных ретинальных клеток при ПОУГ.

Ключевые слова: апоптоз; нейрональный апоптоз; ответ клетки на механический стресс; первичная открытоугольная глаукома; ПОУГ; ANDSystem; ассоциативные генные сети.

Molecular-genetic mechanisms of the interaction between processes of cell response to mechanical stress and neuronal apoptosis in primary open-angle glaucoma

O.V. Saik¹✉, N.A. Konovalova²,
P.S. Demenkov¹, N.V. Ivanisenko^{1,3},
T.V. Ivanisenko¹, D.E. Ivanoshchuk¹,
O.S. Konovalova², O.A. Podkolodnaya¹,
I.N. Lavrik^{1,4}, N.A. Kolchanov¹, V.A. Ivanisenko¹

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Tyumen State Medical University, Tyumen, Russia

³ Novosibirsk State University, Novosibirsk, Russia

⁴ Institute of Experimental Internal Medicine, Otto von Guericke University, Magdeburg, Germany

Glaucoma is a chronic and progressive disease, which affects more than 60 million people worldwide. Primary open-angle glaucoma (POAG) is one of the most common forms of glaucoma. For example, about 2.71 million people in the USA had primary open-angle glaucoma in 2011. Currently POAG is a major cause of irreversible vision loss. In patients with treated open-angle glaucoma the risk of blindness reached to be about 27 %. It is known that the death of optic nerve cells can be triggered by mechanical stress caused by increased intraocular pressure, which induces neuronal apoptosis and is observed in patients with POAG. Currently, there is a large number of scientific publications describing proteins and genes involved in the pathogenesis of POAG, including neuronal apoptosis and the cell response to mechanical stress. However, the molecular-genetic mechanisms underlying the pathophysiology of POAG are still poorly understood. Reconstruction of associative networks describing the functional interactions between these genes/proteins, including biochemical reactions, regulatory interactions, transport, etc., requires the use of methods of automated knowledge

extraction from texts of scientific publications. The aim of the work was the analysis of associative networks, describing the molecular-genetic interactions between proteins and genes involved in cell response to mechanical stress (CRMS), neuronal apoptosis and pathogenesis of POAG using ANDSystem, our previous development for automated text analysis. It was shown that genes associated with POAG are statistically significantly more often represented among the genes involved in the interactions between CRMS and neuronal apoptosis than it was expected by random reasons, which can be an explanation for the effect of POAG leading to the retinal ganglion cell death.

Key words: apoptosis; neuronal apoptosis; cell response to mechanical stress; primary open-angle glaucoma; POAG; ANDSystem; associative gene networks.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Сайк О.В., Коновалова Н.А., Деменков П.С., Иванисенко Н.В., Иванисенко Т.В., Иваношук Д.Е., Коновалова О.С., Подколотная О.А., Лаврик И.Н., Колчанов Н.А., Иванисенко В.А. Молекулярно-генетические механизмы взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открытоугольной глаукоме. Вавиловский журнал генетики и селекции. 2016;20(6):840-847. DOI 10.18699/VJ16.200

HOW TO CITE THIS ARTICLE:

Saik O.V., Konvalova N.A., Demenkov P.S., Ivanisenko N.V., Ivanisenko T.V., Ivanoshchuk D.E., Konvalova O.S., Podkolodnaya O.A., Lavrik I.N., Kolchanov N.A., Ivanisenko V.A. Molecular-genetic mechanisms of the interaction between processes of cell response to mechanical stress and neuronal apoptosis in primary open-angle glaucoma. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):840-847. DOI 10.18699/VJ16.200

Первичная открытоугольная глаукома (ПОУГ) – одна из наиболее часто встречающихся форм глаукомы, характеризующаяся прогрессирующей гибелью ганглионарных клеток сетчатки. Гибель клеток зрительного нерва сопровождается характерным сужением полей зрения с последующей необратимой потерей зрения. Механизм гибели ганглионарных клеток до сих пор в деталях не изучен, однако предполагается, что ключевая роль в этом процессе принадлежит апоптозу (Wang et al., 2013). Апоптоз, или запрограммированная клеточная гибель, играет важную роль в различных генетико-физиологических процессах, в том числе в процессах обновления клеток, развития и функционирования иммунной системы, а также эмбрионального развития (Kerr et al., 1972). Считается, что существуют два основных пути апоптоза: индуцируемый через рецепторы клеточной смерти и митохондриальный. Нарушения функционирования как митохондриального пути, так и пути индукции апоптоза через рецепторы, являются важным молекулярно-генетическим механизмом развития многих заболеваний, таких как нейродегенеративные заболевания, ишемические повреждения, аутоиммунные расстройства, различные виды рака и др. (Elmore, 2007). В большинстве случаев ПОУГ характеризуется повышением внутриглазного давления, в результате которого ганглиозные клетки сетчатки испытывают постоянный механический стресс. В настоящее время полагают, что ответ клетки на механический стресс (ОКМС), наряду с оксидативным стрессом и опосредованной гиперактивацией глутаматных рецепторов, может быть основным фактором, индуцирующим нейрональный апоптоз при глаукоме (Calandrella et al., 2007). Однако повышенное внутриглазное давление хотя и является одним из наиболее важных факторов риска развития глаукомы, не всегда приводит к ней (Matsopoulos et al., 2008). Современные методы лечения глаукомы направлены на

снижение внутриглазного давления, в том числе посредством оперативного или лазерного вмешательства и лекарственного воздействия, но, к сожалению, не позволяют полностью остановить патологический процесс и могут только замедлить потерю функции глаза (Quigley, 1999, 2011; Almasieh et al., 2012; Jindal, 2013). Можно предположить, что нарушение функции генов, ассоциированных с глаукомой, оказывает влияние на взаимодействие между процессами ОКМС и нейронального апоптоза.

Цель работы – анализ ассоциативных сетей, описывающих молекулярно-генетические взаимодействия между белками и генами, ассоциированными с ПОУГ, а также участвующими в процессах ОКМС и нейрональном апоптозе. Выявление генов, вовлеченных в такие взаимодействия, может быть основой для разработки лекарственных препаратов, корректирующих молекулярно-генетические нарушения, вызывающие повышенный апоптоз ганглионарных клеток сетчатки при ПОУГ.

Материалы и методы

Выявление генов, ассоциированных с ПОУГ (согласно литературным данным), а также реконструкция ассоциативных сетей, описывающих молекулярно-генетические взаимодействия между этими генами и белками, и генами, участвующими в процессах ОКМС и нейронального апоптоза, осуществлялись с использованием системы ANDSystem (Demenkov et al., 2012; Ivanisenko et al., 2015). Эта система была разработана для автоматического анализа научных публикаций с целью извлечения знаний о молекулярно-генетических взаимодействиях и ассоциациях белков, генов, метаболитов, лекарств, микроРНК с заболеваниями, биологическими процессами, побочными эффектами лекарств и фенотипическими признаками различных организмов. ANDSystem была построена на основе широкомасштабного анализа более 25 млн аннотаций

научных статей, представленных в базе данных PubMed. Кроме того, в базу знаний интегрирована информация о молекулярно-генетических взаимодействиях из фактографических баз данных, таких как IntAct, MINT и др. Всего в базе знаний ANDSystem представлено более 7 млн фактов о молекулярно-генетических взаимодействиях и ассоциациях. Дополнительно для выявления генов, ассоциированных с ПОУГ, использовалась информация из баз данных OMIM (<http://www.omim.org/>), ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>), GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) и SNPedia (<http://www.snpedia.com/index.php/SNPedia>), полученная по запросу «Primary open angle glaucoma». Списки генов, вовлеченных в нейрональный апоптоз (GO:0051402) и ОКМС (GO:0071260), взяты из базы данных Gene Ontology annotation (<http://geneontology.org/>).

Для оценки статистической значимости сверхпредставленности набора генов в анализируемой выборке с помощью гипергеометрического распределения использовалась программа, доступная по адресу: <https://www.geneprof.org/GeneProf/tools/hypergeometric.jsp>.

Анализ сверхпредставленности Gene Ontology биологических процессов, а также выявление генов, экспрессирующихся в мозге, проводились веб-сервисом DAVID (<https://david-d.ncicrf.gov/>), все настройки использовались по умолчанию.

Результаты и обсуждение

Использование системы ANDSystem позволило выявить 68 генов, ассоциированных с ПОУГ, согласно литературным данным. Среди этого набора для 35 генов приводились данные о наличии полиморфизмов, ассоциированных с ПОУГ. Остальные гены авторы статей ассоциировали с ПОУГ на основе изменения их функции или экспрессии при данном заболевании. С помощью анализа баз данных OMIM, ClinVar, GWAS Catalog и SNPedia дополнительно были выявлены 12, 2, 10 и 9 генов соответственно. Списки генов, вовлеченных в ОКМС и нейрональный апоптоз, содержали 80 и 213 генов соответственно. Оказалось, что среди 213 генов, участвующих в нейрональном апоптозе, представлены 10 генов (*HIF1A*, *CTNNB1*, *FAS*, *GSK3B*, *FADD*, *HDAC4*, *MAP2K4*, *BNIP3*, *CASP2*, *HDAC7*), вовлеченные также в ОКМС. Известно, например, что экспрессия гена *HIF1A* увеличивается при механическом стрессе (Fu-lan et al., 2012), в свою очередь повышенная экспрессия гена *HIF1A* приводит к активации апоптоза (Park et al., 2014). В работе (Dorn, Kirshenbaum, 2008) показано, что при механическом стрессе, вызванном сдавливанием, активируется транскрипция гена *BNIP3*. Кроме того, известно, что подавление активности *BNIP3* оказывает протективное действие по отношению к апоптозу предшественников нервных клеток (Wang et al., 2013). Анализ сверхпредставленности показал, что эти 10 генов статистически значимо сверхпредставлены среди генов, участвующих в нейрональном апоптозе ($p\text{-value} < 10^{-7}$).

Механический стресс может оказывать также влияние на функцию генов апоптоза путем прямых межмолекулярных взаимодействий. С помощью системы ANDSystem было реконструировано 506 взаимодействий между белками, вовлеченными в ответ клетки на механический стресс,

и белками, участвующими в нейрональном апоптозе. Рассматривались такие типы взаимодействий, как регуляция активности, регуляция деградации, регуляция экспрессии, белок-белок взаимодействия и регуляция транспорта. Наибольшее число связей (410) наблюдалось для белок-белок взаимодействий. Интересно, что число взаимодействий типа «expression upregulation» примерно в четыре раза превышало число взаимодействий типа «expression downregulation» и составило 15 и 4 связи соответственно. В качестве примера положительной регуляции экспрессии можно привести белок *GSK3b* и ген *Bim*. Известно, что подавление активности белка *GSK3b*, наблюдающееся при ОКМС (Badorff et al., 2002), активирует экспрессию гена *Bim* (McCubrey et al., 2014), вовлеченного в активацию нейронального апоптоза (Zhao et al., 2016). Кроме того, оказалось, что белки, вовлеченные в ОКМС, способны регулировать деградацию белков нейронального апоптоза по типу связи «degradation downregulation». Например, белок *JNK1*, который активируется под влиянием механического стресса (Zhou et al., 2007), способен увеличивать стабильность белка *P73* (Ozaki et al., 2010), являющегося в нейронах проапоптотическим белком (Klein et al., 2011). Из выявленных взаимодействий видно, что регуляция белков и генов нейронального апоптоза со стороны ответа клетки на механический стресс может быть направлена на активацию процесса апоптоза посредством положительной регуляции экспрессии генов и стабильности белков, вовлеченных в нейрональный апоптоз, что согласуется с известными данными (Calandrella et al., 2007; Slemmer et al., 2008; Uchida et al., 2008). Также следует отметить, что из 213 генов/белков, участвующих в нейрональном апоптозе, 118 были вовлечены во взаимодействие с белками ОКМС, что существенно превышало значение данной величины, ожидаемое по случайным причинам. Согласно гипергеометрическому распределению, перепредставленность участников взаимодействий с белками ОКМС среди генов/белков нейронального апоптоза имела статистическую значимость ($p\text{-value} < 10^{-40}$).

Хорошо известно, что на функцию белка или гена, являющихся представителями какой-либо генной сети, существенное влияние могут оказывать другие участники генной сети посредством прямых взаимодействий, а также путем передачи сигнала по цепочкам вершин сети, имеющих между собой соответствующие связи. В настоящей работе с помощью инструмента Pathway Wizard системы ANDSystem были реконструированы все возможные цепочки длиной 3 вершины, описывающие опосредованные связи между белками ОКМС и белками/генами нейронального апоптоза. Pathway Wizard позволяет проводить поиск цепочек в глобальной ассоциативной сети ANDSystem по заданному шаблону. Для поиска регуляторных цепочек, связывающих ОКМС и нейрональный апоптоз, шаблон задавался следующим образом: 1) первым участником цепочки может быть только белок или ген ОКМС; 2) вторым участником может быть любой белок или ген; 3) третьим участником может быть только белок или ген, вовлеченный в нейрональный апоптоз; 4) тип связей между первым и вторым, а также между вторым и третьим участниками цепочки может быть любой из числа: регуляция активности, регуляция деградации, регуля-

ция экспрессии, белок-белок взаимодействия и регуляция транспорта; 5) направление связи должно быть от первого участника ко второму, а также от второго к третьему (для связей белок-белок взаимодействий направление не рассматривалось). Всего получено более 5000 цепочек, связывающих, согласно заданным требованиям, 79 генов/белков, вовлеченных в ОКМС, и 188 генов/белков, участвующих в нейрональном апоптозе. Число генов/белков – посредников таких взаимодействий – составило 2064.

В найденных цепочках взаимодействий наибольшее число связей (9457) было представлено белок-белок взаимодействиями. Отметим, что число связей типа «activity upregulation» превышало число связей «activity downregulation» и составило 244 и 213 соответственно. Число связей «degradation downregulation» примерно в два раза превышало число связей «degradation upregulation» – 101 и 45 соответственно. Число связей типа «expression upregulation» превышало число связей типа «expression downregulation» и составило 144 и 106 связей соответственно. Этот результат говорит в пользу того, что механический стресс может оказывать активирующее влияние на функцию генов апоптоза даже путем не прямых межмолекулярных взаимодействий через ген/белок-посредник. По данным аннотации системой DAVID, среди 2064 генов/белков-посредников были выявлены 1939 генов/белков, которые экспрессируются в тканях мозга. Эти гены/белки вовлечены в 277 сверхпредставленных Gene Ontology биологических процессов, в том числе во врожденный иммунный ответ (GO:0045087), трансдукцию сигнала, опосредованную малой ГТФазой (GO:0007264), регуляцию процесса апоптоза (GO:0042981) и др.

Следующим шагом нашего анализа было исследование возможного влияния генов, ассоциированных с ПОУГ, на взаимодействия между ОКМС и нейрональным апоптозом. Мы предполагали, что если среди генов, обеспечивающих такое взаимодействие, согласно реконструкции ассоциативных сетей, будут статистически значимо перепредставлены гены, ассоциированные с ПОУГ, то нарушение регуляции нейронального апоптоза со стороны ОКМС, вызванное нарушениями функции или экспрессии этих генов, может быть одним из важных молекулярных механизмов патогенеза ПОУГ.

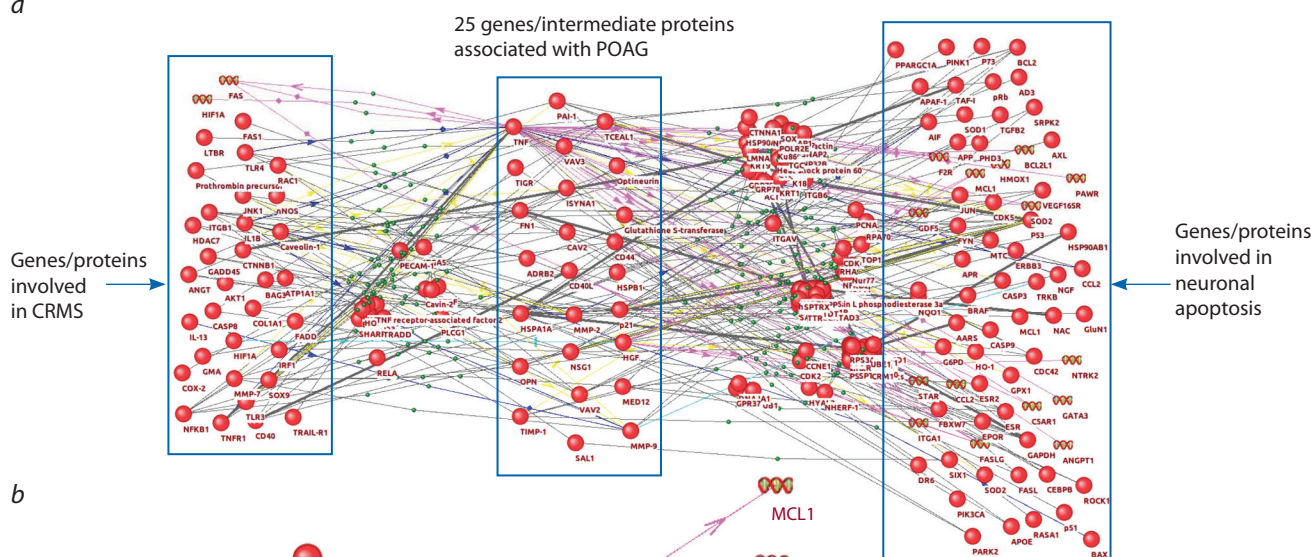
Всего нами выявлено 96 генов, вовлеченных в патогенез ПОУГ, в том числе 68 генов, ассоциированных с ПОУГ согласно данным ANDSystem, а также 37 генов, информация об ассоциации которых с ПОУГ была извлечена из баз данных. Среди них для 63 генов имелась информация о полиморфизмах, а для 33 генов указывалось изменение их экспрессии при ПОУГ. Выяснилось, что среди 80 генов, вовлеченных в ОКМС, были представлены 3 гена (*IL1B*, *TLR4* и *CAVI*), ассоциированных с ПОУГ. Несмотря на то что это число относительно небольшое, гены, ассоциированные с ПОУГ, согласно гипергеометрическому распределению были статистически значимо перепредставлены (p -value < 0,008). Известно, например, что при ПОУГ, вызванной мутациями в гене *MYOC*, наблюдается значительное повышение экспрессии гена *IL1B* (Itakura et al., 2015). Также показано, что механический стресс сопровождается повышенной экспрессией гена *IL1B* (Glossop, Cartmell, 2009). Для гена *TLR4* показано, что

его экспрессия активируется при механическом стрессе (Gawri et al., 2014). Интересно отметить, что генотип (A/A) полиморфизма rs4986790 в этом гене может иметь протективный эффект против ПОУГ (Abu-Amero et al., 2016), а ряд других полиморфизмов (rs10759930, rs1927914, rs1927911, rs2149356), возможно, имеет значение для патогенеза ПОУГ (Takano et al., 2012). В работе (Sun et al., 2003) сообщается о транслокации кавеолина-1 (продукта гена *CAVI*) на клеточной поверхности при ответе на механический стресс, а также о повышении концентрации кавеолина-1 внутри клетки. Также для гена *CAVI* известны мутации (rs4236601 и rs3801994), ассоциированные с ПОУГ (Rong et al., 2016). По-видимому, нарушения структуры кавеолина-1, вызванные мутациями, препятствуют адекватному ответу клетки на механический стресс, что может иметь значение для патогенеза ПОУГ.

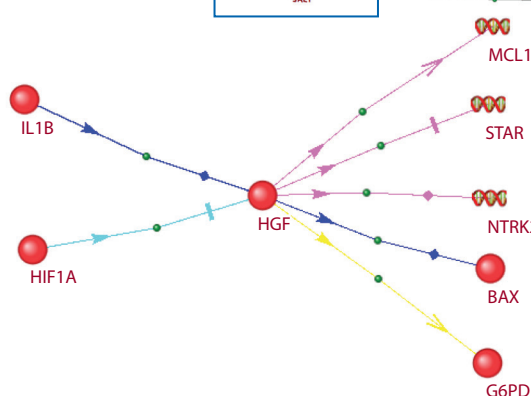
В свою очередь анализ взаимоотношения между ПОУГ и нейрональным апоптозом показал статистически значимое пересечение между списками генов, вовлеченных в нейрональный апоптоз, и генов, ассоциированных с ПОУГ. Среди 213 генов, участвующих в нейрональном апоптозе, было представлено 9 ассоциированных с ПОУГ генов, в том числе *TP53*, *APOE*, *BDNF*, *ESR2*, *SOD2*, *TGFB2*, *NRP1*, *SIX1* и *NTF4*. Данное число генов превышало ожидаемое по случайным причинам, а статистическая значимость их перепредставленности составила p -value < 10^{-7} . Для 6 из этих 9 генов (*TP53*, *APOE*, *BDNF*, *ESR2*, *SOD2*, *TGFB2*) были найдены прямые межмолекулярные взаимодействия с генами/белками, вовлеченными в ОКМС. То есть большинство ассоциированных с ПОУГ генов, участвующих в нейрональном апоптозе, оказались генами, составляющими потенциальный интерфейс для взаимодействия между ОКМС и нейрональным апоптозом. Примером такого взаимодействия является увеличение активности белка *SOD2*, участвующего в нейрональном апоптозе и ассоциированного с ПОУГ, под действием белка *IL1B*, вовлеченного в ОКМС. Известно, что у мышей с нокаутом гена *SOD2* наблюдается индукция нейронального апоптоза (Maity-Kumar et al., 2015), а повышение экспрессии гена *SOD2* приводит к снижению нейронального апоптоза (Jung et al., 2010). В работе (Borg et al., 1992) говорится об увеличении активности *SOD2* под влиянием *IL1B*. Кроме того, при механическом стрессе наблюдается значительное повышение экспрессии гена *IL1B* (Glossop, Cartmell, 2009). Взаимосвязь между *IL1B* и *SOD2* может отражать потенциальный механизм адаптации клеток сетчатки к механическому стрессу. В то же время известно, что у пациентов с ПОУГ наблюдаются мутации в гене *SOD2* (Lascaratos et al., 2012; Zhou et al., 2015). По-видимому, такие мутации могут нарушать механизмы адаптации клеток сетчатки к механическому стрессу и приводить к повышенному нейрональному апоптозу, наблюдаемому при ПОУГ.

Дальнейший анализ генов-посредников, обеспечивающих потенциальные взаимодействия между ОКМС и нейрональным апоптозом через молекулярно-генетические цепочки, включающие трех участников, тоже показал статистически значимую перепредставленность среди них генов, ассоциированных с ПОУГ. Так, оказалось, что среди 1939 генов/белков-посредников, экспрессирую-

a



b



The associative network describing putative interactions between neuronal apoptosis and CRMS through molecular genetic chains, reconstructed using Pathway Wizard of ANDSystem.

(a) The total network including 25 genes/intermediate proteins associated with POAG. (b) An example of chains passing through the hepatocyte growth factor (HGF), which acts as an intermediary associated with POAG.

щихся в тканях мозга, 25 генов/белков были ассоциированы с ПОУГ, что превышало ожидаемое по случайным причинам количество (p -value $< 10^{-5}$). На рисунке (a) показана ассоциативная сеть, включающая все цепочки, проходящие через данные 25 генов/белков-посредников. В качестве примера отдельно взятой цепочки приведена сеть взаимодействия фактора роста гепатоцитов (HGF), выступающего в роли посредника (см. рисунок, б).

Известно, что интерлейкин-1 β (IL1B), который вовлечен в ОКМС, активирует секрецию белка HGF, ассоциированного с ПОУГ (Baroni et al., 1998; Paduch et al., 2010). Другой белок, индуцируемый гипоксией фактор-1 α (HIF1A) и также участвующий в ОКМС, способен повышать стабильность HGF (Chu et al., 2009). В свою очередь HGF оказался связанным с двумя белками и тремя генами, ассоциированными с нейронным апоптозом (см. рисунок, б). Согласно данным работ (Au et al., 2009; Lu et al., 2015), HGF активирует экспрессию генов, кодирующих рецептор 2 типа нейротрофической тирозинкиназы (NTRK2) и белок Mcl1. Экспрессия другого гена стероидогенного острого регуляторного белка (STAR), вовлеченного в нейронный апоптоз, отрицательно регулируется HGF (Taniguchi et al., 2004). HGF также задействован в регуляции активности белка G6PD (Shao et

al., 1996) и транслокации белка Bax (Nakagami et al., 2002), участвующих в нейронном апоптозе.

Согласно анализу перепредставленности Gene Ontology биологических процессов, 25 генов/белков-посредников статистически значимо были вовлечены в 12 процессов (см. таблицу).

Среди этих процессов можно выделить такие, как организация внеклеточного матрикса (GO:0030198 и GO:0022617), регуляция продукции интерлейкина-8 (GO:0032757), Eph-рецепторный сигнальный путь (GO:0048013) и регуляция апоптоза (GO:0043066 и GO:2001240). Например, в процессе отрицательной регуляции пути апоптоза, индуцируемого через рецепторы клеточной смерти, в отсутствие лиганда (GO:2001240) участвовали три гена (*TNF*, *HSPA1A*, *HSPA1B*), а в процессе Eph-рецепторного сигнального пути (GO:0048013), который играет важную роль в аксональном наведении (Egea, Klein, 2007), долговременной потенциации (Kullander, Klein, 2002) и канцерогенезе (Genander, Frisén, 2010), – четыре гена (*MMP9*, *MMP2*, *VAV2*, *VAV3*).

Заключение

Анализ ассоциативных сетей, описывающих молекулярно-генетические взаимодействия между белками и гена-

Biologic processes overrepresented in Gene Ontology for the set of 25 intermediate genes expressed in brain tissues associated with POAG

Gene Ontology: biologic process	Genes	p-value with Benjamini correction
GO:0030198 ~ extracellular matrix organization	<i>TNF, MMP9, MMP2, SERPINE1, CD44, SPP1, FN1, TIMP1</i>	10 ⁻³
GO:0030168 ~ platelet activation	<i>SERPINE1, HGF, FN1, CD40LG, TIMP1, VAV2, VAV3</i>	10 ⁻³
GO:0022617 ~ extracellular matrix disassembly	<i>MMP9, MMP2, CD44, SPP1, FN1, TIMP1</i>	10 ⁻³
GO:0032757 ~ upregulation of interleukin-8 production	<i>TNF, SERPINE1, HSPA1A, HSPA1B</i>	10 ⁻³
GO:0007596 ~ blood coagulation	<i>SERPINE1, CD44, HGF, FN1, TIMP1, VAV2, VAV3</i>	0.01
GO:0002576 ~ platelet degranulation	<i>SERPINE1, HGF, FN1, TIMP1</i>	0.02
GO:0001525 ~ angiogenesis	<i>MMP2, SERPINE1, FN1, VAV2, VAV3</i>	0.02
GO:0010628 ~ upregulation of gene expression	<i>TNF, CD44, FN1, HSPA1A, HSPA1B</i>	0.02
GO:0048013 ~ ephrin receptor signaling pathway	<i>MMP9, MMP2, VAV2, VAV3</i>	0.02
GO:0043066 ~ negative regulation of the apoptotic process	<i>MMP9, CDKN1A, CD44, FN1, CD40LG, TIMP1</i>	0.03
GO:0035987 ~ endodermal cell differentiation	<i>MMP9, MMP2, FN1</i>	0.03
GO:2001240 ~ negative regulation of the extrinsic apoptotic signaling pathway with the absence of the ligand	<i>TNF, HSPA1A, HSPA1B</i>	0.04

ми, вовлеченными в ответ клетки на механический стресс (GO:0071260) и нейрональный апоптоз (GO:0051402), показал, что эти процессы тесно связаны между собой, а их участники имеют статистически значимое пересечение с генами, ассоциированными с ПОУГ. При этом было обнаружено, что гены, ассоциированные с ПОУГ, представлены в большем количестве, чем ожидалось по случайным причинам, среди генов, участвующих в прямых и не прямых взаимодействиях ОКМС/нейрональный апоптоз. Среди генов, ассоциированных с ПОУГ, составляющих потенциальный интерфейс для прямых взаимодействий ОКМС/нейрональный апоптоз, были выявлены гены *TP53*, *APOE*, *BDNF*, *ESR2*, *SOD2*, *TGFB2*. В не прямых взаимодействиях ОКМС/нейрональный апоптоз могут участвовать 25 генов-посредников, включая *TNF*, *MMP2*, *MMP9*, *HGF*, *MYOC*, *VAV2*, *VAV3* и др.

Таким образом, можно предположить, что нарушения во взаимодействиях между ОКМС и нейрональным апоптозом, вызванные нарушением функцией генов, ассоциированных с ПОУГ, играют важную роль в патогенезе заболевания. Результаты анализа могут быть использованы для поиска новых лекарственных мишеней для терапии первичной открытоугольной глаукомы.

Acknowledgments

The filling of the ANDSystem database for associations of genes with POAG through text mining and the reconstruction of POAG associative networks were supported by the Russian Science Foundation program «Computational systems biology: analysis and modeling of network structure–functional organization and evolution», project 14-24-00123.

The reconstruction and analysis of associative networks describing the interaction between CRMS and neuronal ap-

optosis were supported by the Russian Science Foundation program «Programmed cell death induced via death receptors: identification of molecular mechanisms underlying apoptosis by molecular simulation», project 14-44-00011.

Conflict of interest

The authors declare no conflict of interest.

References

- Abu-Amero K.K., Kondkar A.A., Mousa A., Azad T.A., Sultan T., Osman E.A., Al-Obeidan S.A. Analysis of toll-like receptor rs4986790 polymorphism in Saudi patients with primary open angle glaucoma. *Ophthalmic Genetics*. 2016;1-5. DOI 10.3109/13816810.2016.1151900.
- Almasieh M., Wilson A.M., Morquette B., Vargas J.L., Di Polo A. The molecular basis of retinal ganglion cell death in glaucoma. *Progress in Retinal Eye Research*. 2012;31(2):152-181. DOI 10.1016/j.preteyeres.2011.11.002.
- Au C.W., Siu M.K., Liao X., Wong E.S., Ngan H.Y., Tam K.F., Chan D.C., Chan Q.K., Cheung A.N. Tyrosine kinase B receptor and BDNF expression in ovarian cancers – effect on cell migration, angiogenesis and clinical outcome. *Cancer Letters*. 2009;281(2):151-161. DOI 10.1016/j.canlet.2009.02.025.
- Badorff C., Ruetten H., Mueller S., Stahmer M., Gehring D., Jung F., Ihling C., Zeiher A.M., Dimmeler S. Fas receptor signaling inhibits glycogen synthase kinase 3β and induces cardiac hypertrophy following pressure overload. *J. Clin. Investigation*. 2002;109(3):373-381. DOI 10.1172/JCI13779.
- Baroni A., Perfetto B., Ruocco E., Rossano F. Lipoteichoic acid and protein-A from *Staphylococcus aureus* stimulate release of hepatocyte growth factor (HGF) by human dermal fibroblasts. *Archives Dermatol. Research*. 1998;290(4):211-214.
- Borg L.A., Cagliero E.N., Sandler S.T., Welsh N., Eizirik D.L. Interleukin-1 beta increases the activity of superoxide dismutase in rat pan-

- cretic islets. *Endocrinology*. 1992;130(5):2851-2857. DOI 10.1210/endo.130.5.1533363.
- Calandrella N., Scarsella G., Pescosolido N., Risuleo G. Degenerative and apoptotic events at retinal and optic nerve level after experimental induction of ocular hypertension. *Mol. Cell. Biochemistry*. 2007; 301(1-2):155-163. DOI 10.1007/s11010-006-9407-0.
- Chu S.H., Feng D.F., Ma Y.B., Zhu Z.A., Zhang H., Qiu J.H. Stabilization of hepatocyte growth factor mRNA by hypoxia-inducible factor 1. *Mol. Biol. Reports*. 2009;36(7):1967-1975. DOI 10.1007/s11033-008-9406-1.
- Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. In *Silico Biology*. 2012;11(3, 4):149-161. DOI 10.3233/ISB-2012-0449.
- Dorn G.W., Kirshenbaum L.A. Cardiac reanimation: targeting cardiomyocyte death by BNIP3 and NIX/BNIP3L. *Oncogene*. 2008;27: 158-167. DOI 10.1038/onc.2009.53.
- Egea J., Klein R. Bidirectional Eph–ephrin signaling during axon guidance. *Trends in Cell Biology*. 2007;17(5):230-238. DOI 10.1016/j.tcb.2007.03.004.
- Elmore S. Apoptosis: a review of programmed cell death. *Toxicologic Pathology*. 2007;35(4):495-516. DOI 10.1080/01926230701320337.
- Fu-lan W.E., Jie G.E., Chun-ling W.A., Hui W.A., Ben-jun Z.H., Fan Z.H. Expression of HIF-1 α and VEGF in human dental pulp cells under mechanical stretch. *Shanghai J. Stomatology*. 2012;21(5).
- Gawri R., Rosenzweig D.H., Krock E., Ouellet J.A., Stone L.S., Quinn T.M., Haglund L. High mechanical strain of primary intervertebral disc cells promotes secretion of inflammatory factors associated with disc degeneration and pain. *Arthritis Research Therapy*. 2014;16(1):1. DOI 10.1186/ar4449.
- Genander M., Frisén J., Ephrins and Eph receptors in stem cells and cancer. *Current Opinion Cell Biology*. 2010;22(5):611-616. DOI 10.1016/j.ccb.2010.08.005.
- Glossop J.R., Cartmell S.H. Effect of fluid flow-induced shear stress on human mesenchymal stem cells: differential gene expression of IL1B and MAP3K8 in MAPK signaling. *Gene Expression Patterns*. 2009;9(5):381-388. DOI 10.1016/j.gep.2009.01.001.
- Itakura T., Peters D.M., Fini M.E. Glaucomatous MYOC mutations activate the IL-1/NF- κ B inflammatory stress response and the glaucoma marker SELE in trabecular meshwork cells. *Mol. Vision*. 2015; 21:1071.
- Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Systems Biology*. 2015;9(2):1. DOI 10.1186/1752-0509-9-S2-S2.
- Jindal V. Glaucoma: an extension of various chronic neurodegenerative disorders. *Mol. Neurobiology*. 2013;48(1):186-189. DOI 10.1007/s12035-013-8416-8.
- Jung J.E., Kim G.S., Chen H., Maier C.M., Narasimhan P., Song Y.S., Niizuma K., Katsu M., Okami N., Yoshioka H., Sakata H. Reperfusion and neurovascular dysfunction in stroke: from basic mechanisms to potential strategies for neuroprotection. *Mol. Neurobiology*. 2010;41(2-3):172-179. DOI 10.1007/s12035-010-8102-z.
- Kerr J.F.R., Wyllie A.H., Currie A.R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer*. 1972;26(4):239.
- Klein A., Maldonado C., Vargas L.M., Gonzalez M., Robledo F., de Arce K.P., Muñoz F.J., Hetz C., Alvarez A.R., Zanlungo S. Oxidative stress activates the c-Abl/p73 proapoptotic pathway in Niemann-Pick type C neurons. *Neurobiol. Disease*. 2011;41(1):209-18. DOI 10.1016/j.nbd.2010.09.008.
- Kullander K., Klein R. Mechanisms and functions of Eph and ephrin signalling. *Nature Rev. Mol. Cell Biology*. 2002;3(7):475-486. DOI 10.1038/nrm856.
- Lascaratos G., Garway-Heath D.F., Willoughby C.E., Chau K.Y., Schapira A.H. Mitochondrial dysfunction in glaucoma: understanding genetic influences. *Mitochondrion*. 2012;12(2):202-212. DOI 10.1016/j.mito.2011.11.004.
- Lu S., Török H.P., Gallmeier E., Kolligs F.T., Rizzani A., Arena S., Göke B., Gerbes A.L., De Toni E.N. Tivantinib (ARQ 197) affects the apoptotic and proliferative machinery downstream of c-MET: role of Mcl-1, Bcl-xl and Cyclin B1. *Oncotarget*. 2015;6(26):22167. DOI 10.18632/oncotarget.4240.
- Maity-Kumar G., Thal D.R., Baumann B., Scharffetter-Kochanek K., Wirth T. Neuronal redox imbalance results in altered energy homeostasis and early postnatal lethality. *FASEB J*. 2015;29(7):2843-2858. DOI 10.1096/fj.14-265157.
- Matsopoulos G.K., Asvestas P.A., Delibasis K.K., Mouravliansky N.A., Zeyen T.G. Detection of glaucomatous change based on vessel shape analysis. *Comput. Med. Imaging Graphics*. 2008;32(3):183-192. DOI 10.1016/j.compmedimag.2007.11.003.
- McCubrey J.A., Steelman L.S., Bertrand F.E., Davis N.M., Sokolovsky M., Abrams S.L., Montalto G., D'Assoro A.B., Libra M., Nicoletti F., Maestro R. GSK-3 as potential target for therapeutic intervention in cancer. *Oncotarget*. 2014;5(10):2881-2911. DOI 10.18632/oncotarget.2037.
- Nakagami H., Morishita R., Yamamoto K., Taniyama Y., Aoki M., Yamasaki K., Matsumoto K., Nakamura T., Kaneda Y., Ogihara T. Hepatocyte growth factor prevents endothelial cell death through inhibition of bax translocation from cytosol to mitochondrial membrane. *Diabetes*. 2002;51(8):2604-2611.
- Ozaki T., Kubo N., Nakagawa A. p73-binding partners and their functional significance. *Intern. J. Proteomics*. 2010. DOI 10.1155/2010/283863.
- Paduch R., Gil J.J., Niedziela P. Hepatocyte growth factor (HGF), heat shock proteins (HSPs) and multidrug resistance protein (MRP) expression in co-culture of colon tumor spheroids with normal cells after incubation with interleukin-1 β (IL-1 β) and/or camptothecin (CPT-11). *Indian J. Exp. Biol*. 2010;48(4):354-364.
- Park J., Park S.Y., Shin E., Lee S.H., Kim Y.S., Lee D.H., Roh G.S., Kim H.J., Kang S.S., Cho G.J., Jeong B.Y. Hypoxia inducible factor-1 α directly regulates nuclear clusterin transcription by interacting with hypoxia response elements in the clusterin promoter. *Mol. Cells*. 2014;37(2):178-186. DOI 10.14348/molcells.2014.2349.
- Quigley H.A. Neuronal death in glaucoma. *Progress Retinal Eye Res*. 1999;18(1):39-57.
- Quigley H.A. Glaucoma. *Lancet*. 2011;377(9774):1367-1377. DOI 10.1016/S0140-6736(10)61423-7.
- Rong S.S., Chen L.J., Leung C.K., Matsushita K., Jia L., Miki A., Chiang S.W., Tam P.O., Hashida N., Young A.L., Tsujikawa M. Ethnic specific association of the *CAV1/CAV2* locus with primary open-angle glaucoma. *Sci. Rep*. 2016;6:27837. DOI 10.1038/srep27837.
- Shao Q., Arakaki N., Ohnishi T., Nakamura O., Daikuhara Y. Effect of hepatocyte growth factor/scatter factor on lipogenesis in adult rat hepatocytes in primary culture. *J. Biochemistry*. 1996;119(5): 940-946.
- Slemmer J.E., Zhu C., Landshamer S., Trabold R., Grohm J., Ardeschiri A., Wagner E., Sweeney M.I., Blomgren K., Culmsee C., Weber J.T. Causal role of apoptosis-inducing factor for neuronal cell death following traumatic brain injury. *Am. J. Pathol*. 2008;173(6): 1795-1805. DOI 10.2353/ajpath.2008.080168.
- Sun R.J., Muller S., Zhuang F.Y., Stoltz J.F., Wang X. Caveolin-1 redistribution in human endothelial cells induced by laminar flow and cytokine. *Biorheology*. 2003;40(1,2,3):31-39.
- Takano Y., Shi D., Shimizu A., Funayama T., Mashima Y., Yasuda N., Fukuchi T., Abe H., Ideta H., Zheng X., Shiraishi A. Association of Toll-like receptor 4 gene polymorphisms in Japanese subjects with primary open-angle, normal-tension, and exfoliation glaucoma. *Am. J. Ophthalmology*. 2012;154(5):825-832. DOI 10.1016/j.ajo.2012.03.050.
- Taniguchi F., Harada T., Deura I., Iwabe T., Tsukihara S., Terakawa N. Hepatocyte growth factor promotes cell proliferation and inhibits progesterone secretion via PKA and MAPK pathways

- in a human granulosa cell line. *Mol. Reproduction Development*. 2004;68(3):335-344. DOI 10.1002/mrd.20076.
- Uchida K., Nakajima H., Takamura T., Furukawa S., Kobayashi S., Yayama T., Baba H. Gene expression profiles of neurotrophic factors in rat cultured spinal cord cells under cyclic tensile stress. *Spine*. 2008;33(24):2596-2604. DOI 10.1097/BRS.0b013e31818917af.
- Wang F., Xiong L., Huang X., Zhao T., Wu L.Y., Liu Z.H., Ding X., Liu S., Wu Y., Zhao Y., Wu K. miR-210 suppresses BNIP3 to protect against the apoptosis of neural progenitor cells. *Stem Cell Res*. 2013;11(1):657-667. DOI 10.1016/j.scr.2013.04.005.
- Zhao Y., Zhu L., Yu S., Zhu J., Wang C. CaMKII inhibition promotes neuronal apoptosis by transcriptionally upregulating Bim expression. *NeuroReport*. 2016;27(14):1018-1023. DOI 10.1097/WNR.0000000000000648.
- Zhou Y., Millward-Sadler S.J., Lin H., Robinson H., Goldring M., Salter D.M., Nuki G. Evidence for JNK-dependent up-regulation of proteoglycan synthesis and for activation of JNK1 following cyclical mechanical stimulation in a human chondrocyte culture model. *Osteoarthritis Cartilage*. 2007;15(8):884-893. DOI 10.1016/j.joca.2007.02.001.
- Zhou Y., Shuai P., Li X., Liu X., Wang J., Yang Y., Hao F., Lin H., Zhang D., Gong B. Association of *SOD2* polymorphisms with primary open angle glaucoma in a Chinese population. *Ophthalmic Genetics*. 2015;36(1):43-49. DOI 10.3109/13816810.2014.985844.

Математическая модель циркадного осциллятора млекопитающих: взаимодействие с системой NAD⁺/SIRT1 и возрастные изменения экспрессии генов циркадного осциллятора

Н.А. Подкольный^{1, 2}✉, Н.Н. Твердохлеб^{1, 3}, О.А. Подкольная¹

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное бюджетное учреждение науки Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия

³ Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Исследования последнего десятилетия позволяют по-новому посмотреть на возможную связь между процессами старения и циркадным ритмом. Одно из перспективных направлений в этой области появилось в связи с новыми данными относительно участия NAD⁺-зависимой деацетилазы гистонов SIRT1 в интеграции путей регуляции циркадного ритма и метаболизма и новой функции NAD⁺ как «метаболического осциллятора». Представлена модифицированная и расширенная нами наиболее детальная модель циркадного осциллятора, разработанная в 2012 г. J.K. Kim и D.B. Forger. В нее включена дополнительная обратная связь осциллятора с участием генов/белков NAMPT, SIRT1, а также NAM, NAD⁺. Участие транскрипционного фактора CLOCK/BMAL1 в регуляции транскрипции гена *NAMPT* определяет соответствующую ритмичность экспрессии мРНК и белка NAMPT. Поскольку фермент, продукт этого гена, является ключевым в пути биосинтеза и рециклирования NAD⁺, циркадный ритм характерен и для колебания уровня этого кофермента и активности NAD⁺-зависимой деацетилазы гистонов SIRT1. Деацетилирование этим ферментом компонент циркадного осциллятора замыкает обратную связь, опосредуемую этим путем. В частности, среди эффектов SIRT1 в циркадном осцилляторе можно выделить усиление деградации белка *Per2*, усиление транскрипции гена *Bmal1*, деацетилирование хроматина регуляторных областей генов циркадного осциллятора в области E-боксов с последующим подавлением транскрипции. Все эти процессы представлены в предложенной нами расширенной модели циркадного осциллятора. На основе экспериментальных данных об изменениях активности SIRT1 и уровня NAD⁺ с возрастом предпринята попытка исследовать влияние этих возрастных изменений на функционирование циркадного осциллятора. Данные моделирования свидетельствуют о снижении уровня экспрессии ряда генов циркадного осциллятора, в частности *Bmal1* и *Per2*, в старших возрастных группах. Отмечается также увеличение периода циркадных осцилляций. Полученные результаты свидетельствуют, что снижение активности SIRT1, связанное с возрастным нарушением метаболизма NAD⁺, может быть одной из причин нарушений функционирования циркадного осциллятора в супрахизматических ядрах. Такие нарушения могут повлечь за собой и нарушения циркадных ритмов организма в целом.

Ключевые слова: циркадный ритм; математическое моделирование; SIRT1; NAD⁺; старение.

Computational model for mammalian circadian oscillator: interacting with NAD⁺/SIRT1 pathway and age-related changes in gene expression of circadian oscillator

N.L. Podkolodnyy^{1, 2}✉, N.N. Tverdokhlebova^{1, 3}, O.A. Podkolodnaya¹

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk

³ Novosibirsk State University, Novosibirsk, Russia

Studies of the last decade reveal a new sight on the possible link between aging processes and circadian rhythm. New data on the role of the NAD⁺-dependent histone deacetylase SIRT1 in the integration of regulation pathways for circadian rhythms and metabolism as well as data on a new function of the NAD⁺ as the "metabolic oscillator" open a promising direction in this area. In the paper we suggested a modification and extension of the most detailed model for the circadian oscillator developed by Kim and Forger (2012). We included the additional feedback of the oscillator which concerns genes/proteins NAMPT, SIRT1, and also NAM, NAD⁺. The regulation of transcription for gene *NAMPT* by transcription factor CLOCK/BMAL1 determine the appropriate rhythm of mRNA and protein NAMPT expression. Since an enzyme product of this gene is a key in the pathway of biosynthesis and recycling of NAD⁺, therefore the circadian rhythm is also characteristic for the fluctuations in the level of this coenzyme and in the activity of NAD⁺-dependent histone deacetylase SIRT1. The deacetylation of circadian oscillator components by this enzyme closes the feedback mediated through this pathway. In particular, the effects of SIRT1 in circadian oscillator are the gain of degradation of protein *Per2*, increasing of the gene *Bmal1* transcription, deacetylation of chromatin in regulatory regions of circadian oscillator genes in the E-boxes area with subsequent suppression of transcription. We took into account all of these processes in our extended model of the circadian oscillator. Based on the experimental

data on the aging changes in the activity of SIRT1 and the level of NAD⁺, we attempted to study the effect of these age-related changes on the functioning of the circadian oscillator. Simulation data showed a decrease in expression level of several genes of the circadian oscillator, in particular, *Bmal1* and *Per2*, in the older age groups. In addition, our extended model predicted an increase in the period of oscillations. The results indicate that decrease in SIRT1 activity deal with age-related NAD⁺ metabolic disorder may be one of the reasons for the circadian oscillator dysfunctions in the suprachiasmatic nuclei. Such disorders may result in a breaking of the circadian rhythms in the body as a whole.

Key words: circadian oscillator; mathematical modeling; SIRT1; NAD⁺; aging.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Подкольный Н.Л., Твердохлеб Н.Н., Подкольная О.А. Математическая модель циркадного осциллятора млекопитающих: взаимодействие с системой NAD⁺/SIRT1 и возрастные изменения экспрессии генов циркадного осциллятора. Вавиловский журнал генетики и селекции. 2016;20(6):848-856. DOI 10.18699/VJ16.201

HOW TO CITE THIS ARTICLE:

Podkolodnyy N.L., Tverdokhleby N.N., Podkolodnaya O.A. Computational model for mammalian circadian oscillator: interacting with NAD⁺/SIRT1 pathway and age-related changes in gene expression of circadian oscillator. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):848-856. DOI 10.18699/VJ16.201

Исследования последнего десятилетия позволяют по-новому посмотреть на возможную связь между процессами старения и циркадным ритмом. Одно из перспективных направлений в этой области появилось в связи с новыми данными относительно участия NAD⁺-зависимой деацетилазы гистонов SIRT1 в интеграции путей регуляции циркадного ритма и метаболизма и новой функции NAD⁺ как «метаболического осциллятора» (Imai, 2010). Практически все биологические процессы в организме модулируются внутренними циркадными часами, которые оптимизируют физиологические и поведенческие характеристики в соответствии с изменяющимися требованиями внешней среды (Reppert, Weaver, 2002; Albrecht, Eichele, 2003; Ko, Takahashi, 2006).

Главный водитель циркадного ритма локализован в супрахиазматических ядрах (СХЯ) гипоталамуса. Он синхронизирует циркадные часы периферических тканей и органов. Центральные и периферические клеточные циркадные осцилляторы имеют схожую молекулярно-генетическую структуру. В основе ее лежит группа высококонсервативных генов (*Clock*, *Bmal1*, *Per1/2/3*, *Cry1/2*, *Rev-erba*, *RORα*), которая образует ядро циркадного осциллятора (ЦО) (Shearman et al., 2000). Часовой механизм функционирует за счет транскрипционно-трансляционных обратных связей между генами/белками ядра ЦО, формирующих осцилляции с периодом близким к 24 часам (Reppert, Weaver, 2002; Albrecht, Eichele, 2003; Ko, Takahashi, 2006). В основе ЦО лежат две базовые отрицательные обратные связи. Во-первых, гетеродимерный транскрипционный фактор CLOCK/BMAL1, связываясь с регуляторными районами генов *Per1/2/3* и *Cry1/2*, активирует их транскрипцию. Белки PER и CRY при достижении критической концентрации образуют гетеродимер и, связываясь с транскрипционным фактором CLOCK/BMAL1, подавляют его активность, а значит, и транскрипцию собственных генов. Вторая обратная связь направлена на регуляцию транскрипции гена *Bmal1* и вовлекает

транскрипционные факторы REV-ERBα and RORα. Ген *Rev-erba*, так же как и гены *Per* и *Cry*, активируется транскрипционным фактором CLOCK/BMAL1. Продукт этого гена – транскрипционный фактор REV-ERBα – подавляет транскрипцию гена *Bmal1*, функционально конкурируя с активатором транскрипции RORα за RRE элемент в промоторе гена *Bmal1*. Это влечет за собой снижение уровня белка BMAL1, а значит, и транскрипционного фактора CLOCK/BMAL1. Таким образом, транскрипционный фактор REV-ERBα, подавляя транскрипцию гена *Bmal1*, подавляет транскрипцию и собственного гена.

Ядро ЦО, и в первую очередь транскрипционный фактор CLOCK/BMAL1, контролирует множество генов вне осциллятора, обеспечивая регуляцию широкого спектра процессов и функций организма в циркадной манере. Некоторые из этих генов образуют дополнительные обратные связи с генами ЦО. Такие связи обеспечивают устойчивость функционирования осциллятора и его связь с другими молекулярно-генетическими системами и путями организма.

Одна из дополнительных обратных связей формируется с участием NAD⁺-зависимой деацетилазы гистонов SIRT1 (Asher et al., 2008; Nakahata et al., 2009; Chang, Guarente, 2013; Wang et al., 2016). Ген *Sirt1* экспрессируется повсеместно. Белок SIRT1 локализуется преимущественно в ядре, хотя субклеточная его локализация может изменяться во время развития организма и в ответ на определенные физиологические и патологические стимулы (Tanno et al., 2007). Субстратами SIRT1 являются разнообразные структурные, каталитически активные и регуляторные белки, в том числе транскрипционные факторы семейства FOXO, LXR, HIF-1α, NF-κβ, P53, BMAL1, корегуляторы транскрипции p300, PPARγ коактиватор-1α (PGC-1α), Per2 и др. (Vaziri et al., 2001; Yeung et al., 2004; Bouras et al., 2005; Nemoto et al., 2005; Asher et al., 2008; Nakahata et al., 2009; Lim et al., 2010; Defour et al., 2012; Martínez-Redondo, Vaquero, 2013; Kim et al., 2015). Деацетилирование соот-

ветствующих белков влияет на их активность, стабильность и другие свойства (Haigis, Sinclair, 2010). Таким образом, SIRT1 оказывается вовлеченным в регуляцию широкого спектра процессов в клетке и организме. Его эффекты могут быть связаны с выживанием и дифференцировкой клеток, развитием организма и продолжительностью жизни, метаболизмом, воспалением, развитием нейродегенеративных заболеваний и рака (Cho et al., 2015; Masri, 2015). В целом подчеркивается, что активность SIRT1 критична для поддержания здоровья (Chang, Guarente, 2013; Imai, Guarente, 2014). Мишенями SIRT1 в ядре ЦО могут быть ген и белок BMAL1, белок PER2, а также хроматин в области сайтов связывания CLOCK/BMAL1 (Asher et al., 2008; Nakahata et al., 2009; Chang, Guarente, 2013; Wang et al., 2016). Ферментативная активность SIRT1 зависит от уровня NAD⁺, который в данном случае выступает в роли косубстрата для этой деацетилазы (Revollo et al., 2004; Yoshino et al., 2011; Stein, Imai, 2014). Лимитирующим звеном биосинтеза и рециклирования NAD⁺ является фермент NAMPT. Экспрессия гена *Nampt* регулируется транскрипционным фактором CLOCK/BMAL1 и проявляет циркадные осцилляции на уровне мРНК и белка (Revollo et al., 2004; Nakahata et al., 2009; Ramsey et al., 2009; Yoshino et al., 2011; Stein, Imai, 2014; Nikiforov et al., 2015). В соответствии с этим уровень NAD⁺, а значит и активность SIRT1, также изменяется с 24-часовым периодом.

Параметры циркадного ритма и экспрессия ряда генов циркадных часов как в СХЯ, так и в периферических тканях претерпевают существенные изменения с возрастом (Kolker et al., 2003; Wyse, Coogan, 2010; Chang, Guarente, 2013; Bonaconsa et al., 2014; Baburski et al., 2016; Banks et al., 2016). Известно также, что с возрастом снижается ферментативная активность SIRT1 (Herranz et al., 2010; Masri, 2015; Poulouse, Raju, 2015).

Цель нашего исследования – моделирование возрастных изменений функционирования циркадных часов, обусловленных влиянием системы NAMPT/NAD⁺/SIRT1, которая регулируется циркадными часами и воздействует на них, модифицируя компоненты циркадного осциллятора.

Предпосылки

Нами модифицирована и развита детальная модель ЦО, разработанная J.K. Kim и D.B. Forger (Kim, Forger, 2012). Эта модель отличается тем, что включает подробное описание компонентов молекулярно-генетического ЦО и регуляторные связи между ними, обеспечивающие функционирование осциллятора. Кроме того, она содержит все компоненты, необходимые для объединения модели ЦО и подсистемы регуляции активности SIRT1, включающей NAD⁺, NAM, ген *Nampt*, а также белки NAMPT и SIRT1. При моделировании мы использовали следующие положения, полученные из литературы.

1. Ген *Nampt* кодирует фермент NAMPT. Промоторная область гена *Nampt* содержит три высококонсервативных у млекопитающих Е-бокса (E-box), связывающих транскрипционный фактор CLOCK/BMAL1 (Nakahata et al., 2009). Данный транскрипционный фактор, являющийся главным активатором транскрипции ЦО, обеспечива-

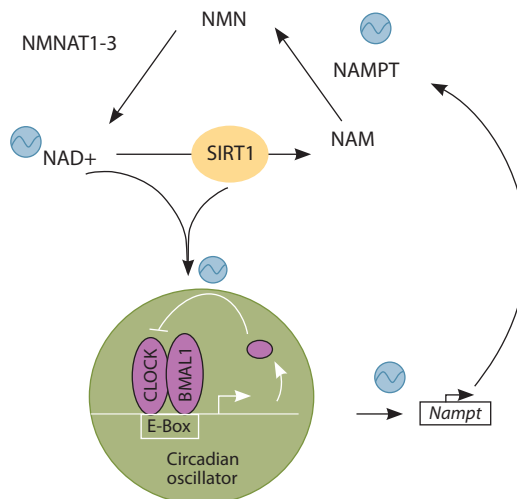


Fig. 1. Schematic representation of the circadian regulation of the NAD⁺ level and SIRT1 activity in mammals.

The *Nampt* gene is directly controlled by the circadian oscillator. Its mRNA and protein levels exhibit circadian oscillations. NAMPT is a key enzyme in NAD⁺ biosynthesis and recycling; therefore, the level of NAD⁺, and hence the SIRT1 activity vary in a rhythmic manner.

ет ритмический характер экспрессии этого гена. Как мРНК, так и белок NAMPT демонстрируют циркадные осцилляции (Nakahata et al., 2009; Ramsey et al., 2009).

2. Фермент NAMPT служит лимитирующим звеном рециклирования и биосинтеза NAD⁺ в клетке. Циклические изменения уровня NAD⁺ в клетке являются отражением циркадного характера экспрессии гена и белка NAMPT (Revollo et al., 2004; Nakahata et al., 2009; Ramsey et al., 2009; Nikiforov et al., 2015).
3. NAD⁺ является косубстратом деацетилазы гистонов SIRT1. Активность SIRT1 определяется уровнем NAD⁺, и поскольку уровень NAD⁺ меняется в циркадной манере, такой же характер носит и изменение активности SIRT1 (Revollo et al., 2004; Nakahata et al., 2009; Nikiforov et al., 2015) (рис. 1). Продукт реакции деацетилирования белков с участием SIRT1 и NAD⁺, никотинамид (NAM) поступает в систему рециклирования и вновь превращается в NAD⁺ с участием фермента NAMPT. NAM подавляет активность SIRT1. Тем самым SIRT1 контролирует колебательный синтез собственного косубстрата.
4. Деацетилирование PER2 снижает его стабильность (Asher et al., 2008) (рис. 2). Гетеродимер CLOCK/BMAL1 активирует транскрипцию генов *Per*, *Cry*. Как только уровень белков PER и CRY достигает до некоего критического значения, они формируют комплекс с транскрипционным фактором CLOCK/BMAL1 и репрессируют его активность, тем самым подавляя транскрипцию собственных генов. SIRT1 деацетилирует PER2 в комплексе CLOCK/BMAL1:PER2/CRY. Деацетилированный PER2 подвергается деградации (Asher et al., 2008). Это снижает репрессирующее действие комплекса PER/CRY на CLOCK/BMAL1.
5. SIRT1 деацетилирует коактиватор транскрипции PGC1 α , тем самым повышая его активность (Nemoto

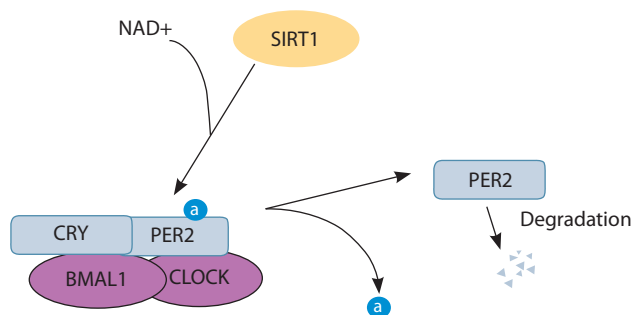


Fig. 2. Schematic representation of the effect of SIRT1 on PER2 deacetylation and degradation.

The PER2/CRY heterodimer (gray bars) binds to CLOCK/BMAL1 transcription factor (open ovals), suppressing its activity. The complex contains acetylated PER2. SIRT1 deacetylates PER2, and the latter is degraded.

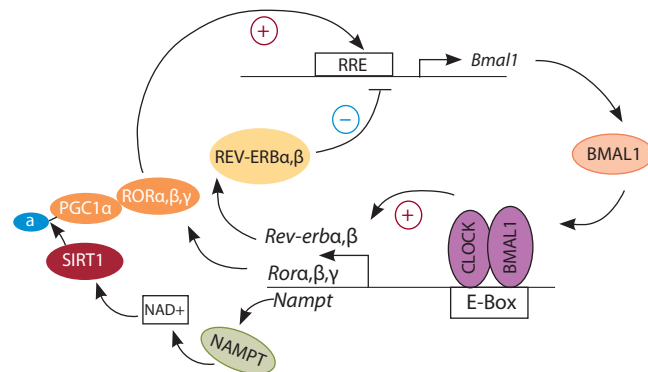


Fig. 3. Schematic representation of SIRT1-mediated *Bmal1* gene activation. Depicted is the oscillating loop that increases the expression of BMAL1 and targets of CLOCK/BMAL1 transcription factor.

et al., 2005; Rodgers et al., 2005; Gurd, 2011). PGC1 α является коактиватором транскрипционного фактора ROR α , участвующего в регуляции транскрипции гена *Bmal1* (Akashi, Takumi, 2005; Liu et al., 2007). Таким образом, SIRT1 участвует в регуляции транскрипции этого гена (Chang, Guarente, 2013) (рис. 3).

6. Деацетилаза SIRT1 привлекается комплексом CLOCK/BMAL1 к E-боксам регуляторных областей циркадных генов и модулирует ацетилирование гистонов, тем самым способствуя ритмическому ремоделированию хроматина и, соответственно, ритмической регуляции транскрипции генов (Nakahata et al., 2008).
7. Уровень NAD⁺ снижается с возрастом (Revollo et al., 2004; Ramsey et al., 2009; Yoshino et al., 2011; Stein, Imai, 2014; Wang et al., 2016).
8. Активность SIRT1 снижается с возрастом (Revollo et al., 2004; Yoshino et al., 2011; Chang, Guarente, 2013; Stein, Imai, 2014; Wang et al., 2016).

Методы

Модификация и расширение математической модели циркадного осциллятора SCN мыши, созданной J.K. Kim и D.B. Forger (Kim, Forger, 2012), выполнена путем добавления в нее следующих подсистем.

1. Путь усиления деградации PER2 после его деацетилирования ферментом SIRT1. Поскольку для подавления активности транскрипционного фактора CLOCK/BMAL1 критичен уровень белка PER2, это действие SIRT1 может быть весьма эффективным регулятором циркадных часов (см. рис. 2). Активность фермента SIRT1 определяется концентрацией NAD⁺. Поэтому данный путь представлен в модели в виде линейной зависимости скорости деградации PER2 в ядре от концентрации NAD⁺.
2. Путь влияния SIRT1 на транскрипцию гена *Bmal1* через активацию корегулятора транскрипции PGC1 α посредством его деацетилирования. Мишенью PGC1 α в системе ЦО, согласно гипотезе (Chang, Guarente, 2013), является транскрипционный фактор ROR α , активирующий транскрипцию гена *Bmal1* (см. рис. 3). В нашей модели влияние активности SIRT1 на транскрипцию

гена *Bmal1* описано линейной зависимостью скорости трансляции гена *Bmal1* от концентрации NAD⁺ с базовым уровнем транскрипции, который соответствует ситуации нокаута гена (мутации *Sirt1*—/—).

3. Путь влияния SIRT1 на скорость освобождения комплекса CLOCK/BMAL1 от E-боксов. В данном случае моделируется эффект SIRT1, проявляющийся в деацетилировании гистонов в области E-боксов, связывающих CLOCK/BMAL1. Для описания этого механизма в модель введена явная зависимость от уровня NAD⁺ нормализованной скорости освобождения комплекса CLOCK/BMAL1 от генов E-бокс *Per1/2*, *Cry1*, гена *Cry2*, гена *Rev-erba* (константы unbin, unbinc, unbir соответственно (Kim, Forger, 2012)).
4. Подсистема, включающая NAMPT/NAD⁺/SIRT1 путь. Этот путь представляет обратную связь, которая через регуляцию транскрипции гена *Nampt* транскрипционным фактором CLOCK/BMAL1 приводит к ритмическому накоплению NAD⁺ — косубстрата деацетилазы SIRT1 (см. рис. 1). Механизм регуляции транскрипции гена *Nampt* определяется наличием трех E-боксов в промоторе этого гена и описан в нашей модели аналогично регуляции генов *Per1*, *Per2* и *Cry1* в модели (Kim, Forger, 2012).

Таким образом, мы включили в модель пути, обеспечивающие циркадную регуляцию активности деацетилазы SIRT1, и пути, обеспечивающие эффекты этого фермента в системе ЦО.

Оценка значений вновь добавляемых параметров выполнялась в два этапа. На первом этапе мы нашли значения параметров, которые обеспечивают соответствие динамики концентраций мРНК основных генов ЦО в ядре клетки экспериментальным данным, используемым при идентификации модели (Kim, Forger, 2012), а также уровни и фазы осцилляций мРНК и белка NAMPT, NAD⁺ и NAM, описанных в базе данных CircaDB (<http://circadb.hogeschlab.org/query>), а также в работах (Nakahata et al., 2009; Ramsey et al., 2009).

На рис. 4 представлены кривые относительного уровня мРНК генов *Per2* и *Bmal1*, рассчитанные с помощью нашей расширенной модели ЦО, и значения уровней

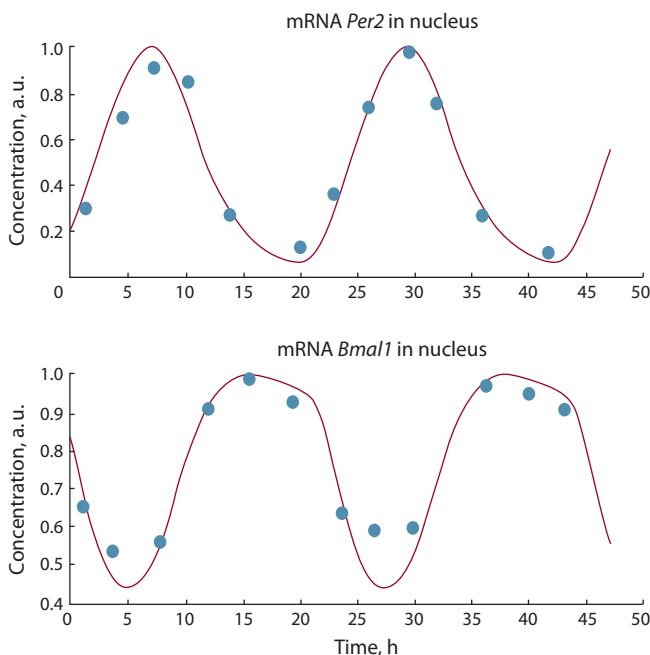


Fig. 4. Comparison of the oscillatory behavior simulated by the extended model for the mRNA *Per2* and *Bmal1* genes with the experimental data. Dots on the curve indicate the experimental data used in Kim & Forger's (2012) model.

экспрессии, полученные в экспериментальных исследованиях (Kim, Forger, 2012). Как видно из рис. 4, изменения, внесенные нами в модель, не нарушили ее адекватности, и кривые осцилляций основных компонент модели соответствуют экспериментальным данным.

На следующем этапе была выполнена подгонка значимых параметров модели с использованием экспериментальных данных, полученных в синхронизированной культуре нативных нейрональных клеток мыши N2a и при деpleции SIRT1 (Chang, Guarente, 2013), а также экспериментальных данных по изменению уровня NAD⁺ у мышей с возрастом (Wang et al., 2016).

Моделирование выполнялось в среде Matlab с использованием решателя ОДУ – ode15s.

Результаты и обсуждение

Математическая модель ЦО млекопитающих, представленная в работе (Kim, Forger, 2012), была модифицирована и расширена нами путем добавления в нее следующих подсистем:

- Подсистема, включающая NAMPT/NAD⁺/SIRT1 путь (см. рис. 1).
- Путь усиления деградации PER2 после его деацетилирования ферментом SIRT1 (см. рис. 2).
- Путь влияния SIRT1 на транскрипцию гена *Bmal1* (см. рис. 3).
- Путь влияния SIRT1 на ингибирование через деацетилирование гистонов функции CLOCK/BMAL1, связанного с E-боксом.

Таким образом, в модель были включены пути, обеспечивающие циркадную регуляцию активности деацетилазы SIRT1, и пути, обеспечивающие эффекты этого фермента в системе ЦО. Расширенная модель представлена в виде си-

стемы 186 обыкновенных дифференциальных уравнений (186 переменных и 81 параметр). Исходная модель (Kim, Forger, 2012) включала 180 переменных и 70 параметров.

Динамика изменения основных компонент расширенной модели ЦО соответствует экспериментальным данным, которые были использованы в исходной модели (Kim, Forger, 2012) (см. рис. 4). Моделирование поведения модели при минимизации активности SIRT1 показало, что экспрессия генов ЦО сохраняет ритмичность в этих условиях (рис. 5). Однако амплитуда колебаний и общий уровень экспрессии ряда генов, в том числе *Per2*, *Bmal1* и *Rev-erb*, снижаются, что согласуется с экспериментальными данными, полученными на синхронизированной культуре нейрональных клеток мыши N2a с деpleцией SIRT1 (Chang, Guarente, 2013). Кроме того, отмечено снижение амплитуды осцилляций уровня белка BMAL1 и комплекса CLOCK/BMAL1 – основного позитивного регулятора транскрипции ЦО.

Отметим, что в литературе приводятся противоречивые сведения о влиянии ингибирования/истощения SIRT1 на экспрессию циркадных генов. Так, в работах (Nakahata et al., 2008, 2009) показано, что экспрессия мРНК *Per2* увеличивается в этих условиях, в то время как другими авторами получен противоположный результат (Asher et al., 2008). Поскольку через NAD⁺/SIRT1 путь сигналы (как активирующие, так и ингибирующие транскрипцию генов) могут передаваться как минимум четырьмя механизмами, теоретически возможны оба эффекта. Для того чтобы учесть особенности регуляции через NAD⁺/SIRT1 путь, требуются дополнительные исследования.

Результаты, полученные с помощью нашей модели, находятся в соответствии с выводами (Asher et al., 2008). К аналогичным выводам пришли авторы работы (Luna et al., 2015) при математическом моделировании влияния скорости биосинтеза NAD⁺, а значит, и изменения активности SIRT1 на экспрессию мРНК *Per/Cry*.

Еще один результат, выявленный нами при моделировании поведения ЦО при минимизации активности *Sirt1*, – удлинение периода осцилляций примерно на 3 %. Это согласуется с экспериментальными данными (Chang, Guarente, 2013), хотя в упомянутом исследовании период оценивался на основе актограммы BSKO (brain specific SIRT1 knockout) мышей, а не на уровне молекулярно-генетического осциллятора, что может объяснять менее выраженное изменение периода циркадного ритма, чем в нашей модели. Таким образом, поведение разработанной нами модели ЦО соответствует результатам, полученным в эксперименте.

Экспериментальные данные свидетельствуют о снижении экспрессии и активности SIRT1 с возрастом (Chang, Guarente, 2013; Wang et al., 2016). Эти изменения наблюдаются как в печени, так и в супрахиазматических ядрах гипоталамуса. Представленные в литературе сведения позволяют предположить, что возрастное снижение активности SIRT1 может быть обусловлено соответствующим снижением уровня NAD⁺ (Revollo et al., 2004; Yoshino et al., 2011; Stein, Imai, 2014).

Основываясь на экспериментальных данных по возрастной динамике активности SIRT1 (Chang, Guarente, 2013; Wang et al., 2016), мы включили в модель в качестве

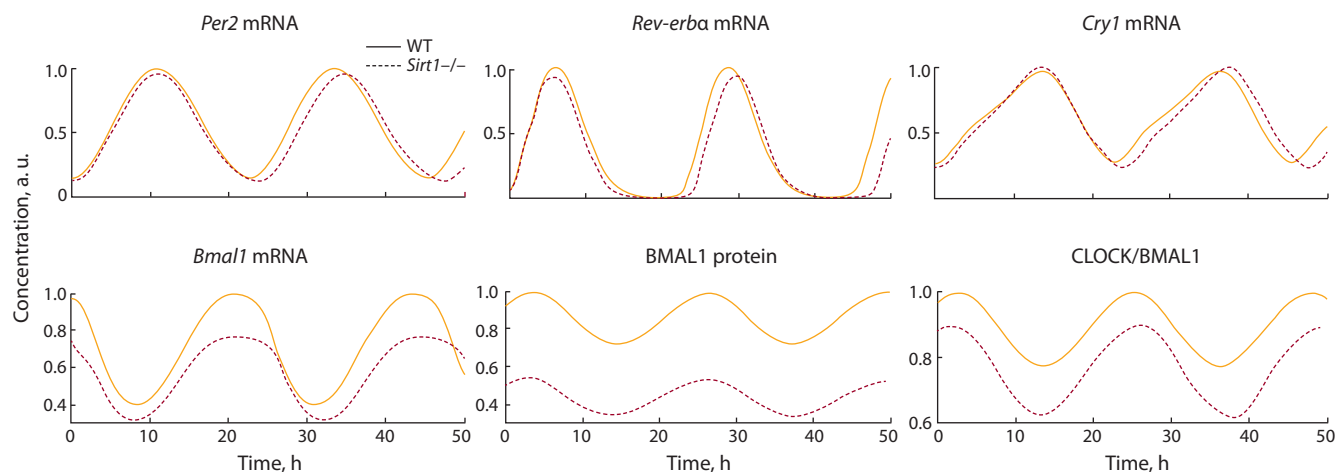


Fig. 5. The oscillatory behavior of *Per2*, *Rev-erba*, *Cry1*, and *Bmal1* mRNAs; BMAL1(B) protein; and the CLOCK/BMAL1 complex in the norm and under the conditions of SIRT1 depletion: simulation with the extended model.

параметра возраст, что позволило моделировать возрастные изменения поведения ЦО. При этом было выявлено снижение экспрессии мРНК генов *Bmal1*, *Per2*, *Rev-erba* и удлинение периода осцилляций с возрастом (рис. 6).

Полученные результаты совпадают с данными целого ряда экспериментальных работ, выполненных на различных организмах и в различных тканях. В частности, при увеличении возраста отмечено снижение экспрессии генов ЦО *Bmal1* и *Per2* в SCN мыши (Chang, Guarente, 2013). Аналогичные изменения выявлены в SCN и других областях мозга сирийского хомячка (Duncan et al., 2013). Наблюдалось снижение экспрессии генов *Bmal1* и *Clock* в SCN мыши в возрасте 16 мес. (Wyse, Coogan, 2010). У крыс в возрасте 24 мес. продемонстрировано снижение экспрессии генов *Bmal1*, *Per1*, *Per2*, *Per3* и *Rev-erba* в клетках Лейдига по сравнению с 3- и 18-месячными животными (Baburski et al., 2016). Следует отметить, что данный результат воспроизводится не во всех исследованиях. Это же касается и удлинения периода осцилляций. Определенное объяснение подобных расхождений экспериментальных данных можно найти в работе (Nakamura et al., 2015), где показано, что с увеличением возраста в клетках SCN происходит удлинение периода свободно текущего циркадного ритма и уменьшение амплитуды. Наиболее выражены эти эффекты при длительном содержании животных в условиях постоянной темноты и менее выражены или не проявляются вовсе при содержании в условиях ритмичной смены света и темноты или при коротком периоде темновой адаптации, предшествующей эксперименту. Авторы объясняют это стимулирующим действием ежедневной циклической смены свет/темнота на молекулярно-генетический ЦО клеток СХЯ, которое маскирует проявление эффектов старения в системе циркадных часов.

Таким образом, наши результаты свидетельствуют, что возрастное снижение активности SIRT1 может быть одной из причин нарушений функционирования ЦО СХЯ, которые могут повлечь за собой и нарушения циркадных ритмов организма в целом.

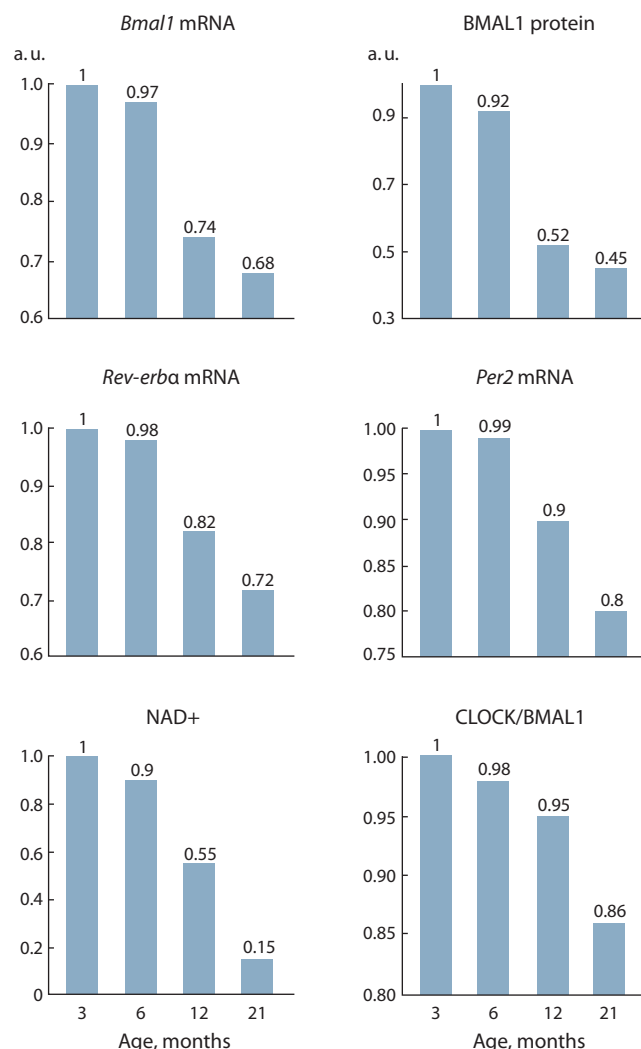


Fig. 6. Variation of the expression of the *Bmal1*, *Per2* and *Rev-erba* genes with age.

Data on the expression at the age of 3, 6, 12, and 21 months calculated with the extended model are presented

Остается открытым вопрос о причинах возрастного снижения активности SIRT1. Как показано в экспериментальных исследованиях, такие изменения, вероятно всего, являются следствием снижения уровня NAD⁺ и/или его доступности для SIRT1, которое может происходить несколькими путями. Прежде всего это возрастные изменения активности фермента NAMPT, зафиксированные экспериментально, происхождение которых пока окончательно не выяснено (Revollo et al., 2004; Yoshino et al., 2011; Stein, Imai, 2014). Вторая причина – увеличивающаяся с возрастом конкуренция за NAD⁺ между SIRT1 и другими NAD⁺ потребляющими ферментами, в частности PARP1, PARP2 и CD38/157 (Imai, Guarente, 2016). Увеличение активности этих ферментов с возрастом происходит из-за нарастания процессов, сопровождающих оксидативный стресс, которые могут проявляться уже в среднем возрасте (Braidy et al., 2011; Massudi et al., 2012). В условиях генотоксического стресса каталитическая активность PARP1 может возрастать более чем в 500 раз, при этом уровень NAD⁺ в клетке снижается до 10–20 % от исходного в течение 5–15 мин (Berger, 1985; Houtkooper et al., 2010). Утверждается, что с учетом соотношения концентрации NAD⁺ и величины констант Михаэлиса (K_m) ферментов PARP1 и CD38/157 для NAD⁺ уровень последнего не является ограничением для их активности в клетке (Houtkooper et al., 2010). Однако указанные ферменты, будучи глобальными потребителями NAD⁺, могут ограничивать его доступность для SIRT1, который характеризуется значительно более высокой K_m . Это, безусловно, может оказывать существенное влияние на активность SIRT1 при увеличении с возрастом экспрессии и активности его конкурентов в потреблении NAD⁺. В подтверждение этого положения можно привести данные об увеличении уровня NAD⁺ и активности SIRT1 у животных с нокаутом или ингибированием генов/белков PARP1 и CD38 (Aksoy et al., 2006; Barbosa et al., 2007; Houtkooper et al., 2010; Bai et al., 2011; Sahar et al., 2011). Кроме того, у CD38^{–/–} мышей отмечается также изменение параметров циркадного ритма. В частности, фиксируют укорочение его периода, а также увеличение экспрессии генов *Per2* и *Nampt*, что соответствует предположениям, которые могут быть получены из нашей модели (Sahar et al., 2011). Наблюдаются изменения некоторых характеристик циркадных часов и у мышей с нокаутом PARP1 (Asher et al., 2010). Таким образом, конкуренция за доступный NAD⁺ может быть одной из причин возрастных изменений активности SIRT1 и соответствующих изменений параметров циркадных часов.

Закключение

SIRT1 является одним из метаболических сенсоров, реагирующих на различные внешние стимулы. Его деацетилирующая активность направлена на широкий спектр субстратов, среди которых есть и структурные, и регуляторные белки, поэтому изменение активности SIRT1 влияет на самые разнообразные процессы как на клеточном, так и на системном уровне. В данной работе мы предприняли попытку исследовать влияние возрастных изменений активности SIRT1 на состояние ЦО с использованием математического моделирования. При построении модели

мы исходили из экспериментальных данных о мишенях SIRT1 в системе молекулярно-генетического ЦО, а также принимали во внимание зависимость активности SIRT1 от уровня его косубстрата NAD⁺. Наши исследования показали, что изменение активности SIRT1 действительно может оказывать влияние на параметры ЦО, изменяя экспрессию его отдельных компонент и период осцилляции. Вероятно, возрастные изменения активности SIRT1, вызванные истощением NAD⁺, способствуют разобщению метаболических процессов и суточного ритма, что само по себе может играть важную роль в процессе старения. Исходя из имеющихся экспериментальных данных, мы планируем продолжить работу над расширением модели. Прежде всего в модель будут внесены дальнейшие изменения, касающиеся эффектов SIRT1, которые он может оказывать на отдельные компоненты ЦО. В частности, это еще один потенциальный эффект SIRT1 в ЦО – деацетилирование BMAL1 в комплексе CLOCK/BMAL1 с последующим усилением его деградации.

С использованием нашей расширенной математической модели ЦО млекопитающих в дальнейшем может быть исследован целый ряд задач. Одна из них – изучение влияния возрастных изменений активности SIRT1 на фазовые характеристики осциллятора, а также на характеристики кривой фазной реакции осциллятора в старших возрастных группах. Это направление имеет большое значение, поскольку считается, что снижение способности циркадных часов захватываться внешними сигналами, такими как свет, поступление пищи, часто выявляемое в старших возрастных группах животных и людей, может вызывать рассогласование метаболических систем организма с последующим формированием патологий, развивающихся при старении. Кроме того, несомненно интересным будет дальнейшее расширение модели с включением PARP1 и CD38, конкурирующих за потребление NAD⁺ с SIRT1. Тем более что все три фермента участвуют в регуляции процессов, нарушение которых приводит к формированию патологических состояний, сопровождающих старение (диабет второго типа, рак, воспалительные и нейродегенеративные заболевания).

Acknowledgments

This work was supported by the Russian Science Foundation, project 14-24-00123.

Conflict of interest

The authors declare no conflict of interest.

References

- Akashi M., Takumi T. The orphan nuclear receptor ROR α regulates circadian transcription of the mammalian core-clock *Bmal1*. *Nat. Struct. Mol. Biol.* 2005;12:441–448.
- Aksoy P., Escande C., White T.A., Thompson M., Soares S., Benech J.C., Chini E.N. Regulation of SIRT 1 mediated NAD dependent deacetylation: a novel role for the multifunctional enzyme CD38. *Biochem. Biophys. Res. Commun.* 2006;349(1):353–359. Epub 2006 Aug 22.
- Albrecht U., Eichele G. The mammalian circadian clock. *Curr. Opinion Genetics Development.* 2003;13(3):271–277.
- Asher G., Gatfield D., Stratmann M., Reinke H., Dibner C., Kreppe F., Mostoslavsky R., Alt F.W., Schibler U. SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell.* 2008;134(2):317–328. DOI 10.1016/j.cell.2008.06.050.

- Asher G., Reinke H., Altmeyer M., Gutierrez-Arcelus M., Hottiger M.O., Schibler U. Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding. *Cell*. 2010; 142(6):943-953. DOI 10.1016/j.cell.2010.08.016. Epub 2010 Sep 9.
- Baburski A.Z., Sokanovic S.J., Bjelic M.M., Radovic S.M., Andric S.A., Kostic T.S. Circadian rhythm of the Leydig cells endocrine function is attenuated during aging. *Exp. Gerontol*. 2016;73: 5-13. DOI 10.1016/j.exger.2015.11.002. Epub 2015 Nov 4.
- Bai P., Cantó C., Oudart H., Brunyánszki A., Cen Y., Thomas C., Yamamoto H., Huber A., Kiss B., Houtkooper R.H., Schoonjans K., Schreiber V., Sauve A.A., Menissier-de Murcia J., Auwerx J. PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metabolism*. 2011;13(4):461-468. DOI 10.1016/j.cmet.2011.03.004.
- Banks G., Nolan P.M., Peirson S.N. Reciprocal interactions between circadian clocks and aging. *Mamm. Genome*. 2016;27(7-8):332-340. DOI 10.1007/s00335-016-9639-6.
- Barbosa M.T., Soares S.M., Novak C.M., Sinclair D., Levine J.A., Aksoy P., Chini E.N. The enzyme CD38 (a NAD glycohydrolase, EC 3.2.2.5) is necessary for the development of diet-induced obesity. *FASEB J*. 2007;21(13):3629-3639. Epub 2007 Jun 21.
- Berger N.A. Poly(ADP-ribose) in the cellular response to DNA damage. *Radiat Res*. 1985;101(1):4-15.
- Bonaconsa M., Malpeli G., Montaruli A., Carandente F., Grassi-Zucconi G., Bentivoglio M. Differential modulation of clock gene expression in the suprachiasmatic nucleus, liver and heart of aged mice. *Exp. Gerontol*. 2014;55:70-79. DOI 10.1016/j.exger.2014.03.011. Epub 2014 Mar 24.
- Bouras T., Fu M., Sauve A.A., Wang F., Quong A.A., Perkins N.D., Hay R.T., Gu W., Pestell R.G. SIRT1 deacetylation and repression of p300 involves lysine residues 1020/1024 within the cell cycle regulatory domain 1. *J. Biol. Chem*. 2005;280(11):10264-10276. Epub 2005 Jan 4.
- Braidy N., Guillemin G.J., Mansour H., Chan-Ling T., Poljak A., Grant R. Age related changes in NAD⁺ metabolism oxidative stress and Sirt1 activity in wistar rats. *PLoS One*. 2011;6(4):e19194. DOI 10.1371/journal.pone.0019194.
- Chang H.C., Guarente L. SIRT1 mediates central circadian control in the SCN by a mechanism that decays with aging. *Cell*. 2013; 153(7):1448-1460. DOI 10.1016/j.cell.2013.05.027.
- Cho S.H., Chen J.A., Sayed F., Ward M.E., Gao F., Nguyen T.A., Krabbe G., Sohn P.D., Lo I., Minami S., Devidze N., Zhou Y., Coppola G., Gan L. SIRT1 deficiency in microglia contributes to cognitive decline in aging and neurodegeneration via epigenetic regulation of IL-1 β . *J. Neurosci*. 2015;35(2):807-818. DOI 10.1523/JNEUROSCI.2939-14.2015.
- Defour A., Dessalle K., Castro Perez A., Poyot T., Castells J., Gallot Y.S., Durand C., Euthine V., Gu Y., Béchet D., Peinnequin A., Lefai E., Freyssen D. Sirtuin 1 regulates SREBP-1c expression in a LXR-dependent manner in skeletal muscle. *PLoS One*. 2012;7(9): e43490. DOI 10.1371/journal.pone.0043490. Epub 2012 Sep 11.
- Duncan M.J., Prochot J.R., Cook D.H., Smith J.T., Franklin K.M. Influence of aging on Bmal1 and Per2 expression in extra-SCN oscillators in hamster brain. *Brain Res*. 2013;1491:44-53. DOI 10.1016/j.brainres.2012.11.008. Epub 2012 Nov 15.
- Gurd B.J. Deacetylation of PGC-1 α by SIRT1: importance for skeletal muscle function and exercise-induced mitochondrial biogenesis. *Appl. Physiol. Nutr. Metab*. 2011;36(5):589-597. DOI 10.1139/h11-070. Epub 2011 Sep 2.
- Haigis M.C., Sinclair D.A. Mammalian sirtuins: biological insights and disease relevance. *Annual Rev. Pathol*. 2010;5:253-295.
- Herranz D., Muñoz-Martin M., Cañamero M., Mulero F., Martínez-Pastor B., Fernandez-Capetillo O., Serrano M. Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer. *Nat. Commun*. 2010;1:3. DOI 10.1038/ncomms1001.
- Houtkooper R.H., Cantó C., Wanders R.J., Auwerx J. The secret life of NAD⁺: an old metabolite controlling new metabolic signaling pathways. *Endocr. Rev*. 2010;31(2):194-223. DOI 10.1210/er.2009-0026. Epub 2009 Dec 9.
- Imai S. "Clocks" in the NAD World: NAD as a metabolic oscillator for the regulation of metabolism and aging. *Biochim. Biophys. Acta*. 2010;1804(8):1584-90. DOI 10.1016/j.bbapap.2009.10.024. Epub 2009 Nov 6.
- Imai S., Guarente L. NAD⁺ and sirtuins in aging and disease. *Trends Cell Biol*. 2014;24(8):464-471. DOI 10.1016/j.tcb.2014.04.002. Epub 2014 Apr 29.
- Imai S.I., Guarente L. It takes two to tango: NAD⁺ and sirtuins in aging/longevity control. *Aging Mechanisms Disease*. 2016;2:16017. DOI 10.1038/npjamd.2016.17; Epub 2016 Aug 18.
- Kim H.N., Han L., Iyer S., de Cabo R., Zhao H., O'Brien C.A., Manolagas S.C., Almeida M. Sirtuin1 suppresses osteoclastogenesis by deacetylating FoxOs. *Mol. Endocrinol*. 2015;29(10):1498-1509.
- Kim J.K., Forger D.B. A mechanism for robust circadian timekeeping via stoichiometric balance. *Mol. Syst. Biol*. 2012;8:630. DOI 10.1038/msb.2012.62.
- Ko C.H., Takahashi J.S. Molecular components of the mammalian circadian clock. *Hum. Mol. Genet*. 2006;15;15(Spec.2):R271-7.
- Kolker D.E., Fukuyama H., Huang D.S., Takahashi J.S., Horton T.H., Turek F.W. Aging alters circadian and light-induced expression of clock genes in golden hamsters. *J. Biol. Rhythms*. 2003;18(2):159-169.
- Lim J.H., Lee Y.M., Chun Y.S., Chen J., Kim J.E., Park J.W. Sirtuin 1 modulates cellular responses to hypoxia by deacetylating hypoxia-inducible factor 1 α . *Mol. Cell*. 2010;38(6):864-878. DOI 10.1016/j.molcel.2010.05.023.
- Liu C., Li S., Liu T., Borjigin J., Lin J.D. Transcriptional coactivator PGC-1 α integrates the mammalian clock and energy metabolism. *Nature*. 2007;447:477-481. PubMed: 17476214.
- Luna A., McFadden G.B., Aladjem M.I., Kohn K.W. Predicted role of NAD utilization in the control of circadian rhythms during DNA damage response. *PLoS Comput. Biol*. 2015;11(5):e1004144. DOI 10.1371/journal.pcbi.1004144.eCollection 2015.
- Martinez-Redondo P., Vaquero A. The diversity of histone versus non-histone sirtuin substrates. *Genes Cancer*. 2013;4(3-4):148-163. DOI 10.1177/1947601913483767.
- Masri S. Sirtuin-dependent clock control: new advances in metabolism, aging and cancer. *Curr. Opinion Clin. Nutrition Metabolic Care*. 2015;18(6):521-527. DOI 10.1097/MCO.0000000000000219.
- Massudi H., Grant R., Braidy N., Gesto J., Farnsworth B., Guillemin G.J. Age-associated changes in oxidative stress and NAD⁺ metabolism in human tissue. *PLoS One*. 2012;7(7):e42357. DOI 10.1371/journal.pone.0042357. Epub 2012 Jul 27.
- Nakahata Y., Kaluzova M., Grimaldi B., Sahar S., Hirayama J., Chen D., Guarente L.P., Sassone-Corsi P. The NAD⁺-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. *Cell*. 2008;134(2):329-340. DOI 10.1016/j.cell.2008.07.002.
- Nakahata Y., Sahar S., Astarita G., Kaluzova M., Sassone-Corsi P. Circadian control of the NAD⁺ salvage pathway by CLOCK-SIRT1. *Science*. 2009;324(5927):654-657.
- Nakamura T.J., Nakamura W., Tokuda I.T., Ishikawa T., Kudo T., Colwell C.S., Block G.D. Age-related changes in the circadian system unmasked by constant conditions. *eNeuro*. 2015;2(4). pii: ENEURO.0064-15.2015. DOI 10.1523/ENEURO.0064-15.2015.
- Nemoto S., Fergusson M.M., Finkel T. SIRT1 functionally interacts with the metabolic regulator and transcriptional coactivator PGC-1 α . *J. Biol. Chem*. 2005;280(16):16456-16460. Epub 2005 Feb 16.
- Nikiforov A., Kulikova V., Ziegler M. The human NAD metabolome: Functions, metabolism and compartmentalization. *Crit. Rev. Biochem. Mol. Biol*. 2015;50(4):284-297. DOI 10.3109/10409238.2015.1028612. Epub 2015 Apr 2.
- Poulose N., Raju R. Sirtuin regulation in aging and injury. *Biochim. Biophys. Acta*. 2015;1852(11):2442-2455. DOI 10.1016/j.bbdis.2015.08.017.
- Ramsey K.M., Yoshino J., Brace C.S., Abrassart D., Kobayashi Y., Marcheva B., Hong H.K., Chong J.L., Buhr E.D., Lee C., Takahashi J.S., Imai S., Bass J. Circadian clock feedback cycle through NAMPT-mediated NAD⁺ biosynthesis. *Science*. 2009;324(5927): 651-654. DOI 10.1126/science.1171641. Epub 2009 Mar 19.

- Reppert S.M., Weaver D.R. Coordination of circadian timing in mammals. *Nature*. 2002;29;418(6901):935-941.
- Revollo J.R., Grimm A.A., Imai S. The NAD biosynthesis pathway mediated by nicotinamide phosphoribosyltransferase regulates Sir2 activity in mammalian cells. *J. Biol. Chem.* 2004;279(49):50754-50763. Epub 2004 Sep 20.
- Rodgers J.T., Lerin C., Haas W., Gygi S.P., Spiegelman B.M., Puigserver P. Nutrient control of glucose homeostasis through a complex of PGC-1 α and SIRT1. *Nature*. 2005;434:113-118.
- Sahar S., Nin V., Barbosa M.T., Chini E.N., Sassone-Corsi P. Altered behavioral and metabolic circadian rhythms in mice with disrupted NAD⁺ oscillation. *Aging (Albany NY)*. 2011;3(8):794-802.
- Shearman L.P., Sriram S., Weaver D.R., Maywood E.S., Chaves I., Zheng B., Kume K., Lee C.C., van der Horst G.T., Hastings M.H., Reppert S.M. Interacting molecular loops in the mammalian circadian clock. *Science*. 2000;288(5468):1013-1019.
- Stein L.R., Imai S. Specific ablation of Nampt in adult neural stem cells recapitulates their functional defects during aging. *EMBO J.* 2014;33(12):1321-1340. DOI 10.1002/embj.201386917. Epub 2014 May 8.
- Tanno M., Sakamoto J., Miura T., Shimamoto K., Horio Y. Nucleocytoplasmic shuttling of the NAD⁺-dependent histone deacetylase SIRT1. *J. Biol. Chem.* 2007;282(9):6823-6832. Epub 2006 Dec 30.
- Vaziri H., Dessain S.K., Eaton Ng.E., Imai S.I., Frye R.A., Pandita T.K., Guarente L., Weinberg R.A. hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. *Cell*. 2001;107:149-159.
- Wang R.H., Zhao T., Cui K., Hu G., Chen Q., Chen W., Wang X.-W., Soto-Gutierrez A., Zhao K., Deng C.-X. Negative reciprocal regulation between Sirt1 and Per2 modulates the circadian clock and aging. *Sci. Rep.* 2016;6:28633.
- Wyse C.A., Coogan A.N. Impact of aging on diurnal expression patterns of CLOCK and BMAL1 in the mouse brain. *Brain Res.* 2010;1337:21-31. DOI 10.1016/j.brainres.2010.03.113. Epub 2010 Apr 9.
- Yeung F., Hoberg J.E., Ramsey C.S., Keller M.D., Jones D.R., Frye R.A., Mayo M.W. Modulation of NF- κ B-dependent transcription and cell survival by the SIRT1 deacetylase. *EMBO J.* 2004;23(12):2369-2380.
- Yoshino J., Mills K.F., Yoon M.J., Imai S. Nicotinamide mononucleotide, a key NAD⁺ intermediate, treats the pathophysiology of diet- and age-induced diabetes in mice. *Cell Metab.* 2011;14:528-536.



Структурное моделирование мод связывания НАД⁺ с ПАРП-1

Н.В. Иванисенко^{1, 2}✉, Д.А. Жечев¹, В.А. Иванисенко¹

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Ядерный белок поли(АДФ-рибозил) полимеразы-1 (ПАРП-1) играет важную роль в механизмах ответа клетки на повреждения ДНК. ПАРП-1 катализирует ковалентное связывание поли(АДФ-рибозил) полимеров со своей субъединицей, а также с другими акцепторными белками, используя НАД⁺ как донора АДФ-рибозы. Ингибиторы поли(АДФ-рибозил) полимераз показали высокую эффективность в улучшении радиотерапии и химиотерапии рака в клинических испытаниях. Разработка новых ингибиторов ПАРП-1 на основе производных природных соединений, таких как НАД⁺, представляет новую перспективную стратегию. Наличие структуры поли(АДФ-рибозил) полимеразы-1 человека в комплексе с НАД⁺ может являться отправной точкой для рационального дизайна класса низкомолекулярных ингибиторов ПАРП-1 на основе производных НАД⁺. Однако на сегодняшний день не опубликовано кристаллической структуры комплекса ПАРП-1 в комплексе с никотинамидадениндинуклеотидом (НАД⁺). В данной работе с использованием методов молекулярного моделирования нами проведено предсказание положений связывания НАД⁺ в донорном сайте каталитического домена ПАРП-1. С использованием структуры гомологов ПАРП-1 в комплексе с НАД⁺ предсказаны фармакофорные ограничения связывания НАД⁺ с ПАРП-1. На основе кластеризации конформаций ПАРП-1 в комплексе совместно кристаллизованными ингибиторами и на основе предсказания фармакофорных ограничений предложено несколько возможных моделей связывания НАД⁺ с ПАРП-1 в донорном сайте связывания каталитического домена. Согласно предсказанным моделям, для пирофосфатной группы НАД⁺ в комплексе с ПАРП-1 в донорном сайте связывания возможно наличие двух конформаций. Валидация предложенных моделей связывания НАД⁺ с ПАРП-1 может быть достигнута количественным анализом структура–активность для производных НАД⁺. Дополнительно предложены структуры двух производных НАД⁺, которые могут быть использованы для валидации предсказанных положений связывания НАД⁺.

Ключевые слова: ПАРП-1; НАД⁺; структурное моделирование.

Structural modeling of NAD⁺ binding modes to PARP-1

N.V. Ivanisenko^{1, 2}✉, D.A. Zhechev¹,
V.A. Ivanisenko¹

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

The nuclear protein poly (ADP-ribose) polymerase-1 (PARP-1) plays an important role in the signaling and repair of DNA. PARP-1 catalyses covalent binding of poly (ADP-ribose) polymers with itself as well as with other acceptor proteins using NAD⁺ as a donor of ADP-ribose. Inhibitors of poly (ADP-ribose) polymerase have been shown to be effective in improvement of radiation therapy and chemotherapy of cancer in clinical testing. Development of new poly (ADP-ribose) polymerase-1 inhibitors based on derivatives of natural compounds such as NAD⁺ represents a novel and promising strategy. The structure of complex of human poly (ADP-ribose) polymerase-1 with NAD⁺ can be a starting point for rational design of small molecule inhibitors based on NAD⁺ derivatives. Indeed there is no crystal structure of complex poly (ADP-ribose) polymerase-1 with nicotinamide adenine dinucleotide (NAD⁺) available yet. In this work using molecular modeling approaches we have predicted NAD⁺ binding modes with PARP-1 at the donor binding site of the catalytic domain. Using structures of PARP-1 homologs in complex with NAD⁺ we predicted pharmacophore restraints of NAD⁺ binding to PARP-1. Based on clustering of PARP-1 conformations in complex with co-crystallized inhibitors and predicted pharmacophore restraints, we proposed several possible models of NAD⁺ binding to PARP-1 at the donor binding site of the catalytic domain. According to the predicted models, two conformations of pyrophosphate group of NAD⁺ in complex with PARP-1 at the donor binding site are possible. Validation of the proposed models of NAD⁺ binding with PARP-1 can be achieved by quantitative structure-activity analysis of NAD⁺ derivatives. We designed two NAD⁺ derivatives, which can be used for validation of predicted NAD⁺ binding models.

Key words: PARP-1; NAD⁺; structural model.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Иванисенко Н.В., Жечев Д.А., Иванисенко В.А. Структурное моделирование мод связывания НАД⁺ с ПАРП-1. Вавиловский журнал генетики и селекции. 2016;20(6):857-862. DOI 10.18699/VJ16.202

HOW TO CITE THIS ARTICLE:

Ivanisenko N.V., Zhechev D.A., Ivanisenko V.A. Structural modeling of NAD⁺ binding modes to PARP-1. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):857-862. DOI 10.18699/VJ16.202

Поли(АДФ-рибозил) полимеразы-1 (ПАРП-1) – один из наиболее хорошо охарактеризованных членов семейства ядерных ферментов ПАРП (Ferraris, 2010). На сегодняшний день известно более 17 белков с подтвержденной или предполагаемой моно(АДФ-рибозил) и поли(АДФ-рибозил) трансферазной активностью, которые содержат «ПАРП сигнатурный мотив» в гомологичном каталитическом домене (Luo, Kraus, 2012). ПАРП-1 участвует более чем в 90 % случаев АДФ-рибозилирования внутри клетки и высоко эволюционно консервативен во всех высших эукариотах (Ferraris, 2010).

Считается, что каталитическая активность ПАРП-1 в клетках повышается при повреждениях ДНК (Zhu et al., 2013). В случае слабых повреждений ДНК ПАРП-1 участвует в процессе репарации и обеспечивает выживание клеток (Martin et al., 2000). При сильных повреждениях ДНК ПАРП-1 гиперактивируется и индуцирует расход клеточных НАД⁺ и АТФ, приводя к клеточной дисфункции или даже некротической клеточной смерти (Martin et al., 2000). Такие особенности ответа ПАРП-1 на повреждение ДНК и вовлеченность этого белка в процесс регуляции клеточной смерти позволили использовать фармакологическую модуляцию активности ПАРП-1 как эффективный подход для увеличения активности ДНК-связывающих противоопухолевых препаратов (Peralta-Leal et al., 2008). Также ингибиторы ПАРП-1 могут применяться в качестве монотерапии раковых опухолей с несколькими нарушенными механизмами репарации ДНК (Peralta-Leal et al., 2008). В частности, ингибиторы ПАРП-1 могут быть эффективными для лечения рака молочной железы, вызванной делециями в генах *BRCA 1/2* (Hay et al., 2005; Domagala et al., 2011).

ПАРП-1 катализирует ковалентное связывание поли(АДФ-рибозил) (ПАР) полимеров со своей субъединицей и с другими акцепторными белками, включая гистоны, белки репарации ДНК, транскрипционные факторы, модуляторы хроматина, используя НАД⁺ как донора АДФ-рибозы (Luo, Kraus, 2012). Активный сайт каталитического домена формируется последовательностью из 50 аминокислотных остатков и может быть разделен на два сайта: акцепторный и донорный. Акцепторный сайт захватывается АДФ-фрагментом поли(АДФ-рибозил) цепей, в то время как с донорным сайтом связывается НАД⁺. Донорный сайт можно разделить на три сайта: никотинамид-рибозил-связывающий, фосфат-связывающий и аденин-рибоза-связывающий (Kinoshita et al., 2004). Акцепторный сайт формируется спиральным N-терминальным доменом, состоящим из остатков 662–784, а донорный сайт включает C-терминальный домен 785–1010 (Ruf et al., 1996).

Большинство существующих ингибиторов ПАРП-1 содержат никотинамид-подобный фрагмент с модифицированной боковой группой (Basu et al., 2012). Такие ингибиторы связываются с сайтом посадки никотинамида и внешней стороной донорного сайта, осуществляя конкурентное ингибирование взаимодействия НАД⁺ с этим сайтом (Basu et al., 2012). Другим подходом к созданию искусственных низкомолекулярных ингибиторов ПАРП-1 является дизайн малых молекул на основе производных

природных соединений, таких как НАД⁺ (Sherstyuk et al., 2016). В частности, перспективным подходом может являться дизайн нового класса ингибиторов ПАРП-1, способных дополнительно с никотинамид-связывающим сайтом занимать фосфат- и аденин-рибоза-связывающие сайты. Такие соединения могут обладать улучшенными фармакокинетическими и фармакодинамическими свойствами. Однако для компьютерного дизайна ингибиторов этого класса необходима пространственная структура комплекса ПАРП-1 с НАД⁺ в донорном сайте, которая на сегодняшний день отсутствует. Описанные в научной литературе работы по предсказанию пространственной структуры такого комплекса позволили получить только грубую модель, в которой положение НАД⁺ определено с большой долей допущения. Так, в работах (Lin, 2007; Lee et al., 2010; Barkauskaite et al., 2015) положение НАД⁺ определяли путем совмещения пространственных структур белков из суперсемейства АДФ-рибозил трансфераз, кристаллизованных в комплексе с НАД⁺, со структурами ПАРП-1 без НАД⁺.

В настоящей работе проведено молекулярное моделирование мод связывания НАД⁺ с ПАРП-1 с учетом конформационных состояний ПАРП-1, что позволило построить четыре вероятные модели комплексов ПАРП-1/НАД⁺.

Материалы и методы

Выравнивание пространственных структур белков проводилось с помощью программного пакета Binding Site Alignment (Schrodinger Drug Discovery Suit). Минимизация потенциальной энергии моделей пространственных структур осуществлялась с использованием Schrodinger Drug Discovery Suit в силовом поле OPLS3 (Harder et al., 2015). Замены боковых остатков аминокислот и расчет энергии MM-GBSA выполнялись с помощью Prime (Schrodinger Drug Discovery Suit) (Jacobson et al., 2002, 2004). При подготовке пространственной структуры белка для молекулярного моделирования использовался Protein Preparation Wizard.

Результаты и обсуждение

Анализ донорного сайта в структуре ПАРП-1

В базе данных PDB (Protein Data Bank) приведены несколько кристаллических структур комплексов ПАРП-1 с различными конкурентными ингибиторами НАД⁺. В результате анализа пространственного совмещения этих структур были выявлены два кластера конформаций ПАРП-1. Их основными свойствами являлось наличие водородной связи, образуемой между остатками TYR889 и ASP770 (рис. 1). Первый кластер включал все структуры ПАРП-1, в которых данная связь наблюдалась, а в структурах второго кластера эта связь была разорвана, что соответствовало разным конформациям петли 883–893 а.о. В результате анализа были выбраны две пространственные структуры (идентификаторы PDB 4ZZZ и 4R6E), полученные с высоким разрешением, каждая из которых являлась представителем одного из рассмотренных кластеров. Сделано предположение, что при связывании с НАД⁺ ПАРП-1 может принять конформацию, характерную для одного из кластеров.

Для подготовки пространственных структур белка ПАРП-1 с идентификаторами 4ZZZ и 4R6E для дальнейшего анализа и молекулярного моделирования были проведены замены аминокислотных остатков, имеющих высокое значение В-фактора и высокую изменчивость своего положения при структурном выравнивании 3D структур ПАРП-1. Такими остатками оказались ARG878 и ASP766, входящие в сайт связывания аденозина, которые были заменены на аланин. Прием с аминокислотными заменами высокоподвижных остатков часто используется при анализе пространственных структур. Ориентация таких остатков может быть достаточно точно определена уже после предсказания положения лигандов в структуре анализируемых рецепторов, после чего осуществляются обратные замены на исходные аминокислоты.

Как следует из анализа пространственных структур гомологичных АДФ-рибозил трансфераз, связывание НАД⁺ может сопровождаться образованием водородных связей между молекулами воды в активном сайте и пирофосфатной группы НАД⁺. Для учета воды при дальнейшем анализе во внимание принимались только те молекулы воды, которые присутствовали более чем в пяти различных структурах ПАРП-1 на расстоянии меньшем 1 Å друг относительно друга после структурного наложения кристаллических структур ПАРП-1. Как видно из рис. 1, в сайте связывания фосфатной группы НАД⁺ наблюдается два стабильных кластера молекул воды. Интересным может быть положение сульфат-иона в пространственной структуре комплекса ПАРП-1 с низкомолекулярным ингибитором хав939 (PDB 4R5W) в сайте связывания пирофосфатной группы донорного НАД⁺. В данной кристаллической структуре сульфат-ион замещает положение одной из стабильных молекул воды (см. рис. 1, б). Из анализа ряда кристаллических структур АДФ-связывающих ферментов в аро-форме известно, что сульфат-ионы часто взаимодействуют с сайтами связывания фосфатных групп. В частности, совмещение пространственной структуры ПАРП *Gallus gallus* (PDB 1A26) (Ruf et al., 1996), полученной в комплексе с ак-

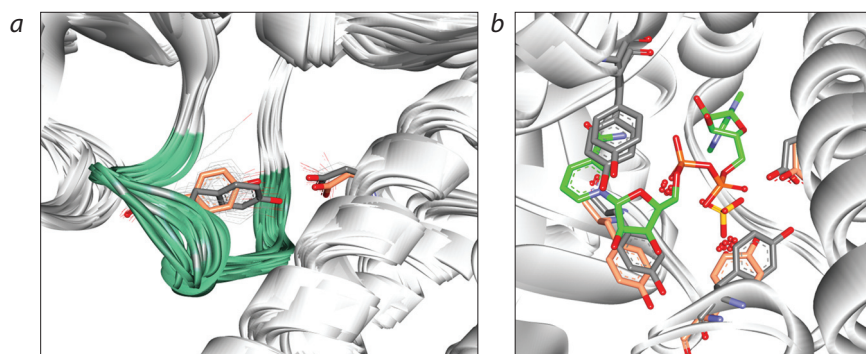


Fig. 1. Conformational states of the PARP-1 donor site.

(a) Structural superimposition of PARP-1 structures available in the PDB database. The mobile loop (883–893 aa) is shown in green, and amino acid residues TYR889 and ASP770, corresponding to PDB structures 4ZZZ and 4R6E, are shown in stick model. (b) Structural superimposition of PARP-1 crystal structures corresponding to PDB ID 4ZZZ and 4R6E, including conservative water molecules (ball-and-stick model, red color), donor NAD⁺ (stick model, green color) predicted from the eEF2 structure (2ZIT), and sulfate ion from the structure with PDB ID 4R5W (stick model).

цепторным НАД⁺, и структуры ПАРП-1 человека (PDB 4R5W) показывает, что сульфат-ион находится в акцепторном сайте посадки пирофосфатной группы НАД⁺.

В результате анализа конформаций донорного сайта ПАРП-1 и предположения о возможном замещении пирофосфатной группой НАД⁺ одной из молекул воды можно выдвинуть следующие модели по взаимодействию НАД⁺ с ПАРП-1:

- 1) НАД⁺ связывается с ПАРП-1 в конформации, соответствующей структуре 4ZZZ; при этом пирофосфатная группа находится в положении, соответствующем сульфат-иону структуры 4R5W;
- 2) НАД⁺ связывается с ПАРП-1 в конформации, соответствующей структуре 4ZZZ; при этом пирофосфатная группа образует водородные связи с двумя стабильными молекулами воды в донорном сайте связывания;
- 3) НАД⁺ связывается с ПАРП-1 в конформации, соответствующей структуре 4R6E; при этом пирофосфатная группа находится в положении, соответствующем сульфат-иону структуры 4R5W;
- 4) НАД⁺ связывается с ПАРП-1 в конформации, соответствующей структуре 4R6E; при этом пирофосфатная группа образует водородные связи с двумя стабильными молекулами воды в донорном сайте связывания.

Предсказание положения НАД⁺ в донорном сайте ПАРП-1

Прямое решение задачи по предсказанию положения НАД⁺ в комплексе с ПАРП-1 путем применения молекулярного докинга оказывается неэффективным в связи с тем, что НАД⁺ имеет большое число вращательных степеней свободы, а рецептор ПАРП-1, в свою очередь, может находиться не в оптимальной конформации для связывания с НАД⁺. В данной работе для предсказания структур комплексов ПАРП-1 с НАД⁺ в донорном сайте связывания использовался протокол фармакофор-ограниченного наложения с последующей минимизацией структур комплексов. На первом шаге, для предсказания начального положения НАД⁺ в донорном сайте ПАРП-1, проводилось структурное выравнивание ПАРП-1 с гомологичным белком – АДФ-рибозил трансферазой eEF2 *Saccharomyces cerevisiae* (PDB 2ZIT), находящейся в комплексе с НАД⁺ (Jørgensen et al., 2008). Положение НАД⁺ относительно ПАРП-1, полученное в результате совмещения пространственных структур ПАРП-1 и eEF2, давало наименьшее число стерических затруднений по сравнению с положениями НАД⁺, предсказанными при структурном выравнивании с другими гомологичными АДФ-рибозил трансферазами, кристаллизованными в комплексе с НАД⁺.

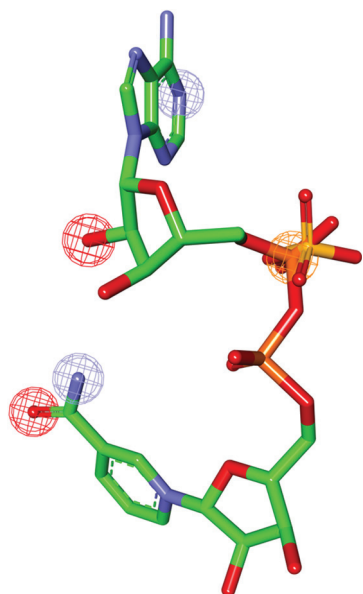


Fig. 2. Pharmacophore restraints used in minimization steps. Spheres represent physical restraints on the mobility of NAD⁺ atoms.

Pharmacophore restraints corresponding to the sulfate ion position were not applied to models 1 or 3.

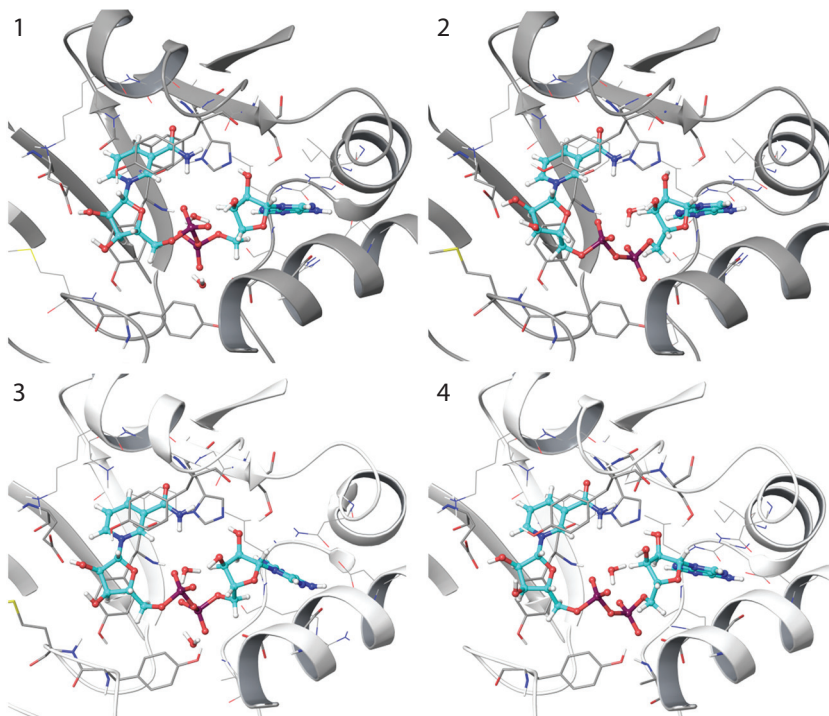


Fig. 3. Spatial structures of the predicted PARP1/NAD⁺ complexes corresponding to models 1–4.

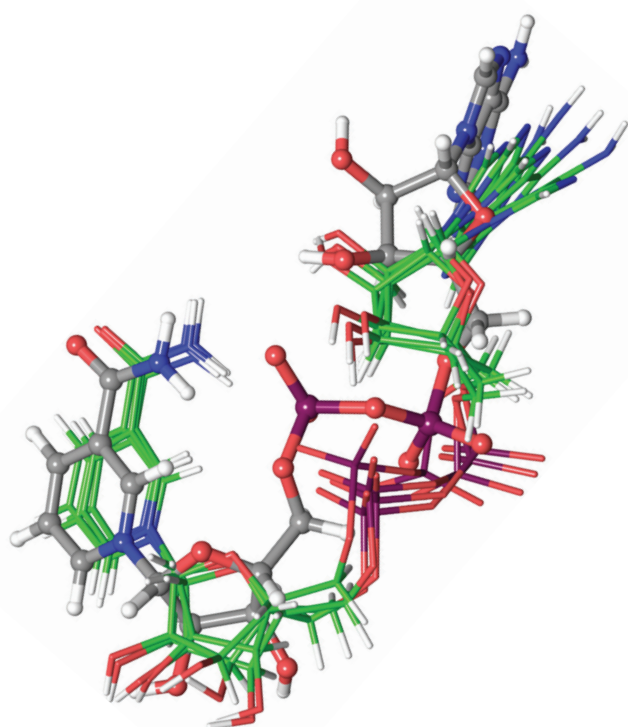


Fig. 4. Structural superimposition of NAD⁺ in conformations corresponding to models 1–4 (green color, stick model) and the NAD⁺ conformation corresponding to the NAD⁺/eEF2 complex (gray color, ball and stick model).

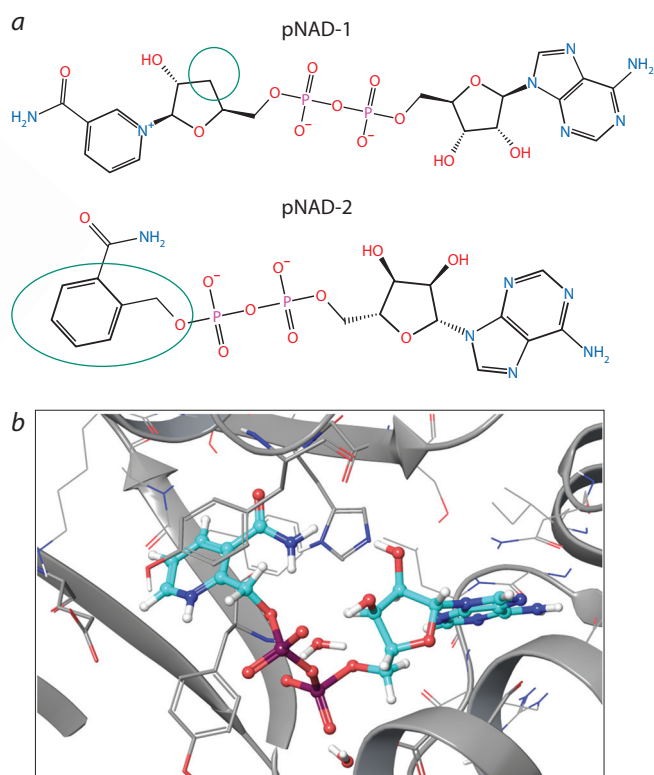


Fig. 5. Chemical structures of NAD⁺ derivatives designed as indicator molecules for functional validation of PARP-1/NAD⁺ binding models.

(a) chemical structures of NAD⁺ derivatives; (b) examples of the PARP-1 complex with indicator molecule pNAD-2 designed on the base of model 1.

Expected activities of pNAD-1 and pNAD-2 compounds in the predicted models

Compound	Model 1	Model 2	Model 3	Model 4
pNAD-1	$\Delta\Delta G = 4.6$ kcal/mol	$\Delta\Delta G = 3.5$ kcal/mol	$\Delta\Delta G = 7.18$ kcal/mol	$\Delta\Delta G = 2.87$ kcal/mol
pNAD-2	Activity	No activity	No activity	No activity

$\Delta\Delta G$ is the difference between the free energies of PARP-1 binding to NAD+ and pNAD-2 estimated by the MM-GBSA method.

На следующем шаге были получены модели пространственных структур для двух основных конформационных состояний с помощью минимизации потенциальной энергии с применением физических ограничений, соответствующих 3D фармакофорам взаимодействия НАД+ с АДФ-рибозил трансферазами. Ограничения на подвижность атомов НАД+, полученные при анализе фармакофорной структуры НАД+, показаны на рис. 2. Эти ограничения последовали из предположения, что НАД+ должен образовывать водородные связи с консервативными аминокислотными остатками, которые могут быть идентифицированы во всех опубликованных комплексах НАД+ с АДФ-рибозил трансферазами. Таким образом, во внимание были приняты водородные связи, образуемые между НАД+ и HIS862 (1 связь), GLY863 (2 связи), а также NH-группой скелета ARG878. Кроме того, во время минимизации накладывались физические ограничения в виде гармонического потенциала с константой жесткости 100 ккал/(моль $\cdot \text{\AA}^2$) на все тяжелые атомы белка и молекул воды. В моделях 2 и 4 были дополнительно рассмотрены ограничения на положение фосфатной группы, соответствующей положению сульфат-иона из структуры с идентификатором PDB 4R5W.

После построения всех четырех обсуждаемых выше моделей комплексов ПАРП-1 с донорным НАД+ была проведена их дальнейшая минимизация с наложением физических ограничений на смещение всех тяжелых атомов не более 0.3 $\text{\AA}$ (рис. 3). Согласно полученным моделям, положение пирофосфатной группы НАД+ стабилизируется образованием водородных связей с молекулами воды в сайте связывания. Как видно из рис. 3, в моделях пространственных структур 3 и 4, построенных с использованием структуры с идентификатором PDB 4R6E, водородная связь TYR889/ASP766 отсутствует, хотя потенциально TYR889 способен образовывать водородные связи с пирофосфатной группой НАД+. В случае же моделей, построенных на основе структуры 4ZZZ (модели 1 и 2), пирофосфатная группа НАД+ потенциально образует водородные связи только со стабильными молекулами воды, а также участвует в образовании внутримолекулярных водородных связей.

Одно из существенных отличий моделей с замещением пирофосфатной группой НАД+ сульфат-иона (модели 1 и 3) от моделей 2 и 4 состоит в отсутствии водородной связи гидроксильной группы аденин-рибозы с остовом аминокислотного остатка MET890. Конформация НАД+ очень схожа для моделей 2, 3 и 4 (рис. 4), однако в модели 1 видна существенная разница в конформации пирофосфатной группы для НАД+. В этой модели конформация пирофосфатной группы НАД+ позволяет образование

двух внутримолекулярных водородных связей с аденин-рибозой и амидной группой никотинамида.

Для экспериментальной проверки того, какая из четырех предложенных структурных моделей ПАРП-1/НАД+ может реально существовать, были предложены два новых соединения, являющихся производными НАД+, названные нами пНАД-1 и пНАД-2 (рис. 5).

Первое соединение (пНАД-1) рассчитывалось как индикатор корректности моделей 1 и 3 на основе проверки активности производных НАД+ без гидроксильной группы, способных образовывать водородные связи с NH-группой MET890. Второе соединение (пНАД-2) строилось на основе модификаций НАД+, при которых данная молекула, с одной стороны, обладала, согласно расчетам, способностью связываться с ПАРП-1, удовлетворяя фармакофорным ограничениям (см. рис. 2), а с другой – имела минимальную длину линкера между пирофосфатной группой и никотинамид-подобной группой. Минимальная длина такого линкера была в модели 1 – одна CH_2 -группа, в то время как для моделей 2, 3 и 4 длина линкера была больше. Расчетные оценки методом MM-GBSA разницы свободных энергий взаимодействия производных НАД+, которые могут использоваться для экспериментальной проверки рассмотренных выше моделей комплексов ПАРП-1/НАД+, приведены в таблице.

Расчеты показали, что пНАД-2 связывается с высокой аффинностью только с ПАРП-1, находящимся в состоянии, соответствующем модели 1. В свою очередь для моделей 1, 2 и 4 наблюдалось малое изменение аффинности связывания пНАД-1 по отношению к НАД+, в то время как для модели 3 этот показатель оказался существенно ниже (см. таблицу).

Таким образом, если в эксперименте наблюдается связывание ПАРП-1 с пНАД-2, то этот факт свидетельствует в пользу модели 1. Если же ПАРП-1 связывается с пНАД-1 значительно хуже, чем ПАРП-1 с НАД+, то это свидетельствует в пользу модели 3. К сожалению, верификация моделей 2 и 4 требует более сложного анализа связи между структурой и активностью лигандов.

Заключение

Для разработки нового класса НАД+-подобных ингибиторов ПАРП-1 чрезвычайно важно иметь корректно предсказанные модели взаимодействия НАД+ с ПАРП-1. В настоящей работе с использованием методов молекулярного моделирования были предсказаны структуры комплексов ПАРП-1 с НАД+ для двух наблюдаемых конформаций рецептора ПАРП-1. Дополнительно сделано предположение, что пирофосфатная группа может замещать положение сульфат-иона, наблюдаемого в структуре комплекса

ПАРП-1 с хав939. В результате были получены четыре модели связывания ПАРП-1 с НАД⁺, которые различаются по конформации пирофосфатной группы НАД⁺. Валидация моделей взаимодействия может быть проведена на основе проверки активности предложенных производных НАД⁺. Полученные модели доступны по адресу <http://www.bionet.sscs.ru/psd/nivanisenko/vogis2016/> и могут быть использованы для рационального дизайна ингибиторов ПАРП-1 на основе производных НАД⁺.

Acknowledgments

The construction and analysis of models of PARP-1–NAD⁺ complexes were supported by the Russian Science Foundation, project 4-44-00011 «Programmed cell death induced via death receptors: identification of apoptosis-inducing molecular mechanisms by molecular simulation».

The template-based comparison of the prototypes of active sites in NAD⁺-binding proteins was supported by the Russian Foundation for Basic Research, project 15-04-08065.

Conflict of interest

The authors declare no conflict of interest.

References

- Barkauskaite E., Jankevicius G., Ahel I. Structures and mechanisms of enzymes employed in the synthesis and degradation of PARP-dependent protein ADP-ribosylation. *Mol. Cell.* 2015;58(6):935-946. DOI 10.1016/j.molcel.2015.05.007.
- Basu B., Sandhu S.K., de Bono J.S. PARP Inhibitors. *Drugs.* 2012; 72(12):1579-1590.
- Domagala P., Huzarski T., Lubinski J., Gugala K., Domagala W. PARP-1 expression in breast cancer including BRCA1-associated, triple negative and basal-like tumors: possible implications for PARP-1 inhibitor therapy. *Breast Cancer Res. Treatment.* 2011;127(3):861-869. DOI 10.1007/s10549-011-1441-2.
- Ferraris D. Evolution of poly (ADP-ribose) polymerase-1 (PARP-1) inhibitors. From concept to clinic. *J. Med. Chemistry.* 2010;53(12): 4561-4584. DOI 10.1021/jm100012m.
- Harder E., Damm W., Maple J., Wu C., Reboul M., Xiang J.Y., Wang L., Lupyan D., Dahlgren M.K., Knight J.L., Kaus J.W., Cerutti D.S., Krilov G., Jorgensen W.L., Abel R., Friesner R.A. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Computation.* 2015;12(1):281-296. DOI 10.1021/acs.jctc.5b00864.
- Hay T., Jenkins H., Sansom O.J., Martin N.M.B., Smith G.C.M., Clarke A.R. Efficient deletion of normal *Brca2*-deficient intestinal epithelium by poly(ADP-ribose) polymerase inhibition models potential prophylactic therapy. *Cancer Res.* 2005;65:10145-10148. DOI 10.1158/0008-5472.
- Jacobson M.P., Friesner R.A., Xiang Z., Honig B. On the role of crystal packing forces in determining protein side-chain conformations. *J. Mol. Biol.* 2002;320:597-608. DOI 10.1016/S0022-2836(02)00470-9.
- Jacobson M.P., Pincus D.L., Rapp C.S., Day T.J.F., Honig B., Shaw D.E., Friesner R.A. A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function Bioinformatics.* 2004;55:351-367. DOI 10.1002/prot.10613.
- Jørgensen R., Wang Y., Visschedyk D., Merrill A.R. The nature and character of the transition state for the ADP-ribosyltransferase reaction. *EMBO Reports.* 2008;9(8):802-809. DOI 10.1038/embor.2008.90.
- Kinoshita T., Nakanishi I., Warizaya M., Iwashita A., Kido Y., Hattori K., Fujita T. Inhibitor-induced structural change of the active site of human poly(ADP-ribose) polymerase. *FEBS Lett.* 2004;556(1-3): 43-46. DOI 10.1016/S0014-5793(03)01362-0.
- Lee Y.M., Babu C.S., Chen Y.C., Milcic M., Qu Y., Lim C. Conserved structural motif for recognizing nicotinamide adenine dinucleotide in poly(ADP-ribose) polymerases and ADP-ribosylating toxins: implications for structure-based drug design. *J. Med. Chemistry.* 2010;53(10):4038-4049. DOI 10.1021/jm1001106.
- Lin H. Nicotinamide adenine dinucleotide: beyond a redox coenzyme. *Org. Biomol. Chemistry.* 2007;5(16):2541-2554. DOI 10.1039/B706887E.
- Luo X., Kraus W.L. On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Gen. Development.* 2012; 26(5):417-432. DOI 10.1101/gad.183509.111.
- Martin D.S., Bertino J.R., Koutcher J.A. ATP depletion + pyrimidine depletion can markedly enhance cancer therapy: fresh insight for a new approach. *Cancer Res.* 2000;60(24):6776-6783.
- Peralta-Leal A., Rodriguez M.I., Oliver F.J. Poly(ADP-ribose) polymerase-1 (PARP-1) in carcinogenesis: potential role of PARP inhibitors in cancer treatment. *Clin. Transl. Oncol.* 2008;10(6):318-323.
- Ruf A., de Murcia J.M., de Murcia G., Schulz G.E. Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken. *Proc. Natl. Acad. Sci.* 1996;93(15):7481-7485.
- Sherstyuk Y.V., Zakharenko A.L., Kutuzov M.M., Chalova P.V., Sukhanova M.V., Lavrik O.I., Silnikov V.N., Abramova T.V. A versatile strategy for the design and synthesis of novel ADP conjugates and their evaluation as potential poly(ADP-ribose) polymerase 1 inhibitors. *Mol. Diversity.* 2016;1-13. DOI 10.1007/s11030-016-9703-x.
- Zhu Q., Wang X., Chu Z., He G., Dong G., Xu Y. Design, synthesis and biological evaluation of novel imidazo[4, 5-c]pyridinecarboxamide derivatives as PARP-1 inhibitors. *Bioorg. Med. Chem. Letters.* 2013;23(7):1993-1996. DOI 10.1016/j.bmcl.2013.02.032.



Протокол анализа количественных характеристик опушения листа картофеля

А.В. Дорошков , М.А. Генаев, Д.А. Афонников

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

Опушение листьев у растений – один из важных признаков, отвечающих за формирование микроклимата у поверхности листа, – участвует в защите от неблагоприятных биотических и абиотических факторов среды. У пасленовых, к которым относится картофель *Solanum tuberosum* L., опушение листа представлено многоклеточными неветвистыми трихомами разного размера и морфологии. Опушение этой культуры способствует сопротивлению растения насекомым-вредителям, в частности колорадскому жуку и тле, которая является переносчиком вирусных заболеваний. В процессе селекционно-генетических экспериментов возникает необходимость оценить интенсивность опушения листьев растений картофеля. Для этого, как правило, используют микроизображения или микрофотографии листьев листа, сделанные под микроскопом. С их помощью оценивают количество трихом разного типа на поверхности листа и затем нормируют на единицу площади. Такой подход трудоемок, поскольку требует визуального подсчета трихом под микроскопом. Разработанный нами протокол описывает технологию быстрой количественной оценки характеристик опушения у картофеля (числа трихом на поверхности листа и их средней длины) для решения задач генетики и селекции этой культуры. Он состоит из описания технологии приготовления препаратов и получения цифровых изображений сгибов листа с помощью оптического микроскопа в проходящем свете и последующей автоматической обработки этих изображений на компьютере с помощью программы LHDetect2.

Ключевые слова: высокопроизводительное фенотипирование; анализ изображений; *Solanum tuberosum* L.; трихомы.

A protocol for analysis of the quantitative characteristics of leaf pubescence in potato

A.V. Doroshkov , M.A. Genaev, D.A. Afonnikov

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Plant leaf pubescence is one of the important features, which is responsible for microclimate formation near the epidermis. It is involved in protection against adverse biotic and abiotic environmental factors. In *Solanaceae*, to which belongs the potato *Solanum tuberosum* L., leaf pubescence appears as multicellular unbranched trichomes of diverse size and morphology. Pubescence of this plants promotes resistance to insect pests, in particular, Colorado potato beetle and aphid, which is a carrier of viral diseases. During the process of breeding and genetic experiments, there is a need to assess the intensity of leaf pubescence of potato plants. For this task, micrographs taken under a microscope are commonly used. They are used to count different types of trichomes on the leaf surface to characterize the intensity of potato leaf pubescence. This approach requires visual counting of trichomes under a microscope and is fairly laborious. This protocol describes a rapid technology for quantitative assessment of the characteristics of potato pubescence (the number of trichomes on the leaf surface and the average length of trichomes) to solve the problems of genetics and breeding of this plant. It consists of a preparation technology, digital imaging of leaf folds with an optical microscope in transmitted light and subsequent automatic processing of images using the LHDetect2 software.

Key words: high-throughput phenotyping; image analysis; *Solanum tuberosum* L.; trichomes.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Дорошков А.В., Генаев М.А., Афонников Д.А. Протокол анализа количественных характеристик опушения листа картофеля. Вавиловский журнал генетики и селекции. 2016;20(6):863-868. DOI 10.18699/VJ16.218

HOW TO CITE THIS ARTICLE:

Doroshkov A.V., Genaev M.A., Afonnikov D.A. A protocol for analysis of the quantitative characteristics of leaf pubescence in potato. Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding. 2016;20(6):863-868. DOI 10.18699/VJ16.218

METHODS AND PROTOCOLS

Received 25.11.2016

Accepted for publication 23.12.2016

© AUTHORS, 2016



e-mail: ad@bionet.nsc.ru

Опушение – это набор поверхностных образований – трихом. Трихомы растений бывают несекретирующие и секретирующие (выделяют секрет во внешнюю среду). Они участвуют в формировании микроклимата у поверхности листа, в защите от неблагоприятных биотических и абиотических факторов среды. Особый интерес представляют железистые трихомы картофеля и его сородичей: они способны обеспечивать защиту растения от насекомых, в частности колорадского жука, картофельной блохи и картофельной цикадки (Gregory et al., 1986; Flanders et al., 1992; Maharijaya, Vosman, 2015). Показана вариация в наборе типов трихом между видами пасленовых (Schilmiller et al., 2008), что определяет разную степень устойчивости к поражению вредителями.

Было выявлено, что опушение дикого картофеля *Solanum berthaultii* Hawkes. значительно более плотное и более эффективно в борьбе с насекомыми, чем *S. tuberosum* (Tingey et al., 1982). Плотность расположения трихом – один из основных факторов, коррелирующих с устойчивостью к насекомым-вредителям (Kennedy, 2003; Simmons, Gurr, 2005). Плотное опушение железистыми трихомами, эффективное в защите от вредителей, – перспективный признак для селекции (Tingey et al., 1982; Фасулати и др., 2011).

Известно, что применение современных методов высокопроизводительного фенотипирования позволяет получить более быстрое точное описание морфологических признаков, чем ранее используемые методы (Афонников и др., 2016). С использованием такого подхода нами было показано, что между сортами картофеля существует большая вариация по признакам плотности и длины опушения (Дорошков и др., 2016). Эти признаки могут быть использованы для предсказания устойчивости генотипов картофеля к насекомым, а также в широкомасштабной селекционной работе.

В процессе анализа опушения картофеля нами проведена оптимизация параметров распознавания изображений с учетом особенностей морфологии трихом картофеля и разработан протокол приготовления препаратов для получения изображений. Полученный метод, описание которого дано в настоящей статье, позволяет производить быструю количественную оценку характеристик опушения листа картофеля для решения задач генетики и селекции этой культуры.

1. Пробоподготовка

1.1. Материалы

Для сбора образцов необходимы ножницы и герметичный контейнер для транспортировки. Для приготовления препаратов потребуются предметные стекла, спирт и салфетка для их протирки, а также липкая лента для фиксации образца. Для микроскопии необходим световой микроскоп, оснащенный цифровой камерой.

1.2. Растения и время сбора образцов

Для сбора образцов используют здоровые растения картофеля, выращенные в поле или теплице. Сбор материала производят в период цветения.

1.3. Сбор и хранение листьев

Сложные листья картофеля собирают в поле или теплице и опускают черешком в воду для поддержания тургора. Оптимальный размер листовой пластинки для анализа 4–6 см в длину. Это соответствует взрослому листу 4–5-го яруса начиная от точки роста побега. Лист должен быть здоровым и не поврежденным болезнями или вредителями. Препараты для анализа необходимо приготавливать на свежем материале. Допускается хранение листьев, опущенных основанием в воду не более шести часов.

1.4. Подготовка микропрепаратов сгиба листа

Для приготовления препаратов сгибов листа вырезаются фрагменты, заключенные между жилок (рис. 1, а). Фрагмент размером около 10×20 мм сгибают пополам так, чтобы анализируемая поверхность листа (абаксиальная или адаксиальная) оказалась снаружи и помещается на предметное стекло (см. рис. 1, б, в). Фрагмент фиксируется липкой лентой так, чтобы сгиб оказался свободен. Отступ края ленты от сгиба не более 2–3 мм. Для надежной оценки опушения растения следует подготовить и проанализировать препараты как минимум трех сгибов листа. Для минимизации ошибки, связанной с условиями формирования листа, рекомендуется использовать для приготовления каждого сгиба новый лист с того же растения.

2. Получение изображений

Изображения получают при помощи светового микроскопа в проходящем свете (рис. 2, а). Для фотографирования используют объектив малого увеличения (3×–10×). Оптимальное значение увеличения объектива подбирается с учетом увеличения адаптера и размера матрицы камеры. Программа LHDetect2 (Genaev et al., 2012) оптимизирована для работы с изображениями, пиксель в которых примерно соответствует размеру 2×2 мкм. Микроскоп должен иметь возможность обеспечить равномерное освещение всего поля зрения объектива. Это достигается использованием специального конденсора для малого увеличения либо выводом из оптической системы фронтальной линзы конденсора (см. рис. 2, б, в).

Такая методика позволяет получать в проходящем свете контрастное изображение трихом на светлом фоне. Во время съемки апертурная диафрагма должна быть закрыта до минимального значения. Это обеспечит четкое изображение трихом, искривленных в пространстве. Полевая диафрагма, напротив, должна быть максимально открыта для предотвращения появления затемнения углов кадра. После настройки микроскопа необходимо найти сгиб и обеспечить позиционирование границы листа в вертикальном направлении так, чтобы область листа (темная) располагалась справа, а область фона – слева (рис. 3).

Настройкой интенсивности света и экспозиции камеры и позиции тубуса (микро- и макровинтом) необходимо добиться контрастного изображения границы сгиба и трихом. При этом фон должен быть белым (иметь максимальные значения интенсивности, т. е. переэкспонирован), а лист черным (иметь минимальные значения интенсивности, т. е. недоэкспонирован) (см. рис. 3).

Изображения должны быть сохранены в формате TIFF, BMP, PNG, GIF или JPG без сжатия. Рекомендуется по

возможности избегать использования формата JPG, так как даже минимальное сжатие ухудшает распознавание границ объектов.

3. Компьютерная обработка изображений

3.1. Описание модифицированного метода LHDetect2 для распознавания маленьких секретирующих трихом

Обработка изображений проводится программой LHDetect2. Эта программа распознает на стандартном изображении (см. рис. 3, б) вершины и основания трихом и измеряет их длину.

Алгоритм, позволяющий анализировать трихомы на изображении сгиба листа, был опробован на пшенице (Genaev et al., 2012) и табаке. Опушение у пшеницы формируется совокупностью несекретирующих одноклеточных трихом. Опушение картофеля формируется трихомами двух типов: несекретирующими, которые по форме сходны с трихомами пшеницы, и маленькими секретирующимися трихомами (см. рис. 3). Ранее нами для распознавания секретирующих головок трихом, характерных для картофеля, табака и томата, были внесены модификации параметров, позволившие их распознавать.

Рассмотрим особенности применения LHDetect2 для картофеля. Размер секретирующих трихом этого растения относительно мал, и вершины трихом имеют кривизну, обусловленную наличием секретирующей головки. Для преодоления проблемы размера был изменен постоянный параметр «минимальная длина трихомы» с 8 до 4 пикселей. Для поиска вершин трихом в объекте алгоритм последовательно один за другим обходит пиксели, принадлежащие его границе. На каждом шаге обхода для текущего пикселя границы определяются еще два, удаленных вдоль границы от него на N соседних пикселей: один по направлению обхода, другой – в противоположном направлении. После этого измеряют величину угла с вершиной в центральном пикселе, образованном двумя соседними пикселями. Если она меньше порогового значения α , средний пиксель считается потенциальной вершиной

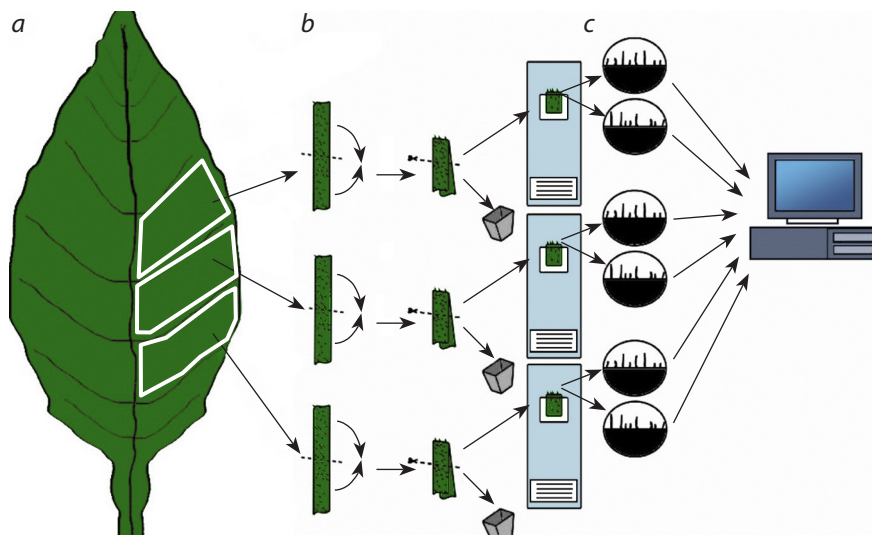


Fig. 1. Step-by-step preparation of potato trichome images for phenotyping: (a) excision of fragments; (b) slide preparation; (c) photography and image processing. From (Doroshkov et al., 2016), with modifications.



Fig. 2. Equipment for imaging: Zeiss Axioscop 2 plus microscope with an Axiocam HRC camera (2/3 inch CCD sensor) and an adapter with 0.63 magnification. Field lens 5x.

(a) With this magnification, a pixel in an image with the resolution 1300×1030 corresponds to the size $2.1 \times 2.1 \mu\text{m}$. (b, c) Removal of the frontal condenser lens from the optical path.

трихомы. Данный алгоритм, следовательно, относит к потенциальным вершинам пиксели, в которых линия границы объекта имеет большую кривизну. В случае анализа опушения пшеницы параметр N был зафиксирован: $N = 2$. Размеры головок у трихом табака могут сильно различаться, что приводит к тому, что использование значения параметра N (определяющего расстояние между тремя пикселями вдоль границы листа для определения угла ее изгиба), одинакового для всех трихом, не позволяет точно идентифицировать положение их вершин. Поэтому они плохо определялись версией алгоритма, разработанной для пшеницы. Решить эту проблему позволила модификация алгоритма, разработанного первоначально для пшеницы. В результате модификации значения параметра N для каждой трихомы в процессе поиска ее вершины могут меняться в пределах от N_{min} до N_{max} , где N_{min} и N_{max} – целые

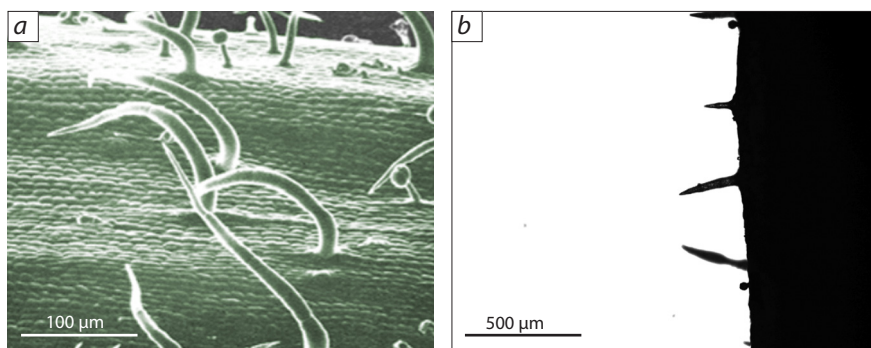


Fig. 3. Morphology of cultivated potato *S. tuberosum* trichomes: (a) in an electron microscopic image from (Tingey et al., 1982) and (b) in an image of a leaf fold in transmitted light, suitable for computerized processing.

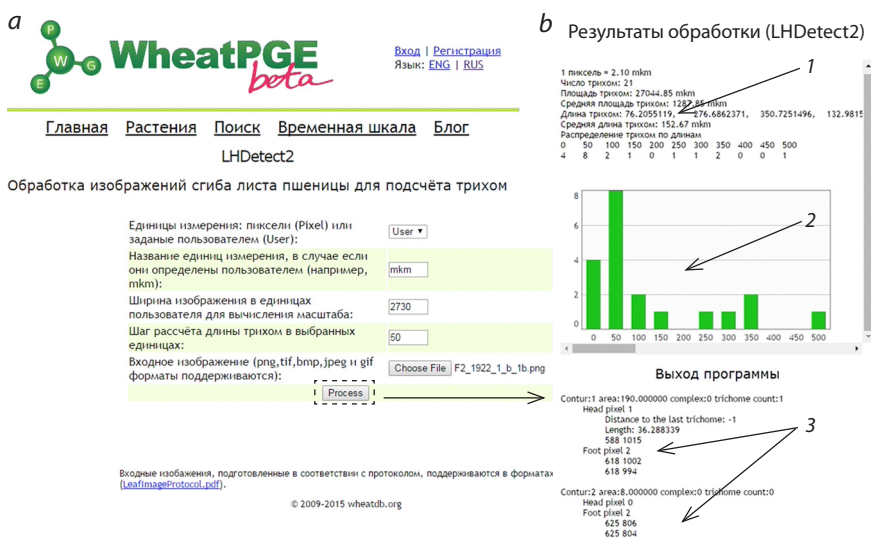


Fig. 4. LHDetect2_potato program: (a) interface; (b) exemplary image processing output.

числа, которые являются параметрами алгоритма. На этапе поиска вершины трихомы, как и в случае с анализом трихом пшеницы, производится поиск пикселей границы трихомы, которые расположены в области максимальной кривизны линии границы. Для текущего пикселя границы определяются пары пикселей, удаленные от анализируемого пикселя на N пикселей (вперед и назад по направлению обхода). При этом значения N последовательно увеличиваются от N_{min} до N_{max} с шагом 1. Для каждой такой пары пикселей измеряется величина угла α_i с вершиной в тестируемом пикселе. Если на определенном шаге α_i меньше порогового значения α , перебор останавливается и тестируемый пиксель считается потенциальной вершиной трихомы.

3.2. Работа с программой LHDetect2

Интерфейс веб-версии программы, открытой и доступной по адресу http://ru.wheatdb.org/LHDetect2_potato, показан на рис. 4, а. Для запуска анализа необходимо выбрать параметры и загрузить изображение, пригодное для анализа (http://wheatdb.org/static/images/lhdetect/F2_1922_1_b_1b.png). В поле 1 программа предлагает выбрать единицы измерения: Pixel или User. В первом случае программа приведет измерения размера трихом в пикселях, во втором пересчитает в величину, заданную пользователем. Для задания собственных единиц измерения нужно ввести их название в поле 2, например мкм, и ширину изображения в этих единицах в поле 3 для вычисления масштаба.

Подразумевается, что оптическая система микроскопа не вносит искажений, длина пикселя изображения соответствует его ширине.

Важно: Во избежание появления ошибок при сравнении данных в пределах одного эксперимента необходимо использовать изображения одинакового масштаба, например, полученные с идентичных систем микроскоп – камера.

После выбора единиц измерения следует в поле 4 задать интервал расчета длины трихом в выбранных единицах. Этот параметр отвечает за предварительную обработку. Данные для каждого изображения будут представлены в виде распределения трихом по интервалам длин, т. е. числового массива, каждым элементом которого является число трихом в соответствующем интервале длин. Интервал выбирают согласно задачам анализа: при малом количестве измерений для описания выборки (1–5 препаратов) следует выбирать шаг в 40–60 мкм, при более детальном анализе (более 20 препаратов) возможно уменьшение шага вплоть до 10 мкм. Шаг меньшего размера использовать нецелесообразно, поскольку большое число интервалов гистограммы, содержащих нулевые значения, приведет к неинформативности анализа. После задания всех параметров следует загрузить файл для обработки кнопкой Choose File и нажать Process.

3.3. Структура результатов программы LHDetect2

Программа LHDetect2 определяет на изображении совокупность объектов, каждый из которых является одной или несколькими трихомами на поверхности листа. Объект содержит несколько трихом, если на изображении они перекрываются. После идентификации объектов программа измеряет длину каждой распознанной в составе объекта трихомы и выводит эти значения в единицах измерения, указанных пользователем. На основании этих данных рассчитывают число трихом и их распределение по интервалам длин. Программа автоматически приводит гистограмму этого распределения. Помимо этого, LHDetect2 измеряет площадь, занятую каждым объектом на изобра-

жении, и общую площадь трихом. Площадь трихомы на изображении пропорциональна ее объему в пространстве, этот параметр позволяет сравнить объемы трихом разных образцов.

Пример получения количественных характеристик опушения из изображения сгиба листа

Используя стандартное изображение сгиба листа, мы выставили следующие параметры (см. рис. 4, а): единицы измерения – User (заданы пользователем), название единиц – мкм (mkm), размер изображения в 2730 мкм. Шаг расчета длины трихом – 50 мкм.

В результате анализа изображения получаем суммарные данные (см. рис. 4, б, 1), гистограмму распределений трихом по длине (см. рис. 4, б, 2), а также данные о каждой трихоме (см. рис. 4, б, 3).

Характеристика полученных данных:

- 1 пиксель = 2.10 мкм – размер пикселя изображения в единицах, заданных пользователем (микрометры);
- число трихом: 21 – число трихом, распознанных на изображении;
- площадь трихом: 27044.85 мкм² – суммарная площадь трихом в единицах, заданных пользователем;
- средняя площадь трихом: 1287.85 мкм² – усредненная площадь трихом в единицах, заданных пользователем;
- длина трихом:

76.205,	276.686,	350.725,	132.981,	68.015,
33.665,	23.1,	149.797,	27.766,	95.068,
87.035,	86.902,	48.795,	336.104,	84.583,
189.426,	61.395,	500.732,	374.189,	50.247

- длины всех распознанных на изображении трихом в единицах, заданных пользователем;
- средняя длина трихом: 152.67 мкм – усредненная длина трихом в единицах, заданных пользователем;
- распределение трихом по длинам:

0	50	100	150	200	250	300	350	400	450	500
4	8	2	1	0	1	1	2	0	0	1

В первой строке отмечены значения длины начала интервала, во второй – количество трихом, попавших в этот интервал. На приведенном интервале можно заметить, что четыре трихомы принадлежат интервалу длин (0; 50) мкм, восемь трихом – интервалу (50; 100) мкм, две трихомы имеют длину в интервале (100; 150) мкм.

Далее приведены значения площади (Area), длины (Length), расстояния до соседней трихомы (Distance to the last trichome), координат верхушки (Head), основания (Foot) для каждой трихомы. Значение Complex = 1 означает, что трихома пересекается на изображении с другой трихомой. Пример для двух трихом:

Contour:3 area:923.500 complex:0 trichome count:1

Head pixel 1

Distance to the last trichome: 203.899

Length: 131.755

507 753

Foot pixel 2

625 802

625 792

Contour:4 area:2292.000000 complex:1 trichome count:2

Head pixel 2

Distance to the last trichome: 77.242

Length: 167.011

457 719

Distance to the last trichome: 0.000

Length: 63.324

563 755

Foot pixel 4

625 747

624 729

624 726

624 717

4. Анализ данных

В результате обработки серии изображений получаются серии данных, структурированных, как показано на рис. 4, б. Основные характеристики опушения – длина трихом и их количество – являются удобными для работы признаками, а распределение трихом по длинам позволяет провести подробный анализ морфологии опушения листа. Для надежной оценки опушения листа конкретного растения или листа необходимо сделать серию сгибов (не менее трех) и получить по два изображения каждого сгиба. Это уменьшит ошибку, связанную с вероятностью попадания трихом на сгиб (Genaev et al., 2012).

Рассмотрим пример анализа данных, полученных в ходе обработки образцов 35 сортов картофеля из работы А.В. Дорошкова и коллег (2016). Для работы с данными распределения трихом по длинам необходимо перенести в файл Excel и усреднить значения серии измерений одного и того же генотипа по каждому интервалу. Пример таблицы данных приведен в Доп. материалах¹ (рис. 5). Полученные результаты могут быть использованы для сравнения количества трихом разных интервалов длин между генотипами. На рисунке видно, что у сортов Брянский деликатес и Гусар отсутствуют трихомы менее 42 мкм, тогда как у сортов Великан и Вымпел эти трихомы представлены. Из таких распределений легко получить оценку среднего числа трихом для сорта (сумма всех элементов строки) и средней длины трихом (сумма произведений элементов строки на середину интервала длин). Примеры далее приведены по данным (Дорошков и др., 2016).

Пример расчета средней длины и числа трихом

Среднее число трихом для сорта Великан = $G3 + H3 + I3 + J3 + K3 + L3 + \dots$ (см. Доп. материалы (рис. 5)); средняя длина трихом для сорта Великан = $G3 \times G1 + H3 \times H1 + I3 \times I1 + J3 \times J1 + K3 \times K1 + L3 \times L1 + \dots$

Для любой экспериментальной выборки сортов или растений эти данные информативно представлять в виде распределений и диаграммы рассеяния.

Пример построения гистограмм распределения сортов по количеству и средней длине трихом

Пример приведен на рис. 6 (см. Доп. материалы). Для описания распределений и их сравнения между выборками необходимо рассчитывать характеристики каждого из распределений, такие как минимум, максимум, медиана, среднее значение, дисперсия, стандартное отклонение. Взаимодействие двух признаков можно оценить с помо-

¹ Дополнительные материалы см. в Приложении 3 по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2016-20/appx6.pdf>

щью корреляционного анализа и диаграммы рассеяния. Диаграмма рассеяния средней длины и числа трихом приведена в Доп. материалах (рис. 7). В выборке данных из примера не наблюдается значимой линейной зависимости между числом трихом и их средней длиной: коэффициент корреляции составил 0.147 ($p < 0.39$).

Анализ сходства сортов по распределениям трихом по длинам можно проводить при помощи метода главных компонент либо кластерного анализа. В Доп. материалах (см. рис. 7, б) приведен пример анализа главных компонент для характеристик опушения у 35 сортов картофеля из коллекции СибНИИРС в многомерном пространстве числа трихом в различных интервалах длин. На две главные компоненты приходится 58 % дисперсии: 43 % на первую и 15 % на вторую.

Коэффициент корреляции Пирсона между весами первой компоненты и числом трихом составил 0.9, а длиной трихом -0.43 . Таким образом, увеличение первой компоненты характеризует увеличение числа трихом при уменьшении их длины. Из рисунка видно, что наибольшее значение первой компоненты наблюдается для наиболее опушенных сортов (Великан, Танай, Кемеровчанин, Танго), которые преимущественно имеют умеренные и малые длины трихом. Что касается второй компоненты, то оказалось, что ее веса отрицательны для числа трихом в интервалах длин 42–200 мкм и положительны для числа трихом в интервалах длин 200–400 мкм. Таким образом, вторая компонента отражает преобладание в опушении трихом умеренной длины. Например, у сорта Танго преобладают короткие трихомы (~100 мкм), а у сорта Великан – умеренной длины (~200 мкм). Сорта Великан, Танай, Кемеровчанин и Танго характеризуются наибольшим разбросом в пространстве главных компонент, т.е. демонстрируют наибольшие отклонения по свойствам опушения листьев по отношению к основной массе исследованных растений. Умеренная доля объясненной дисперсии для первых двух компонент (~60 %) указывает, что изменчивость опушения листьев картофеля является комплексной величиной и не может быть сведена к изменению лишь небольшого числа параметров, таких как число трихом и их средняя длина.

Методы высокопроизводительного фенотипирования заняли важную нишу в генетической и селекционной работе. Они позволяют делать масштабные эксперименты со значительной экономией ресурсов (Афонников и др., 2016). Настоящий подход к анализу количественных характеристик опушения был неоднократно использован нами и показал себя как эффективный метод не только в задачах анализа коллекций (Дорошков и др., 2016), но и в решении эволюционных (Pshenichnikova et al., 2016) и генетических (Genaev et al., 2012; Doroshkov et al., 2016) задач. Представленный протокол делает доступным использование метода анализа количественных характеристик опушения листа картофеля широкой аудитории генетиков и селекционеров, работающих с картофелем.

Acknowledgments

This work was supported by State Budgeted Project 0324-2016-0006 “Collection of crop genotypes for academic

research in plant genetics and development of genetic technologies for marker-assisted and genomic selection”. The development of the modified algorithm for phenotyping potato trichomes and of LHDetect2_potato software was supported by the Russian Foundation for Basic Research, project 16-37-00304-mol_a. The authors are grateful to the staff of the Shared Access Center for Microscopy of Biologic Objects, Novosibirsk for access to the equipment.

Conflict of interest

The authors declare no conflict of interest.

References

- Afonnikov D.A., Genaev M.A., Doroshkov A.V., Komyshev E.G., Pshenichnikova T.A. Methods of high-throughput phenotyping of plants for large-scale breeding experiments. *Genetika = Genetics* (Moscow). 2016;52(7):788-803. DOI 10.7868/s001667581607002x. (in Russian)
- Doroshkov A.V., Afonnikov D.A., Dobrovolskaya O.B., Pshenichnikova T.A. Interactions between leaf pubescence genes in bread wheat as assessed by high throughput phenotyping. *Euphytica*. 2016; 207(3):491-500. DOI 10.1007/s10681-015-1520-2.
- Doroshkov A.V., Simonov A.V., Safonova A.D., Afonnikov D.A., Likhnenko I.E., Kolchanov N.A. Quantitative assessment of potato pubescence by analysis of digital microimages. *Dostizheniya nauki i tekhniki APK = Achievements of Science and Technology of AIC*. 2016;30(10):12-14. (in Russian)
- Fasulati S.R., Limantseva L.A., Ivanova O.V., Rogozina E.V. Extensive drug resistance to Colorado potato beetle, potato ladybird, and yellow potato cyst nematode. *Zashchita i karantin rasteniy = Plant Protection and Quarantine*. 2011;10. (in Russian)
- Flanders K.L., Hawkes J.G., Radcliffe E.B., Lauer F.I. Insect resistance in potatoes: sources, evolutionary relationships, morphological and chemical defenses, and ecogeographical associations. *Euphytica*. 1992;61(2):83-111. DOI 10.1007/BF00026800.
- Genaev M.A., Doroshkov A.V., Pshenichnikova T.A., Kolchanov N.A., Afonnikov D.A. Extracting quantitative characteristics of wheat leaf hairiness using image processing technique. *Planta*. 2012;236(6): 1943-1954. DOI 10.1007/s00425-012-1751-6.
- Gregory P., Tingey W.M., Ave D.A., Bouthyette P.Y. Potato glandular trichomes: a physicochemical defense mechanism against insects. In: Maurice B., Green P., Hedin A. (eds) *Natural resistance of plants to pests*. Am. Chemical Society. Washington, 1986;296:160-167.
- Kennedy G.G. Tomato, pests, parasitoids, and predators: tritrophic interactions involving the genus *Lycopersicon*. *Ann. Rev. Entomol.* 2003;48(1):51-72. DOI 10.1146/annurev.ento.48.091801.112733.
- Maharajaya A., Vosman B. Managing the Colorado potato beetle; the need for resistance breeding. *Euphytica*. 2015;204(3):487-501. DOI 10.1007/s10681-015-1467-3.
- Pshenichnikova T.A., Doroshkov A.V., Simonov A.V., Afonnikov D.A., Börner A. Diversity of leaf pubescence in bread wheat and relative species. *Gen. Resour. Crop Evol.* 2016;1-13. DOI 10.1007/s10722-016-0471-3.
- Schilmiller A.L., Last R.L., Pichersky E. Harnessing plant trichome biochemistry for the production of useful compounds. *Plant J.* 2008; 54(4):702-711. DOI 10.1111/j.1365-3113.2008.03432.x.
- Simmons A.T., Gurr G.M. Trichomes of *Lycopersicon* species and their hybrids: effects on pests and natural enemies. *Agricultur. Forest Entomol.* 2005;7(4):265-276. DOI 10.1111/j.1461-9555.2005.00271.x
- Tingey W.M., Plaisted R.L., Laubengayer J.E., Mehlenbacher S.A. Green peach aphid resistance by glandular trichomes in *Solanum tuberosum* x *S. berthaultii* hybrids. *Am. Potato J.* 1982;59(6):241-251. DOI 10.1007/BF02856560.



Исследование влияния электромагнитного излучения терагерцового диапазона на протеом экстремофильной археи *Halorubrum saccharovororum*

Т.Н. Горячкова¹✉, С.Г. Константинова¹, И.А. Мещерякова¹, С.В. Банникова¹, Е.А. Демидов¹, А.В. Брянская¹, М.А. Щеглов², А.И. Семенов², Д.Ю. Ощепков¹, В.М. Попик², С.Е. Пельтек¹

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное бюджетное учреждение науки Институт ядерной физики им. Г.И. Будкера Сибирского отделения Российской академии наук, Новосибирск, Россия

В последние десятилетия в связи с появлением новых источников терагерцового излучения и разработкой приборов на его основе проводятся исследования воздействия излучения этого диапазона на живые объекты. Использование терагерцового излучения в системах безопасности контроля и досмотра, диагностическом медицинском оборудовании и научных исследованиях теоретически обусловлено малой энергией кванта этого излучения, не вызывающей негативных последствий при контакте с живыми системами, как это происходит при применении излучения более высокоэнергетических диапазонов. В этой связи изучение эффектов нетермического воздействия терагерцового излучения на живые объекты представляет актуальную задачу. Целью настоящей работы была идентификация комплекса белков, участвующих в ответе клеток археи *Halorubrum saccharovororum* H3, выделенной из природной экстремальной среды обитания, на электромагнитное излучение терагерцового диапазона. Для облучения терагерцовым излучением бактериальных и архейных культур нами разработан микрофлюидная система. Проведено 5-часовое облучение галофильной термостабильной археи *H. saccharovororum* терагерцовым излучением с длиной волны 130 мкм при плотности мощности излучения 0.8 Вт/см². Методами протеомного анализа, включающими двумерный электрофорез с последующей MALDI-TOF масс-спектрометрией, проведена идентификация белков, изменяющих уровень экспрессии в результате нетермического воздействия терагерцового излучения. Выявлено 16 белковых фракций, которые достоверно различаются по уровню концентрации белка (более чем в полтора раза). Полученные данные свидетельствуют о том, что клетки *Halorubrum* реагируют на нетермическое воздействие терагерцового излучения путем изменения экспрессии генов, контролирующих регуляцию системы трансляции.

Ключевые слова: лазер на свободных электронах; терагерцовое излучение; микрофлюидная система; галофильная термостабильная архея.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Горячкова Т.Н., Константинова С.Г., Мещерякова И.А., Банникова С.В., Демидов Е.А., Брянская А.В., Щеглов М.А., Семенов А.И., Ощепков Д.Ю., Попик В.М., Пельтек С.Е. Исследование влияния электромагнитного излучения терагерцового диапазона на протеом экстремофильной археи *Halorubrum saccharovororum*. Вавиловский журнал генетики и селекции. 2016;20(6):869-875. DOI 10.18699/VJ16.205

HOW TO CITE THIS ARTICLE:

Goryachkovskaya T.N., Konstantinova S.H., Meshcheriakova I.A., Bannikova S.V., Demidov E.A., Bryanskaya A.V., Scheglov M.A., Semenov A.I., Oshchepkov D.Yu., Popik V.M., Peltek S.E. The impact of terahertz radiation on an extremophilic archaean *Halorubrum saccharovororum* proteome. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):869-875. DOI 10.18699/VJ16.205

ORIGINAL ARTICLE

Received 03.11.2016

Accepted for publication 09.12.2016

© AUTHORS, 2016

✉ e-mail: tanago@bionet.nsc.ru

The impact of terahertz radiation on an extremophilic archaean *Halorubrum saccharovororum* proteome

T.N. Goryachkovskaya¹✉, S.H. Konstantinova¹, I.A. Meshcheriakova¹, S.V. Bannikova¹, E.A. Demidov¹, A.V. Bryanskaya¹, M.A. Scheglov², A.I. Semenov², D.Yu. Oshchepkov¹, V.M. Popik², S.E. Peltek¹

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Budker Institute of Nuclear Physics SB RAS, Novosibirsk, Russia

Nonthermal effects of terahertz radiation on living objects are currently intensely studied, as more sources of this radiation type and devices employing it are being constructed. Terahertz radiation is increasingly used in security and inspection systems, medical and scientific appliances due to its low quant energy, which does not cause severe effects on organisms as other radiation types with higher quant energies do. The aim of this study was the identification of protein complexes participating in the response of the archaea *Halorubrum saccharovororum* H3 isolated from an extreme natural environment to terahertz radiation. We developed a microfluidic system for irradiation of bacterial and archaeal cultures with terahertz radiation and performed a 5-hour-long exposure of *H. saccharovororum* to terahertz radiation at a wavelength of 130 μm and a power density of 0.8 Wt per cm² for 5 h. We identified under- or overexpressed proteins in response to terahertz radiation using 2D electrophoresis with subsequent MALDI-TOF mass spectrometry. A total of 16 differentially expressed protein fractions with at least 1.5-fold changes in expression level were detected. The obtained data suggest that *Halorubrum* cells respond to exposure to terahertz radiation by expression changes in gene products involved in translation regulation.

Key words: free electron laser; terahertz radiation; microfluidic system; halophilic thermotolerant archaea.

В последние десятилетия в связи с появлением новых источников терагерцового излучения и разработкой приборов на его основе появилась необходимость исследовать воздействие излучения этого диапазона на живые объекты (Wilmink, Grundt, 2011; Weightman, 2012).

Использование терагерцового (ТГц) излучения в системах безопасности контроля и досмотра, диагностическом медицинском оборудовании и научных исследованиях обусловлено малой энергией кванта этого излучения, не вызывающей негативных последствий при контакте с живыми системами, как это происходит в случае применения излучения других, более высокоэнергетичных диапазонов. При этом сложно ожидать наличия врожденных механизмов защиты от этого излучения, поскольку в атмосфере Земли оно практически полностью поглощается парами воды.

Применение лазеров на свободных электронах (ЛСЭ) открывает уникальные возможности для изучения воздействия ТГц излучения на живые объекты, поскольку позволяет проводить эксперименты в широком диапазоне длин волн и мощностей.

Существует множество гипотез о воздействии электромагнитного излучения на живые системы разного уровня организации (Fröhlich, 1982; Бецкий, 1994; Гапеев, Чемерис, 2000; Бецкий, Лебедева, 2001), но целостного представления, объясняющего многообразные эффекты воздействия излучения с единых позиций, до сих пор нет. Это относится и к терагерцовому диапазону. К настоящему времени большинством исследований не обнаружено влияния ТГц излучения на первичную структуру ДНК. Проведенное нами исследование на уровне генома микроорганизмов также не выявило генотоксических эффектов воздействия ТГц излучения. Полученные данные свидетельствуют об отсутствии мутагенного эффекта ТГц излучения (Sergeeva et al., 2016).

Нами также изучено нетермическое воздействие ТГц излучения на митотический индекс и экспрессию митохондриальных генов стволовых клеток человека (hESCs). Показано, что ТГц излучение не вызывает двунитевых разрывов в хромосомах стволовых клеток человека, однако повышает транскрипцию генов митохондрий (Bogomazova et al., 2015).

При изучении нетермического воздействия ТГц излучения на культуры клеток микроорганизмов нам удалось с помощью геносенсорных конструкций показать, что в клетках *Escherichia coli* индуцируется стрессочувствительный промотор гена каталазы. Известно, что он маркирует сеть ответа клеток *E. coli* на окислительный стресс (Demidova et al., 2013). Также были исследованы геносенсоры *E. coli*/pCOP-A-GFP и *E. coli*/pEmrR-GFP на основе промоторов генов *copA* и *EmrR*. Согласно нашим данным, одними из первых на нетермическое воздействие ТГц излучения реагируют системы контроля окислительного стресса и поддержания гомеостаза переходных металлов, а система контроля устойчивости клеток *E. coli* к присутствию антибиотиков не реагирует на излучение. Индукция ответа геносенсоров носит ярко выраженный пороговый характер по поглощенной дозе и слабо зависит от длины волны терагерцового излучения (Demidova et al., 2016).

В задачу настоящей работы входила идентификация комплекса белков, участвующих в ответе клеток архей *Halorubrum saccharovorum* H3, выделенной из природной экстремальной среды обитания, на электромагнитное излучение терагерцового диапазона. Культуры экстремофильных архей интересны тем, что имеют иные метаболические характеристики в сравнении с бактериями *E. coli* и обладают большей устойчивостью к внешним стрессам. Методами протеомного анализа, включающими двумерный электрофорез с последующей MALDI-TOF масс-спектрометрией, проведена идентификация белков, изменяющих уровень экспрессии в результате нетермического воздействия терагерцового излучения.

Материалы и методы

Источник излучения. Эксперименты по изучению воздействия терагерцового излучения на клетки архей *H. saccharovorum* H3 проведены с использованием уникальной установки «Новосибирский лазер на свободных электронах» Института ядерной физики (ИЯФ) СО РАН. В отличие от других источников ТГц излучения, применение ЛСЭ для исследования биологических эффектов излучения имеет ряд преимуществ. Прежде всего, это возможность плавно перестраивать длину волны излучения, что осуществляется изменением энергии электронов и параметров ондулятора. Другим преимуществом является возможность получать очень высокие средние мощности излучения благодаря тому, что в качестве рабочей среды используется электронный пучок (Kulipanov et al., 2008, 2015).

Культура архей. В эксперименте использована культура экстремофильных архей вида *Halorubrum saccharovorum* (штамм H3, коллекция лаборатории молекулярных биотехнологий Института цитологии и генетики (ИЦиГ) СО РАН), выделенная из природного образца оз. Бурлинское (Алтайский край), для которого установлено высокое содержание солей в природном растворе (Леонова и др., 2007).

Была отработана методика отбора и культивирования архей. Чистую культуру, полученную из коллекции, высевали на твердую среду S4 следующего состава (г/л): NaCl – 200, MgCl₂ – 5, KCl – 1, CaCl₂ – 1, триптон – 4, дрожжевой экстракт – 2, агар – 15; pH – 7,5. Предварительно чистую культуру разводили более чем в 1 000 раз, что обеспечивало рост колоний из единичных клеток. Эти колонии служили исходным материалом для получения культур и осуществления их периодического культивирования в жидкой среде S4 состава (г/л): NaCl – 200, MgCl₂ – 5, KCl – 1, CaCl₂ – 1, триптон – 4, дрожжевой экстракт – 2; pH – 7,5.

Концентрацию клеток в периодических культурах определяли спектрофотометрически (Epoch Microplate Spectrophotometer, BioTek) по оптической плотности. Исследование показало, что экспоненциальная фаза роста культуры при 37 °C длится не более 48 ч – с 24 по 72-й час культивирования.

Из полученных колоний были отобраны две, имеющие наибольшую скорость роста при периодическом культивировании на жидкой среде S4. Эти колонии послужили исходным посевным материалом для двух биологических

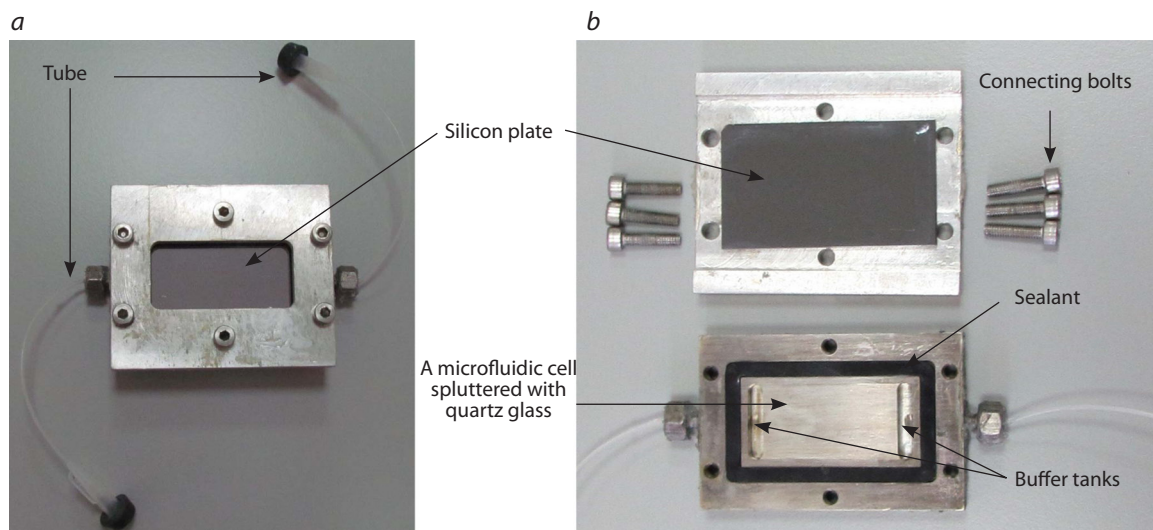


Fig. 1. (a) A microfluidic cuvette for irradiation of liquid cultures; (b) cuvette components.

повторов в дальнейших экспериментах. Исходя из характера кривой роста культур, была выбрана точка отбора образцов – на 60 ч культивирования – для облучения терагерцовым излучением.

Микрофлюидная кювета. Принимая во внимание, что глубина поглощения терагерцового излучения водными растворами составляет несколько десятков микрон, для облучения больших объемов культуры *H. saccharovororum* была сконструирована и изготовлена микрофлюидная проточная кювета с крышкой из высокоомного кремния (с малым поглощением терагерцового излучения), обеспечивающая ламинарный поток слоя жидкости толщиной 40 мкм перпендикулярно потоку излучения.

Материалом для тела кюветы послужил инвар (сплав на основе железа и никеля марки 36Н) с напылением тонкого слоя (~1 мкм) инертного материала – кварцевого стекла, для исключения влияния материала кюветы на состав культуральной среды.

Линейный коэффициент температурного расширения кварцевого стекла близок к таковому для инвара и составляет $1 \times 10^{-6} \text{ K}^{-1}$, что обеспечивает равномерный контакт этих материалов в широком диапазоне температур.

Крышка кюветы, через которую производилось облучение, представляла собой пластину высокоомного кремния толщиной 2 мм, закрепленную в рамке-держателе также из инвара (рис. 1). Проток жидкости через микрофлюидную систему осуществляли при помощи перистальтического насоса (ЛКВ, Швеция), который соединяли с микрофлюидной кюветой полипропиленовыми трубками. Ламинарное течение культуральной жидкости через кювету обеспечивалось скоростью потока и специально сделанными в теле кюветы (на входе и выходе) буферными емкостями.

Облучение клеток *H. saccharovororum*. Эксперименты проводили на биологической станции Сибирского центра синхротронного и терагерцового излучения ИЯФ СО РАН, куда по специальному каналу подается излучение Новосибирского ЛСЭ.

Параметры излучения Новосибирского ЛСЭ приведены ниже:

Длина волны	0.05–0.24 мм
Длительность импульса	50 пс
Частота следования импульсов	2.8–11.2 МГц
Средняя мощность	до 400 Вт
Пиковая мощность	до 1 МВт
Минимальная относительная ширина линии	$3 \cdot 10^{-3}$

Клетки облучали терагерцовым излучением с длиной волны 130 мкм в течение 5 ч при плотности мощности излучения 0.8 Вт/см^2 в микрофлюидной проточной кювете, перпендикулярной потоку излучения (рис. 2). Температу-

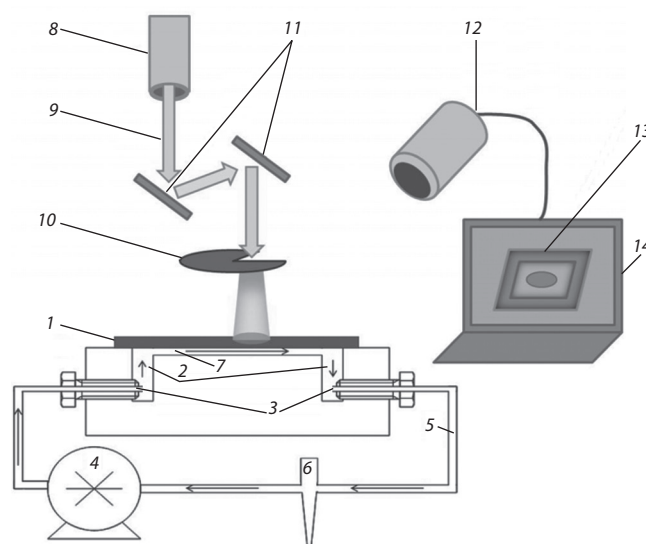


Fig. 2. Experimental facility: 1, cuvette lid (a quartz plate); 2, buffer tanks; 3, tube connection points; 4, peristaltic pump; 5, pump tubes; 6, aeration tube; 7, chamber for cell culture; 8, the Novosibirsk FEL; 9, THz radiation; 10, shutter; 11, mirrors; 12, thermal camera; 13, thermal image of the cuvette; 14, computer.

ру 42 ± 2 °C на поверхности кюветы поддерживали путем регулировки средней мощности излучения обтюратором, вращающимся при помощи электромотора. Обтюратор представлял собой два медных круга, имеющих общую ось вращения, в каждом из которых было сделано секторальное отверстие. Поворот кругов друг относительно друга регулировал площадь открытого отверстия и, таким образом, среднюю мощность излучения при облучении образцов. Контроль температуры осуществляли с помощью тепловизора ТКВр-СВИТ101 (Kuryshv et al., 1998). Контрольные клетки циркулировали в такой же кювете в течение 5 ч в термостате при 42 °C. Для проведения протеомного анализа облученные и контрольные клетки собирали центрифугированием и лизировали добавлением лизирующего буфера.

Выделение общего белка при высоком содержании солей в среде. Белки галофильных архей относятся к экстремально кислым (средняя рI 4.5) и имеют высокий отрицательный поверхностный заряд, что является адаптивным приспособлением к существованию в условиях осмотического стресса. Это делает их непростым объектом для протеомных исследований. Для проведения двумерного электрофореза подобных белков необходимо избавиться от солей и преодолеть сильный катодный дрейф, приводящий к искажению картины двумерного электрофореза, особенно для белков с высоким молекулярным весом.

Для выделения общего белка была использована методика C.W. Cho et al. (2003) с модификациями. Клетки осаждали центрифугированием при 5900 g в течение 15 мин. Полученный осадок ресуспендировали в десятикратном объеме лизирующего буфера, содержащего 10 mM ЭДТА, 20 mM трис-HCl pH 8.0, 1 mM PMSF, 0.5 mM ДТТ, 0.2 % CHAPS¹. После этого клетки дезинтегрировали на ультразвуковом дезинтеграторе Ultrasonic processor CPX 130 (Cole Parmer, США) в течение 10 мин (режим – 2 с импульс, 8 с выключен) на льду. К полученному гомогенату добавляли ДНКазу в концентрации 100 мкг/мл и инкубировали на льду в течение часа. Гомогенат осветляли центрифугированием при 4 °C в течение 20 мин на 15 700 g. Для удаления солей и низкомолекулярных примесей супернатант пропускали через хроматографическую колонку MicroBioSpin6 (BioRad, США). Дополнительно белок очищали с помощью ReadyPrep 2-D Cleanup Kit (BioRad, США). Осадок растворяли в буфере для изоэлектрофокусирования (7 M мочевины, 2 M тиомочевины, 4 % CHAPS, 65 mM ДТТ, 0.4 % амфолит pH 3-10, 1 mM PMSF). Концентрацию белка определяли с помощью Quick Start Bradford Dye Reagent (BioRad, США).

Проведение двумерного электрофореза. Для проведения двумерного электрофореза выделенный белок наносили на 17 см стрип, pH 3–6 (ReadyStrip™ IPG Strip, BioRad, США). Для аналитических гелей брали по 100 мкг белка, для препаративного геля – 300 мкг. Стрипы помещали в камеру для изоэлектрофокусирования Protean IEF Cell (BioRad, США). Регидратировали при 50 мА на стрип при 20 °C в течение 12 ч. Изоэлектрофокусирование

проводили при 50 мА на стрип при 20 °C в следующем режиме: линейное повышение напряжения до 250 В в течение 20 мин, затем до 5700 В в течение 2.5 ч, и далее выдерживали при 5700 В до достижения 40000 вольт-часов. Стрипы выдерживали 15 мин в уравнивающем буфере (6 M мочевины, 2%-й ДСН, 20%-й глицерин, 0.375 M трис-HCl pH 8.8, 130 mM ДТТ), затем 15 мин в уравнивающем буфере (6 M мочевины, 2%-й ДСН, 20%-й глицерин, 0.375 M трис-HCl pH 8.8, 200 mM иод-ацетамид), промывали в ДСН трис-глициновом буфере (pH 8.3) и накладывали на 12%-й ПААГ-ДСН. Разделение белка во втором направлении проводили в 12 % ПААГ-ДСН при 16 мА/гель в течение 30 мин, затем при 24 мА/гель в течение 5 ч в камере PROTEAN II xi multi-cell (BioRad, США). Аналитические гели окрашивали флуоресцентным красителем SYPRO Ruby Protein Gel Stain (BioRad, США), препаративный гель – флуоресцентным красителем Flamingo (BioRad, США). Гели фотографировали при помощи геледокументирующей системы Versadoc MP4000 (BioRad, США). Данные анализировали, используя PDQuest Advanced 2-D Analysis Software (BioRad, США).

Фракции белка с достоверно меняющейся концентрацией (критерий Стьюдента, $p < 0.01$) вырезали из препаративного геля при помощи прибора EXQuest Spot Cutter (BioRad, США) и промывали 0.2 M бикарбонатом аммония в 50 % ацетонитриле при 37 °C 10 мин. После высушивания с помощью ацетонитрила проводили гидролиз белка, регидратируя гели раствором модифицированного свиного трипсина (Trypsin Gold, Mass Spectrometry Grade, Promega, США) с концентрацией 0.02 mM/мл в 40 mM бикарбонате аммония в 10 % ацетонитриле. Регидратированные кусочки геля инкубировали при 37 °C в течение 16 ч.

Идентификация белков. Для экстракции пептидов в каждую пробирку с отдельными фракциями белка добавляли по 20 мкл 5 % ТФУ (Sigma) в воде. Инкубировали 45 мин на шейкере при 36 °C. Выделение и очистку триптических пептидов проводили при помощи Millipore ZIPTIP C18 (Millipore, США). Очищенные пептиды смешивали с насыщенным раствором матрицы – α -циано-4-гидроксикоричной кислоты (Bruker, Германия) в 70 % AcN (Merck) и 0.1 % ТФУ (Sigma). Аликвоту (1 мкл) полученной смеси наносили на металлическую мишень масс-спектрометра и высушивали на воздухе. Измерения выполняли на приборе Ultraflex III (Bruker, Германия) со следующими параметрами: ускоряющее напряжение 25 кВ, напряжение рефлектрона 26.3 кВ, диапазон масс 500–3500 Да, задержка экстракции 200 нс. Калибровку прибора производили с использованием стандартного триптического гидролизата бычьего сывороточного альбумина (Bruker, Германия) по следующим массам (Да): 937.48, 1163.63, 1283.71, 1305.71, 1399.69, 1439.81, 1479.79, 1567.74, 1639.93, 1724.84, 1880.92, 1907.92, 2045.02.

Белки были идентифицированы с использованием поискового алгоритма Mascot (<http://matrixscience.com/home.html>) и базы данных SwissProt. Параметры поиска: ошибка масс ± 0.05 Да; число сайтов недогидролиза трипсином – 1; возможные модификации – окисление метионина и карбамидометилирование цистеинов. Для корректной идентификации белков и определения их функциональной

¹ Здесь и далее: ДСН – додецилсульфат натрия; ДТТ – дитиотреитол; CHAPS – 3-[(3-холомидопропил)-диметиламмоний]-1-пропансульфонат; ПААГ – полиакриламидный гель; ЭДТА – этилендиаминтетраацетат; PMSF – фенилметилсульфонилфторид.

роли в геноме были использованы результаты проведенного нами ранее полногеномного секвенирования и аннотации генома *H. saccharovorum* H3 (Rozanov et al., 2015).

Биоинформационный анализ данных. Для проведения анализа обогащения по терминам Gene Ontology (GO) нами осуществлен поиск наиболее близких гомологов *E. coli* K-12 для выявленных белков *H. saccharovorum* H3. Поиск проводили путем выравнивания полученных белковых последовательностей с белковыми последовательностями *E. coli* K-12 инструментом BLASTP 2.3.0+, реализованным в базе данных EcoCyc (<http://www.ecocyc.org/>).

Анализ терминов GO проводили инструментом PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System (Mi et al., 2016) (<http://pantherdb.org/>).

Результаты и обсуждение

Проведено две серии экспериментов с тремя техническими повторами. Результаты одного из разделений белка культуры клеток *H. saccharovorum* H3 представлены на рис. 3. При сравнении электрофореграмм белков, выделенных из облученных и контрольных культур, было обнаружено 16 белковых фракций протеома архей с достоверными различиями концентрации более чем в полтора раза (см. рис. 3). В результате нетермического воздействия ТГц излучения концентрация белков в основном снижалась – 13 фракций. Из 13 фракций, изменяющих свою экспрессию под действием ТГц излучения, методами масс-спектрометрии были идентифицированы семь фракций (табл. 1). Увеличение концентрации установлено только в трех фракциях (№ 4, 5, 10), одна из которых была идентифицирована как гликозилтрансфераза (см. табл. 1, № 5). Клеточная стенка архей сложена белковыми глобулами, образующими так называемый S-слой. Большая часть белков S-слоя имеет гликозилированную форму. Гликозилтрансфераза осуществляет сборку N-связанных пентасахаридов, которые покрывают гликопротеины S-слоя (Schäffer, Messner, 2004). Увеличение концентрации гликозилтрансферазы в протеоме бактерии может быть вызвано возбуждением за

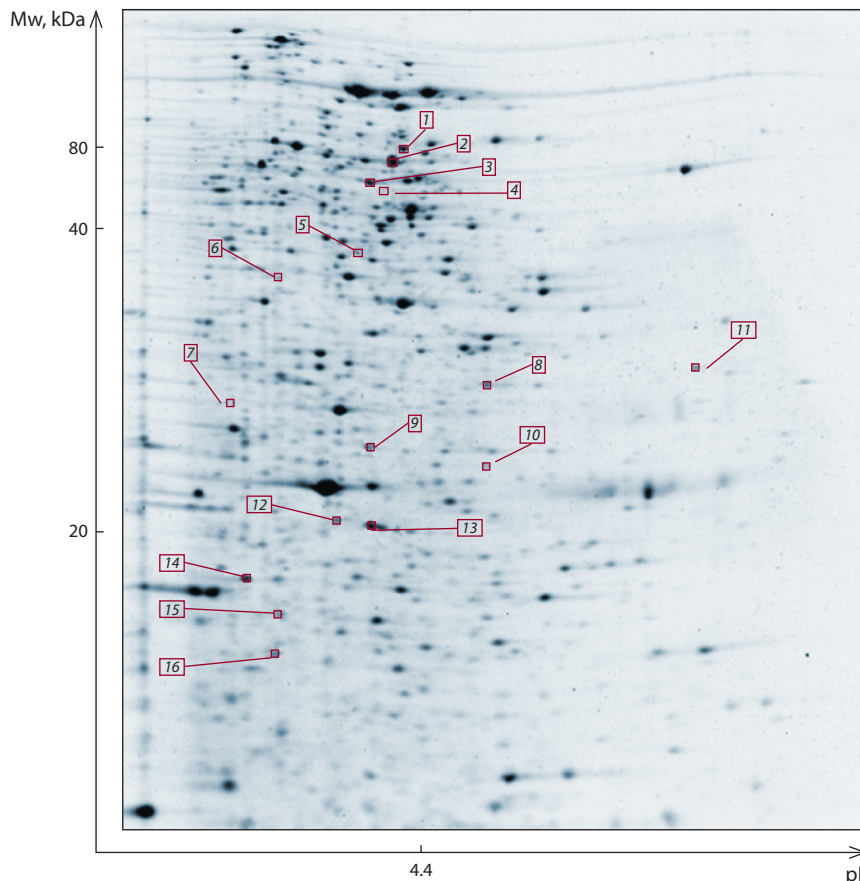


Fig. 3. 2D SDS-PAGE of the total protein fraction of *H. saccharovorum* H3 stained with Flamingo.

Enumerated (1–16) arrows point to fractions with at least 1.5-fold increase or decrease in response to the irradiation.

счет ТГц излучения водородных связей в клеточной мембране *H. saccharovorum* H3, которое приводит к нарушению структуры S-слоя (Abu-Qarn et al., 2008; Kandiba et al., 2013).

При помощи базы данных UniProt Knowledgebase (UniProtKB, <http://www.uniprot.org/>) для идентифицированных белков были установлены молекулярная функция и биологический процесс, в котором они участвуют (см. табл. 1). Определение проводили по термину «название белка» и термину «*Halorubrum*» в поле «организм».

Эксперименты проводились в условиях строгого контроля температуры опытных и контрольных образцов, поэтому наблюдаемые изменения в протеомном профиле вызваны нетермическим воздействием ТГц излучения. Как известно, тепловой шок индуцирует продукцию нескольких белков теплового шока. О нетермическом характере воздействия ТГц излучения свидетельствует и отсутствие белков теплового шока среди идентифицированных в эксперименте белковых фракций.

Результаты поиска наиболее близких гомологов *E. coli* K-12 для выявленных белков *H. saccharovorum* H3 представлены в табл. 2. Здесь же приведены Score и E-value выравнивания белковых последовательностей. Поиск выполнялся путем выравнивания последовательностей пептидов, полученных в результате масс-спектрометрической идентификации, с белковыми последовательностями *E. coli* K-12, представленными в базе данных EcoCyc (<http://www.ecocyc.org/>).

По выявленным гомологам *E. coli* K-12 был проведен анализ обогащения, согласно терминам базы данных Gene Ontology (PANTHER Protein Class). При этом значимое обогащение наблюдалось по факторам элонгации трансляции

Table 1. Identified *H. saccharovorum* H3 protein bands

No.	Name	Mw/pi	Score	Sequence coverage, %	Pathway	Function	Δ
1	Translation_elongation_factor_2	41473/4.67	137	41	Protein biosynthesis	Translation elongation factor	–
2	Translation_elongation_factor_1α subunit	47629/4.36	181	53			–
3	Translation_elongation_factor_1α subunit	47629/4.36	140	46			–
5	Glycosyltransferase	37455/4.33	113	41	Biosynthesis	Transferase	+
6	Eukaryotic_translation_initiation_factor_2α subunit	26724/4.15	54	26	Protein biosynthesis	Translation initiation factor	–
12	Thiosulfate_sulfurtransferase 2 C_rhodanese	34809/4.34	75	38	Metabolic processes	Transferase	–
13	Adenine phosphoribosyl transferase	20841/4.33	78	49	Nucleotide metabolism	Adenine phosphoribosyl transferase	–

Abbreviations: –, downregulation in response to THz; +, upregulation. Fractions 2 and 3 are identified as a single protein.

Table 2. Search for *E. coli* K-12 genes expressing proteins homologous to *H. saccharovorum* H3 whose concentrations are altered by THz irradiation

<i>H. saccharovorum</i> H3 protein homolog	<i>H. saccharovorum</i> protein sequence	Homologous <i>E. coli</i> K-12 gene	Score	E-value
Translation_elongation_factor_2	1 MGRKKIVQE CERLMDSPEN IRNIAIAHV DHGKTTLSDN LLAGAGMISQ 51 DTAGEQLAMD TKEDEQERGI TIDAANVSMT HEYEDTNHLI NLIDTPGHVD 101 FGGDVTRAMR AVDGALVVVD AVEGAMPQTE TVLRQALREG VKPALFINKV 151 DRLISELQEG PQEMQERLMS VIADVNELIR GMAENMDDIP EDWTVSVEDG 201 TVGFGSALYK WGVSMPSMQR TGMDFADIME LEQNDERQEL HERTPLSDVV 251 LDMVCEHFPN PVDAQPRRVP RIWRGDAESE LADTMRMVNE DGEVVMVMTD 301 ISMDPHAGEI ATGRVFSGTL EKGQELYVSG TAGKTASRAS GSSWGSPARK 351 WTASRRGTSR PSPACVTPSP APPSPPPWR	bipA	117	2e–29
Translation_elongation_factor_1α subunit	1 MESTETYSQS ALARDYPMSD KPHQNLAIG HVDHKGKSTLV GRLLFETGSV 51 PEHVIEQHRE EAEKGGKGGF EFAYVMDNLA EERERGVTID IAHQEFDTDN 101 YYFTIVDCPG HRDFVKNNMIT GASQADNAVL VVAADDGVAP QTREHVFLLAR 151 TLGINELIIG VNKMDLVDYS EDSYKQVIDE VEDLLNQVRF ATEDTSFVPI 201 SAFEQDNVAE ASENTWPYDG DTLLESLNDL PESEPPTDAP LRLPIQDVYT 251 ISGIGTVPGV RVETGILNTG DNVSFQPSDV GGEVKTVMH HEEVPKAEFG 301 DNVGFNVRGI GKDDIRRGDV CGPADDPSPV AETFQAQVVV MQHPSVITAG 351 YTPVFHAHTA QVACTIESID QKIDPSSGEV AEENPDFIKS GDAAVTVTRP 401 QKPLSIEPSG EIPELGSFAI RDMGQTIAAG KVLEVNER	tufB	186	1e–54
Glycosyltransferase (EC_2.4.1.-)	1 MLGWGYPPNI TGGLDVHVGE LFSGLRDDFG VDATLVLPAE FAPDEPGIE 51 PVETGAGDVA ARVDRLSDRF AELAPDHDII HTHDWFGYGP GRKAARASDA 101 TWVSSFHSLA SDRNIDPPTR EVETERRLAN AADTNIAVSE LVREDIRELY 151 DADSRVYNG FSTPTFSGKD VREDLGIDGE MLFFVGRHTD QKGISHLLYA 201 MRKLHRPGVT LVVGGSGHQT DQLKRFAELL GIEDRVEFVG YVPEAELGDY 251 YAAADAFVSP SYAEPFGITI TEALESQTQV VATRSGVAEV LPDDCLVEVE 301 VDSEIVDGI TTALDREPP EYERREWSDV TEDTLAVYED VA	wcaL	46.2	2e–06
Eukaryotic_translation_initiation_factor_2α subunit	1 MFVDLDEYED KRGLCHISEV ASGWIKNVRD HVREGQTVVA KVLVDVESSN 51 QIDLSIKDVN EHQRKDKIQD WKNSQKADNW MLIALGEDVD DDRYTEVANA 101 LLVEYDSLYD AFESAAISGE EALDVDIDD DALDAVVEAA RDNVSVPYVD 151 VTGYVDLESM GPDGVDDVKA ALEAAEGNGE VPDGVDLEVG YVGSPEYRIK 201 VRAPDYKTAE DQLEAAAAARA REAIEAAGGV GEFHRERRED DE	pnp	43.1	1e–05
Adenine phosphoribosyl transferase (EC_2.4.2.7)	1 MDQLRQSLLD APIIEKGEYQ YFVHPISDGV PMLKPELLRE IVIRIRKAE 51 LENVDKIVTP AAMGIHISTA LSLMTDIPLV VIRKROYGLD GEVPLFQETG 101 YSESEMYIND VEEGDRVLVL DDVLSTGGTM KAILDALTNE VGAEVVDVVA 151 VIKKAGENEL DQTDYNVKTL INVTVEDGEV VIVDAKGDD	apt	41.2	1e–05
Rhodanese (EC_2.8.1.1)	1 MPEDSLVTAD WVEENLEKFE SDDPAYRLVE INNPTVTDES DYPYEEGHI 51 PGALNFEWDE VFTDKTERDI VSKEKFAERN GEGGIEADTT VVYGGGRVP 101 NWFALFGYWI YKYYGHDDVR VIDGGKGYWV ENDYPLSTEE PDFTPREYEA 151 RGPFESVRAY KDDIDNAIED GLPMVDVRSP EEFSGEVIAP EGLMETAQRG 201 GHIPGASSVP IGTTLNEDGT FKSREELREL YADAGVDGNE STITYCRVGE 251 RSSIEWFLH ELLGYDDVRN YDGSWTEWGN LVGARSRRAS KEVRGRDRTD 301 VFLNSRC	sseA	103	2e–26

(более чем в 86.98 раза, p -value 2.26E-02) и по ГТФ-связывающим белкам (более чем в 64.09 раза, p -value 4.15E-02).

Клетки микроорганизмов осуществляют постоянный мониторинг физико-химических условий среды, в которой они находятся, с тем чтобы координировать экспрессию генов в соответствии с внешними условиями. Эта регуляция происходит на разных уровнях – на уровне инициации транскрипции, на уровне синтеза и деградации мРНК, на уровне синтеза белка. Выживание клетки в экстремальных условиях требует адекватной регуляции экспрессии генов, включая контроль трансляции. Анализ протеома показывает, что четыре из шести генов *Halorubrum*, которые снижают экспрессию под воздействием ТГц излучения, связаны с контролем процесса трансляции.

Ранее нами было изучено влияние нетермического воздействия ТГц излучения на стресс-реактивные системы клеток *E. coli*. Было показано, что на ТГц излучение реагируют в первую очередь и сильнее всего гены, индуцируемые в ответ на окислительный стресс, слабее – гены, связанные с поддержанием гомеостаза меди, и совсем не реагируют гены, контролирующие устойчивость к антибиотикам. Таким образом, ТГц излучение селективно воздействует на разные стрессовые системы *E. coli* и существенно влияет на жизненно важный гомеостаз окислительных процессов.

Сравнение результатов анализа протеомного ответа на нетермическое воздействие терагерцового излучения показало, что для клеток *E. coli* наблюдается ответ систем, связанных прежде всего с окислительным стрессом, в то время как реакция клеток *H. saccharovororum* обусловлена регуляцией синтеза белка и мембранными процессами.

Бактерия *E. coli* – облигатный анаэроб, обитающий в кишечнике млекопитающих, а архея *H. saccharovororum* – свободно живущий экстремофил. Возможно, столь разные реакции этих микроорганизмов на нетермическое воздействие ТГц излучения объясняются различием занимаемых ими экологических ниш и способов адаптации к ним.

Acknowledgments

This work was supported by the Siberian Branch of the Russian Academy of Sciences, Integrated Program II.2P, project 0324-2015-0025. Irradiation experiments were performed on a unique device, the Novosibirsk free-electron laser, at the Siberian Synchrotron and Terahertz Radiation Center (Budker Institute of Nuclear Physics, Novosibirsk).

Conflict of interest

The authors declare no conflict of interest.

References

Abu-Qarn M., Eichler J., Sharon N. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* 2008;18:544-550.
Betskiy O.V. On the mechanisms of the interaction between low-intensity millimeter waves and biologic objects. *Izvestiya vysshikh uchebnykh zavedeniy. Radiofizika* = News of Institutes of Higher Education. *Radiophysics*. 1994;37:30-41. (in Russian)
Betskiy O.V., Lebedeva N.N. Present-day notions of the action of low-intensity millimeter waves on biologic objects. *Millimetrovye volny v biologii i meditsine* = Millimeter waves in biology and medicine. 2001;24:5-19. (in Russian)

Bogomazova A.N., Vassina E.M., Goryachkovskaya T.N., Popik V.M., Sokolov A.S., Kolchanov N.A., Lagarkova M.A., Kiselev S.L., Peltek S.E. No DNA damage response and negligible genome-wide transcriptional changes in human embryonic stem cells exposed to terahertz radiation. *Sci. Rep.* 2015;5:7749.
Cho C.-W., Lee S.-H., Choi J., Park S.-J., Ha D.-J., Kim H.-J., Kim C.-W. Improvement of the two-dimensional gel electrophoresis analysis for the proteome study of *Halobacterium salinarum*. *Proteomics*. 2003;3:2325-2329.
Demidova E.V., Goryachkovskaya T.N., Malup T.K., Bannikova S.V., Semenov A.I., Vinokurov N.A., Kolchanov N.A., Popik V.M., Peltek S.E. Studying the non-thermal effects of terahertz radiation on *E. coli*/pKatG-GFP biosensor cells. *Bioelectromagnetics*. 2013;34:15-21.
Demidova E.V., Goryachkovskaya T.N., Mescheryakova I.A., Malup T.K., Semenov A.I., Vinokurov N.A., Kolchanov N.A., Popik V.M., Peltek S.E. Impact of terahertz radiation on stress-sensitive genes of *E. coli* cell. *IEEE Transact. Terahertz Sci. Technol.* 2016;6(3):435-441.
Fröhlich H. What are non-thermal electric biological effects? *Bioelectromagnetics*. 1982;3(1):45-46.
Gapeev A.B., Chemeris N.K. The model-based approach to the analysis of the action of modulated electromagnetic radiation on animal cells. *Biofizika* = Biophysics. 2000;45:299-312. (in Russian)
Mi H., Poudel S., Muruganujan A., Casagrande J.T., Thomas P.D. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucl. Acids Res.* 2016;44:D336-D342. DOI 10.1093/nar/gkv1194.
Kandiba L., Aitio O., Helin J., Guan Z., Permi P., Bamford D.H., Eichler J., Roine E. Diversity in prokaryotic glycosylation: an archaeal-derived N-linked glycan contains legionaminic acid. *Mol. Microbiol.* 2013;84:578-593.
Kulipanov G.N., Bagryanskaya E.G., Chesnokov E.N., Choporova Yu.Yu., Gerasimov V.V., Getmanov Ya.V., Kiselev S.L., Knyazev B.A., Kubarev V.V., Peltek S.E., Popik V.M., Salikova T.V., Scheglov M.A., Seredniakov S.S., Shevchenko O.A., Skrinisky A.N., Veber S.L., Vinokurov N.A. Novosibirsk Free Electron Laser – facility description and recent experiments. *IEEE Transact. Terahertz Sci. Technol.* 2015;5:798-809.
Kulipanov G.N., Gavrilov N.G., Knyazev B.A., Kolobanov E.I., Kotenkov V.V., Kubarev V.V., Matveenko A.N., Medvedev L.E., Miginsky S.V., Mironenko L.A., Ovchar V.K., Popik V.M., Salikova T.V., Scheglov M.A., Seredniakov S.S., Shevchenko O.A., Skrinisky A.N., Tcheskidov V.G., Vinokurov N.A. Research highlights from the Novosibirsk 400 W average power THz FEL. *Terahertz Sci. Technol.* 2008;1(2):107-125.
Kuryshv G.L., Kovchantsev A.P., Vainer B.G. Medical thermal imager based on matrix photodetector 128 × 128 operating in the spectral range 2.8–3.05 μm. *Avtometriya*. 1998;4:5-12. (in Russian)
Leonova G.A., Bogush A.A., Bobrov V.A., Bychinskiy V.A., Trofimova L.B., Malikov Yu.I. Ecogeochemical assessment of salt lakes of Altai Krai. *Geografiya i prirodnye resursy* = Geography and Natural Resources. 2007;1:51-59. (in Russian)
Rozanov A.S., Bryanskaya A.V., Malup T.K., Kotenko A.V., Peltek S.E. Draft genome sequence of a *Halorubrum* H3 strain isolated from the Burlinskoye salt lake (Altai Krai, Russia). *Genome Announc.* 2015;3(3):e00566-15.
Schäffer C., Messner P. Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology*. 2004;14:31R-42R.
Sergeeva S., Demidova E., Sinitsyna O., Goryachkovskaya T., Bryanskaya A., Semenov A., Meshcheryakova I., Dianov G., Popik V., Peltek S. 2.3 THz radiation: Absence of genotoxicity/mutagenicity in *Escherichia coli* and *Salmonella typhimurium*. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 2016;803-804:34-38.
Weightman P. Prospects for the study of biological systems with high power sources of terahertz radiation. *Phys. Biol.* 2012;9:053001.
Wilmink G.J., Grundt J.E. Invited review article: Current state of research on biological effects of terahertz radiation. *J. Infrared Milli. Terahz. Waves*. 2011;32:1074-1122.

Стрессовые системы *Escherichia coli* и их роль в реакциях на воздействие терагерцового излучения

С.Е. Пельтек¹✉, Е.В. Демидова¹, В.М. Попик², Т.Н. Горячкова¹

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное бюджетное учреждение науки Институт ядерной физики им. Г.И. Будкера Сибирского отделения Российской академии наук, Новосибирск, Россия

Изложены результаты последних лет по изучению реакции стрессовых систем *Escherichia coli* в ответ на нетермическое воздействие терагерцового (ТГц) излучения. Наиболее простым и удобным объектом для изучения воздействия ТГц излучения на живые объекты является бактерия *E. coli*. Это связано с ее высокой изученностью как на молекулярно-генетическом, так и на метаболическом уровне, а также с возможностью создания на основе промоторов ее стресс-активируемых генов и репортерного белка GFP геносенсорных конструкций. Введение этих конструкций в клетки *E. coli* позволяет исследовать реакцию конкретной стрессовой системы бактерии на ТГц излучение по интенсивности синтеза белка GFP, легко определяемого флуориметрически. В работе представлены данные литературных источников и собственные результаты по нетермическому воздействию ТГц излучения на конкретные стрессовые системы *E. coli*. Обсуждаются экспериментальные данные, полученные с использованием геносенсоров *E. coli*/pKatG-GFP, *E. coli*/pCopA-GFP и *E. coli*/pEmrR-GFP, которые являются маркерами генных сетей *E. coli*, активирующихся в условиях окислительного стресса, при нарушении гомеостаза ионов меди и в присутствии антисептиков соответственно. Обзор проведенных исследований показал, что нетермическое воздействие ТГц излучения индуцирует в клетках *E. coli* генные сети окислительного стресса и поддержания гомеостаза меди, но не влияет на активность стрессовых систем защиты от антибиотиков, протонифоров и супероксид-анионов. Наличие динамических особенностей в развитии стрессорного ответа у геносенсоров *E. coli*/pKatG-GFP и *E. coli*/pCopA-GFP на ТГц излучение в сравнении с естественными индукторами позволяет предположить специфичность ответа систем окислительного стресса и поддержания гомеостаза меди в реакции адаптации клеток *E. coli* к ТГц излучению.

Ключевые слова: терагерцовое излучение; стрессовые системы *E. coli*; геносенсорные конструкции.

Escherichia coli stress response systems and their reaction to terahertz radiation

S.E. Peltek¹✉, E.V. Demidova¹, V.M. Popik², T.N. Goryachkovskaya¹

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Budker Institute of Nuclear Physics SB RAS, Novosibirsk, Russia

In this review, we summarize the latest data concerning the reactions of *Escherichia coli* to nonthermal terahertz radiation and the underlying molecular mechanisms. *E. coli* is the most simple and convenient model object for studying the effects of terahertz radiation: both its genetics and metabolism are well studied, and it is easily amenable to genetic engineering allowing one to create biosensors using promoters of genes activated by certain stress factors and the reporter GFP protein. Transformed *E. coli* cells containing biosensors can be used to visualize their reactions to terahertz radiation based on the intensity of GFP fluorescence. In this review, we present data on the response of certain *E. coli* stress response systems to terahertz radiation obtained by us, as well as by other authors. We discuss experimental results for *E. coli*/pKatG-GFP, *E. coli*/pCopA-GFP, and *E. coli*/pEmrR-GFP biosensors that are used to detect *E. coli* genetic networks responding to oxidative stress, copper ion homeostasis failures, and antiseptics, respectively. The obtained data indicate that exposure to nonthermal terahertz radiation induces *E. coli* gene networks of oxidative stress and copper ion homeostasis, but does not activate those responding to antibiotics, protonophores, or superoxide anions. The fact that *E. coli*/pKatG-GFP and *E. coli*/pCopA-GFP biosensors have different activation and reaction periods when exposed to terahertz radiation and natural inducers suggests that reactions of oxidative stress and copper ion homeostasis systems to terahertz radiation are specific.

Key words: terahertz radiation; *E. coli* stress response systems; biosensors.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Пельтек С.Е., Демидова Е.В., Попик В.М., Горячкова Т.Н. Стрессовые системы *Escherichia coli* и их роль в реакциях на воздействие терагерцового излучения. Вавиловский журнал генетики и селекции. 2016;20(6):876-886. DOI 10.18699/VJ16.206

HOW TO CITE THIS ARTICLE:

Peltek S.E., Demidova E.V., Popik V.M., Goryachkovskaya T.N. *Escherichia coli* stress response systems and their reaction to terahertz radiation. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):876-886. DOI 10.18699/VJ16.206

REVIEW

Received 16.11.2016

Accepted for publication 21.12.2016

© AUTHORS, 2016

В связи с тем, что терагерцовый (ТГц) диапазон частот электромагнитного спектра активно используется в научных и прикладных областях деятельности, в дальнейшем предполагается все более интенсивный контакт человека с этим диапазоном частот, особенно в инспекционных системах безопасности и диагностическом медицинском оборудовании. Главная идея эксплуатации ТГц диапазона частот заключается в его теоретически обоснованной безопасности для живых систем – низкой энергии кванта, не способной к ионизации молекул и, следовательно, не вызывающей отрицательных последствий при контакте живых систем с другими излучениями более высокоэнергетичных диапазонов электромагнитного спектра.

Появление таких источников терагерцового излучения, как Новосибирский лазер на свободных электронах (ЛСЭ), открывает уникальные возможности для изучения воздействия излучения ТГц диапазона на живые объекты. В частности, можно проводить эксперименты при строго контролируемых параметрах излучения (длина волны, средняя и пиковая мощности).

При исследовании механизмов воздействия ТГц излучения на живые объекты необходимо учитывать, что излучение этого диапазона будет поглощаться водородными связями воды, присутствие которой в разных формах присуще всем живым объектам. В ходе эволюции живые организмы не сталкивались с данным видом излучения (из-за сильного его поглощения парами воды в атмосфере) и, следовательно, могут не иметь специфичных защитных систем. Поэтому при выполнении экспериментов по изучению возможности воздействия ТГц излучения на живые объекты критически важно обеспечить строгий температурный контроль.

Наиболее простым и удобным объектом для исследования воздействия ТГц излучения на живые объекты является бактерия *Escherichia coli*, генетика, молекулярная биология и метаболизм которой изучены наиболее подробно. Возможно создание геносенсорных конструкций, которые будут сигнализировать о наличии или отсутствии реакции конкретной стрессовой системы на ТГц излучение у *E. coli* синтезом специального репортерного белка GFP, легко определяемого флуориметрически. Применение таких геносенсорных конструкций позволит не только выявить возможное влияние ТГц излучения на функционирование отдельных генетических систем у *E. coli*, но и подойти к изучению некоторых механизмов этого воздействия.

В основе геносенсора лежит способность клетки отвечать на конкретное воздействие активацией своих специфичных систем. Каскады генов активируются в клетке в присутствии агентов, повреждающих структуру и функции клетки (Zheng et al., 2001). Чувствительным звеном геносенсорной конструкции являются промоторы таких генов, связанные с репортерным геном, появление продукта которого свидетельствует об активации системы. В качестве репортерных генов используют гены, кодирующие флуоресцентные или люминесцентные белки, которые легко детектируются (Hakkila et al., 2002). В результате внешнего воздействия клетка активирует экспрессию этой генетической конструкции, что приводит к синтезу флуоресцентных белков.

Известно, что иногда один и тот же стрессирующий фактор способен индуцировать как специфичную генную сеть, характерную именно для этого фактора, так и сопряженные генные сети, характерные для других видов стресса, что помогает клетке справиться с последствиями воздействия, снизить или даже ликвидировать повреждающий эффект токсических агентов на ее структуру и функции.

Целью настоящей работы является попытка описать основные свойства и элементы стрессовых систем *E. coli*, задействованных в ответе на нетермическое воздействие ТГц излучения.

Определение стрессовых систем *E. coli*, задействованных в ответе на терагерцовое излучение

В основу геносенсорных конструкций для изучения влияния ТГц излучения на живые системы нами положены известные регуляторные элементы хорошо изученных стрессовых систем *E. coli*, связанные с основными видами внешних воздействий на жизнедеятельность бактерий (окислительный стресс, поддержание гомеостаза ионов меди, присутствие антисептиков).

Эксперименты проводили с использованием уникальной установки – Новосибирского лазера на свободных электронах терагерцового диапазона Сибирского центра синхротронного и терагерцового излучения, спроектированного и запущенного в действие Институтом ядерной физики СО РАН. Облучение биологических объектов осуществляли на специально оборудованной биологической станции, позволяющей регулировать точную дозу облучения и строго контролировать температуру образца (Demidova et al., 2013; Bogomazova et al., 2015).

В работе использованы созданные в Институте цитологии и генетики СО РАН геносенсоры на основе промоторов генов, чувствительных к окислительному стрессу – *E. coli*/pKatG-GFP (Khlebodarova et al., 2007), присутствию в среде ионов меди и серебра – *E. coli*/pCopA-GFP, и геносенсор на основе промотора гена *emrR E. coli* (*E. coli* multidrug resistance regulator), чувствительного к присутствию таких противомикробных агентов, как салициловая или налидиксовая кислоты, – *E. coli*/pEmrR-GFP (Demidova et al., 2016). В качестве репортерного белка использовался белок GFP_{aaV}, характеризующийся интенсивной флуоресценцией и временем полужизни 60 мин (Andersen et al., 1998).

При сравнении динамики развития ответа наблюдается различный характер реакции геносенсоров на естественные индукторы: наработка флуоресцентного белка геносенсором *E. coli*/pKatG-GFP на перекись водорода достигает 1.25 отн. ед. относительно отрицательного контроля примерно через час после начала эксперимента и затем держится на этом уровне с некоторыми колебаниями, тогда как ответ геносенсора *E. coli*/pCopA-GFP на ионы меди развивается быстро и постоянно и достигает уровня 2 отн. ед. за три часа; в ответ на присутствие салициловой кислоты флуоресценция геносенсора *E. coli*/pEmrR-GFP медленно возрастает первые 2.5 часа и затем стабилизируется на уровне примерно 1.5 отн. ед. Характер реакции геносенсоров на ТГц излучение также

Table 1. Results of genosensor induction

Genosensor	Hydrogen peroxide 2.5 mM	Heat shock, 42 °C	THz	Copper sulfate 25 µM	Salicylic acid 0.1 mM
<i>E. coli</i> /pKatG-GFP	Natural inducer	No induction	Induction greater than with the natural inducer	Induction greater than with the natural inducer	No induction
<i>E. coli</i> /pCopA-GFP	No induction	»	Induction lower than with the natural inducer	Natural inducer	»
<i>E. coli</i> /pEmrR-GFP	»	»	No induction	No induction	Natural inducer

различен: наибольшее значение флуоресценции отмечается у *E. coli*/pKatG-GFP, при этом первые два часа ответ развивается медленно, затем ускоряется и достигает 1.5 отн. ед. за три часа от начала эксперимента, тогда как ответ *E. coli*/pCopA-GFP на излучение развивается с постоянной скоростью, но медленнее, чем *E. coli*/pKatG-GFP, и достигает 1.5 отн. ед. более чем за три часа; *E. coli*/pEmrR-GFP вовсе не развивает флуоресцентного ответа на воздействие излучения. Уровень индукции геносенсора *E. coli*/pKatG-GFP ТГц излучением оказался выше уровня индукции естественным индуктором, а уровень индукции геносенсора *E. coli*/pCopA-GFP ТГц излучением – ниже уровня индукции естественным индуктором. Концентрация клеток во всех образцах (облученных, отрицательного и положительного контроля), отобранных для измерения флуоресценции, была одинакова. Разная динамика ответа геносенсоров на естественные индукторы и ТГц излучение свидетельствует о различных механизмах индукции стрессового ответа. Понятно, что ответ *E. coli* на стрессы окружающей среды всегда является комплексным. На первом этапе стресса включаются механизмы нейтрализации воздействия, далее включается система защиты от повреждений, в конце – возможно, система репарации.

Для выявления перекрестных механизмов индукции имеющихся геносенсоров была проведена обработка клеток геносенсоров неспецифическими индукторами, включая тепловой шок. Результаты тестирования приведены в табл. 1.

Реакция геносенсоров на ТГц излучение и отсутствие реакции на тепловой шок однозначно свидетельствуют о том, что нами выявлена именно нетермическая составляющая воздействия. Чтобы выявить участие систем защиты от окислительного стресса в реакции на ТГц излучение, был использован геносенсор *E. coli*/pKatG-GFP, чувствительный к перекиси водорода. Однако этот геносенсор оказался чувствительным и к другим индукторам. Активация ТГц излучением геносенсоров, не чувствительных к перекиси водорода, а также геносенсора *E. coli*/pKatG-GFP свидетельствует о других механизмах индукции. В частности, регуляция экспрессии гена *CopA* не связана с окислительным стрессом. Добавление в среду с геносенсором *E. coli*/pCopA-GFP перекиси водорода в концентрациях, вызывающих развитие окислительного стресса, не приводит к активации этого геносенсора.

В настоящее время определено более десятка генных сетей, которые активируются в ответ на различные виды стресса. Известно, что в ответ на один стрессовый фактор могут активироваться несколько генных сетей (Hengge-Aronis, 1999). Воздействие высоких доз ТГц излучения

является новым фактором, отсутствующим в природных условиях, следовательно, живые организмы не имеют сложенных и устоявшихся систем сигнализации и защиты при его воздействии. Тем не менее реакцию бактериальной клетки на ТГц излучение можно зарегистрировать и интерпретировать, опираясь на уже известные механизмы стрессовых ответов *E. coli*. Рассмотрим более подробно стрессовые системы *E. coli*, реагирующие на перечисленные в табл. 1 естественные индукторы (перекись водорода, тепловой шок, сульфат меди и салициловая кислота).

Механизмы регуляции стрессовых реакций *E. coli*

Метаболизм перекиси водорода (H₂O₂)

Окислительному стрессу подвержены все организмы с аэробным дыханием, поскольку активные формы кислорода (АФК) образуются вследствие нормального метаболизма. Большинство АФК, такие как пероксид водорода, супероксидный и гидроксильный радикалы, являются побочными продуктами в электронно-транспортной цепи в процессе окислительного фосфорилирования (González-Flecha, Demple, 1995; Lushchak, 2011).

В метаболических путях у *E. coli* встречается два пути расщепления перекиси водорода – с образованием кислорода и воды; окисление L-аскорбата до L-дегидроаскорбата (рис. 1).

Помимо участия в метаболических путях, перекись водорода участвует в реакциях окисления и реакциях Фентона, происходящих в клетках *E. coli* (табл. 2).

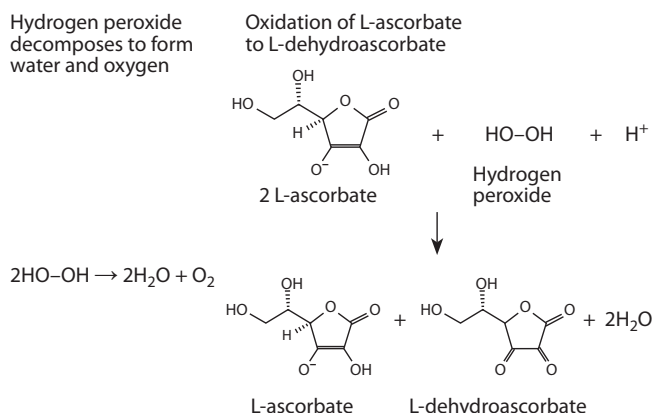


Fig. 1. Hydrogen peroxide decay pathways in *E. coli* (from <http://ecocyc.org>).

Table 2. Hydrogen peroxide reactions in *E. coli* cells (<http://ecocyc.org>)

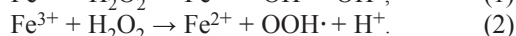
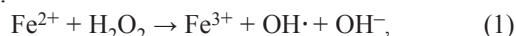
Reduction	$2\text{Fe}^{2+} + \text{H}_2\text{O}_2 + 2\text{H}_2\text{O} \rightarrow 2[\text{FeO}(\text{OH})] + 4\text{H}^+$
L-ascorbate oxidation	$2\text{L-ascorbate} + \text{H}_2\text{O}_2 + 2\text{H}^+ \rightarrow 2 \text{ monodehydroascorbate radical} + 2\text{H}_2\text{O}$
Fenton reaction	$\text{Fe}^{2+} + \text{H}_2\text{O}_2 \rightarrow \text{Fe}^{3+} + \text{OH}\cdot + \text{OH}^-$ $\text{Fe}^{3+} + \text{H}_2\text{O}_2 \rightarrow \text{Fe}^{2+} + \text{OOH}\cdot + \text{H}^+$
Oxidation	$\text{L-methionine} + \text{H}_2\text{O}_2 \rightarrow \text{L-methionine superoxide} + 2\text{H}_2\text{O}$ $\text{Reduced electron acceptor} + \text{H}_2\text{O}_2 \rightarrow \text{oxidized electron acceptor} + 2\text{H}_2\text{O}$ $2 \text{ glutathione} + \text{H}_2\text{O}_2 \rightarrow \text{glutathione disulfide} + 2\text{H}_2\text{O}$ $\text{Reduced thioredoxin} + \text{H}_2\text{O}_2 \rightarrow \text{oxidized thioredoxin} + 2\text{H}_2\text{O}$

АФК оказывают негативное действие на все биомолекулы – ДНК, РНК, белки и липиды, вызывая изменения в их функциональности, например «дисульфидный стресс» у белков в результате образования нежелательных дисульфидных связей (Aslund et al., 1999) или перекисное окисление липидов, приводящее к повышению текучести мембран и, в конечном итоге, к нарушению целостности и гибели клеток (Cabiscol et al., 2000). Закономерно, что в бактериальных клетках, подверженным в том числе и множеству сторонних воздействий, выработались механизмы поддержания АФК на определенном уровне со стационарной концентрацией $<10^{-8}$ М (Gutteridge, Halliwell, 1992), а также способы репарации повреждений вследствие воздействия АФК.

Адаптивный ответ клеток *E. coli* на окислительный стресс заключается в активации двух механизмов, координирующихся глобальными транскрипционными регуляторами SoxRS и OxyR, в результате чего происходит индукция экспрессии множества генов, кодирующих белки с антиоксидантной и репарирующей активностью.

Механизм с участием OxyR регулона. OxyR регулон играет центральную роль в защите клеток *E. coli* против эндогенного окислительного стресса от пероксида водорода, образующегося в результате активного аэробного роста (González-Flecha, Demple, 1995, 2000).

Повреждающее действие пероксида водорода связано прежде всего с реакцией Фентона – прямым окислением несвязанных внутриклеточных ионов восстановленного железа, что приводит к образованию гидроксильного радикала и супероксид-радикала (Imlay, Linn, 1988; Henle et al., 1999):



Гидроксильный радикал, образующийся в этой реакции, короткоживущий, но чрезвычайно активный и способный к дальнейшим реакциям с биомолекулами вблизи сайта своего образования (Park et al., 2005; Lushchak, 2011). Восстановителями окисленного железа являются НАД(Ф)Н, тиолы, свободные восстановленные флавины (Rowley, Halliwell, 1982; Woodmansee, Imlay, 2002). Также эффективным восстановителем железа считается цистеин (Park, Imlay, 2003).

Супероксидный радикал оказывает повреждающее действие, связанное со способностью высвобождать железо из центров железо-серных дегидратаз. В ходе этого процесса происходит инактивация железо-серных

кластеров ферментов и накопление несвязанного железа в цитоплазме (Flint et al., 1993; Liochev, Fridovich, 1994; Jang, Imlay, 2007).

Под контролем OxyR находятся 9 из примерно 30 белков, которые являются основными звеньями в метаболических путях разложения пероксида водорода в бактериальных клетках. Прежде всего это каталаза HPI (hydroperoxidase I) – продукт гена *katG*, активный при аэробном росте, и каталаза HPII (hydroperoxidase II) – продукт гена *katE*, индуцирующийся в стационарной фазе роста под контролем σ S (RpoS) субъединицы РНК-полимеразы (Loewen et al., 1998). Также к OxyR регулону относятся Mn-супероксиддисмутаза, глутаредоксин I (*grxA*), глутатионредуктаза (*gorA*), алкилгидропероксидредуктаза (*ahpFC*), ДНК-связывающий белок (*dps*), копропорфириноген III редуктаза, арилсульфатаза, регулятор синтеза капслярного полисахарида и несколько белков теплового шока (Zheng et al., 2001). Также OxyR в окисленной форме активирует транскрипцию OxyS – маленькой нетранслируемой РНК, которая косвенно регулирует примерно 20 генов, нарушая стабильность РНК или эффективность трансляции (Altuvia et al., 1997). В частности, OxyS является репрессором *hyp* оперона *shlA* и гена сигма-фактора стационарной фазы RpoS (Altuvia et al., 1998; Zhang et al., 1998).

Транскрипционный фактор (ТФ) OxyR – хорошо изученный белок с массой 34.4 кДа, принадлежит к LysR семейству бактериальных ТФ. Очищенный белок OxyR существует в растворе как тетрамер (Kullik et al., 1995) и относится к активаторам класса I, которые взаимодействуют с альфа-субъединицей РНК-полимеразы на промоторе, активируя транскрипцию (Tao et al., 1993).

Синтез OxyR происходит и в отсутствие воздействия перекиси, более того, при воздействии перекиси его экспрессия не увеличивается. Этот факт свидетельствует о том, что OxyR активируется посттрансляционно (Storz et al., 1990). Восстановленная форма белка OxyR обеспечивает авторегуляцию, связываясь с промотором гена *oxyR*.

Существуют две различные реакции, определяющие статус OxyR. Экспериментально показано, что глутаредоксин I восстанавливает OxyR *in vivo*. Это выявлено по пролонгированному ответу мутантных клеток по гену *grxA* (глутаредоксин I) на воздействие перекиси. В условиях нормального роста осуществляется равновесие окисленной и восстановленной форм с преобладанием первого пути. При увеличении внеклеточной концен-

трации перекиси $>5 \mu\text{M}$ доминирует вторая реакция, и окисленная форма становится основной (Aslund et al., 1999). Эта окислительная реакция происходит очень быстро, примерно за 30 с; после удаления перекиси из среды окисленную форму OxyR можно обнаружить в клетке еще в течение 5 мин:



В условиях нормального роста при низкой концентрации пероксида водорода преобладает прямая реакция с участием глутатиона – уравнение (3). При увеличении внеклеточной концентрации перекиси $>5 \mu\text{M}$ доминирует вторая реакция – уравнение (4).

С окислительным стрессом связан и метаболизм железа: OxyR является положительным регулятором железосвязывающего репрессора переноса железа Fur (Zheng et al., 1999). Таким образом, через OxyR может осуществляться непрямая активация генов, регулятором которых является Fur.

Механизм с участием SoxRS регулона. SoxRS – это система из двух транскрипционных факторов SoxR и SoxS, вместе они составляют регулон SoxRS, активация которого происходит исключительно в условиях окислительного стресса и может быть вызвана агентами, генерирующими супероксид-анионы, такими, например, как паракват и менадион (González-Flecha, Demple, 2000).

Гены *soxS* и *soxR* кодируют два транскрипционных активатора, которые в конечном счете модулируют экспрессию более 16 других генов, таких как *sodA* (Mn-содержащая супероксиддисмутаза), *zwf* (глюкозо-6-фосфатдегидрогеназа), *fldA* и *fldB* (два флаводоксина), *fpr* (НАД(Ф)Н-ферредоксин редуктаза), *fur* (железо-связывающий репрессор переноса железа), *nfo* (ДНК репаративная эндонуклеаза IV), *acrAB* (экспортирующий насос) и *micF* (нетранслируемая РНК, которая уменьшает экспрессию порына OmpF). Белки, экспрессия которых индуцируется системой SoxRS, действуют совместно и устраняют возможный ущерб от окислительного стресса, используя механизмы удаления оксидантов (супероксиддисмутаза), ДНК репарацию (эндонуклеаза IV), восстановление окисленных металлов в протетических группах (флаводоксин- и ферредоксинредуктаза) и системы НАД(Ф)Н (глюкозо-6-фосфатдегидрогеназа), снижение проницаемости (*micF*) и экскрецию токсинов (порины). Активация генов SoxRS регулона увеличивает устойчивость клетки не только к супероксид-генерирующим агентам, но и к органическим растворителям, а также оксиду азота (NO), который может генерироваться антибиотиками (Pomposiello, Demple, 2001).

Чувствительным звеном этой системы является белок SoxR, который продуцируется постоянно на низком уровне – примерно 100 молекул на клетку. В ответ на появление супероксид-образующих агентов он приобретает свойства транскрипционного фактора и активирует экспрессию гена *soxS* (Hidalgo et al., 1995), продукт которого, в свою очередь, индуцирует экспрессию всех вышеперечисленных генов.

SoxR – полипептид размером 17 кДа (Tsaneva, Weiss, 1990), принадлежащий к MerR семейству белков. К нему

же относятся некоторые регуляторы устойчивости к антибактериальным веществам, например TipA и BmrR (Wu et al., 1995) и металл-чувствительные регуляторы, в том числе CueR – регулятор гомеостаза меди в *E. coli*, выделяющиеся внутри семейства наличием консервативных цистеинов в фиксированных позициях (Nies, 1999; Brown et al., 2003). Характерными особенностями белков, принадлежащих к семейству MerR, являются ДНК-связывающий домен спираль–поворот–спираль и способ активации транскрипции: промотор регулируемого гена обладает определенным спейсером (19–20 п. н.) и палиндромом, в обычных условиях это слабый промотор (Brown et al., 2003). Транскрипционный фактор увеличивает его силу путем изгибания ДНК и изменения расстояния между боксами.

SoxR формирует в растворе димер, каждый мономер которого содержит [2Fe-2S] кластер (Hidalgo et al., 1995; Wu et al., 1995). В этом состоянии белок неактивен. Окисленная форма [2Fe-2S] кластера – $(\text{Fe}^{3+}-\text{Fe}^{3+})$ – изменяет структуру белка SoxR, и он приобретает способность к активации транскрипции. В условиях окислительного стресса SoxR связывается со своим сайтом, блокирует собственную транскрипцию и изменяет структуру промотора SoxS таким образом, что полимеразы становится способной к активации его транскрипции (Hidalgo, Demple, 1994).

В отличие от SoxR, SoxS имеет сайты связывания в промоторах других генов, которые он и активирует в условиях стресса (González-Flecha, Demple, 2000). Однако уровень аффинности связывания SoxS с этими промоторами на два порядка ниже, чем у транскрипционного фактора SoxR, и для активации его требуется намного больше, чем SoxR. В условиях длительного стресса накопление мРНК SoxS прекращается и даже падает в связи с тем, что ее стабильность резко уменьшается. Не исключено, что и низкая аффинность SoxS (т. е. необходимость высокой концентрации ТФ SoxS для активации генов-мишеней) является необходимым компонентом прекращения стрессорной реакции в условиях длительного стресса.

Активация SoxR с помощью NO не зависит от присутствия кислорода (Nunoshiba et al., 1995). *In vivo* нитрозилированный SoxR имеет активность, сравнимую с той, которую имеет окисленный SoxR (Pomposiello, Demple, 2001).

В антиоксидантную защиту клеток вовлечен еще один регулятор – альтернативный транскрипционный фактор sigma S (RpoS, σs , KatF), регулирующий адаптационный ответ в стационарной фазе. Он кодируется геном *rpoS* и регулирует активность большого числа генов, в том числе антиоксидантных *katG*, *dps* и *gor* (Hengge-Aronis, 2002).

Основные защитные системы в *E. coli*, активируемые окислительным стрессом, можно представить следующим образом (табл. 3).

Гомеостаз ионов меди

Медь – важнейший микроэлемент, необходимый для нормального протекания множества биопроцессов. Механизмы, благодаря которым медь поступает в *E. coli*, остаются неясными, но известно, что в клетке содержится около 10^4 атомов меди – так называемая «медная квота», необходимая для нормальной физиологии (Finney, O'Halloran,

Table 3. Major mechanisms protecting *E. coli* cells from oxidative stress

soxRS (O ₂ ⁻ -inducible)	oxyR (H ₂ O ₂ -inducible)	o ^s (H ₂ O ₂ , O ₂ ⁻ -inducible)
<i>sodA</i> – superoxide dismutase	<i>katG</i> – catalase (hydroperoxidase I)	<i>katE</i> – hydroperoxidase II
<i>nfo</i> – endonuclease IV	<i>ahpCF</i> – alkylhydroperoxide reductase	<i>xthA</i> – endonuclease III
<i>acnA</i> – aconitase	<i>grxA</i> – glutaredoxin I	<i>sodC</i> – superoxide dismutase
<i>tolC</i> – outer membrane protein	<i>gorA</i> – glutathione reductase	<i>sodB</i> – superoxide dismutase
<i>micF</i> – regulatory RNA, repression of outer membrane protein expression	<i>oxyS</i> – regulatory PHK (protection from mutagenesis)	
<i>fldA</i> – flavodoxin A	<i>dps</i> – nonspecific DNA-binding protein	
<i>fpr</i> – ferredoxin/flavodoxin NADP ⁺ reductase	<i>hemF</i> – coproporphyrinogen-III oxidase	
<i>ribA</i> – GTP cyclohydrolase II	<i>rcsC</i> – regulator of genes for capsular polysaccharide synthesis	
<i>zwf</i> – glucose-6-phosphate dehydrogenase	<i>f497</i> – protein homologous to arylsulfatase enzymes	
<i>inaA</i>	<i>fur</i>	
<i>pqi5</i>		
<i>fumC</i>		
<i>fur</i>		
<i>acrAB</i>		

2003). Регуляция гомеостаза ионов меди в *E. coli* осуществляется с помощью транскрипционного фактора CueR и двухкомпонентной системы CusRS.

Транскрипционный фактор CueR, так же как и SoxR, принадлежит к MerR семейству. Для CueR тоже характерно наличие ДНК-связывающего домена спираль–поворот–спираль и способ активации транскрипции, когда в обычных условиях промотор регулируемого гена слабый, а транскрипционный фактор увеличивает его силу путем изгибания ДНК и изменения расстояния между боксами.

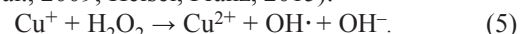
Под регуляторным контролем CueR находятся ген мультимедной оксидазы *sueO*, осуществляющей детоксикацию ионов меди, и ген АТФазного насоса Р-типа *copA*, который является центральным компонентом поддержания гомеостаза меди и осуществляет экспорт ионов в периплазматическое пространство (Outten et al., 2000; Petersen, Moller, 2000). В аэробных условиях уровень индукции системы поддержания гомеостаза меди выше через регуляторный ТФ CueR, чем через систему CusRS (Outten et al., 2001). При пролонгированном воздействии 0.5 mM CuSO₄ уровень CueR повышается примерно в 5 раз и составляет 320–400 молекул на клетку (Yamamoto, Ishihama, 2005).

Двухкомпонентная система CusRS регулирует транскрипцию оперона *cusCFBA*.

Гены, кодирующие ТФ CusR и CusS, организованы в оперон, направленный противоположно регулируемым ими генам, при этом CusR – ответный регулятор, CusS – сенсорная киназа. Продукты генов оперона *cusCFBA* образуют систему из трех компонентов (CusCBA), охватывающую внутреннюю и внешнюю мембраны, и при помощи CusF осуществляют функцию экспорта ионов меди из цитоплазматического и периплазматического пространства во внешнюю среду. Функции белков оперона *cusCFBA* распределены следующим образом: CusC – белок, который обеспечивает канал во внутренней мембране; CusA – пермеаза, осуществляющая транспорт ионов меди в цитоплазму и периплазму; CusB – член семейства белков, слитых с мембраной (membrane fusion protein – MFP); CusF – периплазматический шаперон, взаимодействующий с CusC и CusB (Franke et al., 2003). Нуль-мутанты

по *cusR* показывают высокий уровень чувствительности к меди только в анаэробных условиях. Уровень индукции CusC при низких концентрациях меди выше также в анаэробных условиях, в аэробных условиях незначительная активация происходит при концентрации ионов меди выше 0.5 mM. В среднем количество молекул CusR при стрессе от избытка ионов меди меняется мало и составляет 50–100 молекул на клетку. Таким образом, система CusRS участвует в обеспечении устойчивости к ионам в анаэробных условиях, где происходит увеличение их токсичности, но только в дополнение к эффлюксу меди при помощи CopA, экспрессия которого в анаэробных условиях также увеличивается (Outten et al., 2001; Thieme et al., 2008).

Несмотря на то что избыток ионов меди токсичен для бактерий, фундаментальные механизмы клеточных повреждений вследствие такого воздействия до сих пор остаются не до конца понятыми. Предполагается, что из-за особенностей свойств окислительно-восстановительного статуса ионы меди похожи на ионы железа. Они аналогично и даже легче железа могут участвовать в реакциях, в том числе в реакции Фентона (см. табл. 2), приводящих к образованию супероксид-аниона, перекиси водорода, а также гидроксил-радикала, крайне токсичного для клетки (Thomas et al., 2009; Helsel, Franz, 2015):



Однако в работе (Macomber et al., 2007) показано, что большая часть H₂O₂-окисленной меди локализована в периплазме. Окислительный стресс при добавлении перекиси водорода индуцировал белки секвестрации железа, но не индуцировал белки, контролирующие уровень меди. Таким образом опровергалась гипотеза об окислительном повреждении ДНК при участии внутриклеточной меди.

В другом исследовании (Macomber, Imlay, 2009), посвященном выяснению механизмов токсичности меди, использовались мутантные линии *E. coli* по генам *copA*, *sueO*, *cus*. В работе показана блокировка ферментов, содержащих Fe-S-кластеры, таких как ферменты биосинтеза аминокислот – дегидратаз (изопротилмалат и дегидратаза и дигидрокси-кислая дегидратаза), а также блокировка фумаразы А, насыщение субстратом (малатом) которой

защищало активный центр фермента от инактивации. Аналогичные данные получены в работе (Fung et al., 2013).

Взаимодействие систем гомеостаза разных металлов в клетках бактерий остается предметом тщательного изучения. Известно, что при анаэробном росте бактерий одновременное воздействие меди и железа в концентрациях, характерных для окружающей среды (1 мкМ и 200 мкМ соответственно) имеет синергический эффект, приводящий к задержке роста бактерий (Bird et al., 2013). При взаимодействии систем гомеостаза меди и железа важную роль играет мультимедная оксидаза CueO. Помимо ионов меди естественным субстратом для этого фермента является энтеробактин – соединение, синтезируемое бактерией для захвата железа из окружающей среды. Показано, что смесь энтеробактина и меди *in vitro* токсична для клетки, но добавление очищенной оксидазы CueO способствовало выживанию клеток. Делеция гена *fur* – центрального регулятора метаболизма железа в *E. coli* – приводила к гиперчувствительности бактерий к воздействию ионов меди, которая ослаблялась при делеции гена энтеробактина *entC*. Таким образом, CueO может играть центральную роль во взаимодействии между гомеостазом меди и железа (Grass et al., 2004).

EmrRAB оперон (мультилекарственная устойчивость)

В *E. coli* существует несколько систем, формирующих устойчивость клеток к различным антибактериальным веществам (Grkovic et al., 2002).

EmrRAB – оперон, участвующий в формировании мультилекарственной устойчивости *E. coli* посредством удаления из клетки таких структурно непохожих веществ, как протонофоры и антибиотики типа тиолактомицина и налидиксовой кислоты. *EmrR* – первый ген оперона, кодирующий 20.6 кДа белок. Мутации в гене *emrR* приводят к повышенной экспрессии *EmrAB*, что в свою очередь увеличивает устойчивость бактерий к антимикробным агентам. EmrR является негативным регулятором оперона и принадлежит к marR семейству транскрипционных репрессоров (Lomovskaya et al., 1995; Seoane, Levy, 1995; Miller, Sulavik, 1996).

EmrB – белок, выполняющий функцию насоса при транспорте соединений через цитоплазматическую мембрану. EmrA – белок, связывающий соединения и, вероятно, осуществляющий их передачу от EmrB к TolC (Borges-Walmsley et al., 2003). TolC – порин наружной мембраны, вовлеченный в эффлюкс некоторых гидрофобных и алифатических молекул; является общим компонентом для нескольких систем экспорта лекарств из клетки, в том числе EmrRAB, но не находится под регуляторным контролем EmrR (Morona et al., 1983).

Эффлюкс соединений системой EmrAB–TolC осуществляется с использованием градиента протонов (Putman et al., 2000).

Негативная регуляция оперона *emrRAB* осуществляется путем прямого связывания ТФ EmrR с последовательностью ДНК промотора оперона *emrRAB*. Индукторы системы *emrAB*, такие как карбонилцианид, 2,4-динитрофенол, тетрахлорсалициланилид и налидиксовая кислота, связываются с EmrR и ингибируют взаимодействие промотор–репрессор (Xiong et al., 2000).

Недавно для *E. coli* было показано, что полиамины (путресцин, спермидин и спермин) вовлечены в преодоление окислительного стресса путем стимулирования синтеза на уровне трансляции ТФ SoxR, EmrR и GshA вместе с RpoS. При этом стимуляция синтеза SoxR и EmrR зависела от наличия неспецифически расположенной последовательности Шайна–Дальгарно в мРНК SoxR и EmrR. Синтез GshA был следствием существования неспецифического иницирующего кодона UUG вместо AUG. Полиамины стимулировали синтез EmrR главным образом в логарифмической фазе роста, тогда как для SoxR и GshA стимуляция синтеза была характерна для стационарной фазы. Таким образом, *emrR* входит в полиаминный модулон – группу генов, регулируемых полиаминами (Yoshida et al., 2004; Sakamoto et al., 2015).

Тепловой шок

Тепловой шок является гомеостатическим ответом клетки на индуцированное стрессовыми условиями изменение конформации белков. Тепловой шок у *E. coli* развивается при температуре 42 °С. Механизмы защиты достаточно хорошо описаны (Arsène et al., 2000). Этот феномен свойствен всем организмам, и его молекулярный механизм практически идентичен у бактерий (Arsène et al., 2000), архей (de Macario, Macario, 1994) и эукариот (Хлебодарова, 2002). При повышении температуры клетка начинает синтезировать белки теплового шока, к которым прежде всего относятся молекулярные шапероны (DnaK, GroEL) и АТФ-зависимые протеазы (Lon, ClpAP). Эти белки выполняют две основные функции – обеспечивают фолдинг (сворачивание в нативную конформацию) и деградацию поврежденных белков. Гены стрессовых промоторов, использованных в настоящей работе, не относятся к системе защиты клетки от теплового шока. Проверенные геносенсоры не отвечали на тепловой шок.

При прохождении через водные среды ТГц излучение ими активно поглощается и диссипирует в тепло, что в свою очередь приводит к повышению температуры. Именно этот процесс тщательно контролировали при помощи тепловизора и регулировали изменением средней мощности излучения. Для исключения индукции изучаемых элементов стрессовых систем бактериальных клеток вследствие теплового шока, в ходе экспериментов для геносенсоров *E. coli*/pKatG-GFP, *E. coli*/pCopA-GFP и *E. coli*/pEmrR-GFP проводилось дополнительное тестирование на активацию после температурного воздействия 42 °С в течение 15 мин.

Активация стрессовых систем *E. coli* в условиях терагерцового облучения

С целью определения стрессовых систем *E. coli*, задействованных в ответе на нетермическое воздействие ТГц излучения, нами была проведена серия экспериментов с геносенсорами, маркирующими генные сети *E. coli*, активирующиеся в условиях окислительного стресса, присутствия антисептиков, поддержания гомеостаза ионов меди.

Одним из центральных регуляторов системы окислительного стресса является транскрипционный фактор OxyR, под положительным регуляторным контролем которого находится ген пероксидазы *katG* (Belkin et al.,

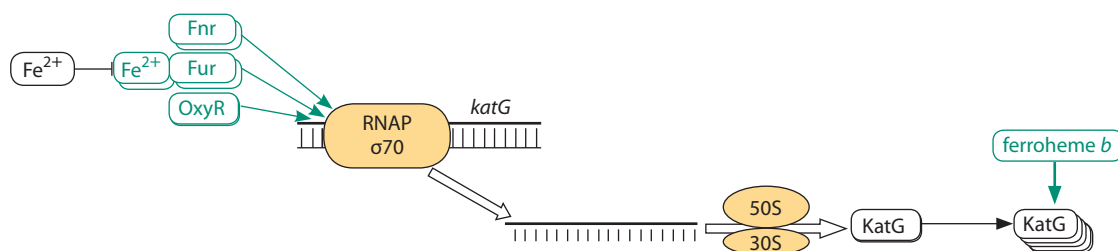


Fig. 2. Regulation of the *katG* gene (<http://ecocyc.org>).

1996; Khlebodarova et al., 2007). Появление в клетке *E. coli* перекиси водорода или супероксид-радикала приводит к активации экспрессии гена *katG* (рис. 2). KatG осуществляет реакцию восстановления супероксид-радикала, кофактором реакции выступает гем, имеющий в своем составе железо (Fe^{2+} или Fe^{3+}).

В результате серии экспериментов было показано, что воздействие ТГц излучением приводит к увеличению экспрессии гена-репортера под контролем промотора гена *katG*, чувствительного к окислительному стрессу, по сравнению с образцами, не подвергавшимися нетермическому воздействию ТГц излучения.

Поддержание концентрации кислорода внутри клетки – один из главных факторов устойчивости внутриклеточного метаболизма. На поддержание этого гомеостаза направлены несколько стрессовых систем – в частности, функционирование регулонов OxyR и SoxRS, а также сопряженного с ними гомеостаза переходных металлов. Одним из таких металлов является медь. Медь входит в состав супероксиддисмутазы SodC, или CuZnSOD, которая играет важную роль в преобразовании супероксидных радикалов до перекиси водорода и кислорода (Imlay K., Imlay J., 1996; Pena et al., 1999), однако ее избыток, так же как и недостаток, может вызывать серьезные патологии. По свойствам окислительно-восстановительного статуса ионы меди похожи на ионы железа, однако роль меди в клетке более специфична, как и сама реакция клетки на ионы меди, для которых существует специальная система эффлюкса. Известно, что ионы меди в бактериальных клетках могут влиять на гомеостаз железа, дестабилизируя белковые Fe-S кластеры (Chillappagari et al., 2010; Fung et al., 2013), в свою очередь белки, содержащие железосерные кластеры, обладают окислительно-восстановительным потенциалом и являются важнейшими участниками окислительно-восстановительных реакций в клетках. В стрессовых условиях, связанных прежде всего с окислительно-восстановительными процессами, возможна активация эффлюкса ионов меди, которые могут

оказаться токсичными для клетки. Поэтому стрессовая система поддержания гомеостаза ионов меди, а именно ген *copA*, кодирующий АТФазную помпу (Rensing et al., 1999), была выбрана для разработки и создания гено-сенсорной конструкции для дальнейшего тестирования эффектов нетермического воздействия ТГц излучения. CopA – центральный компонент поддержания гомеостаза меди, осуществляющий экспорт ионов меди в периплазматическое пространство. Активируют экспрессию ионы меди и серебра и транскрипционный фактор CueR. Для активации промотора транскрипционный фактор CueR должен связать ион Cu^+ , Ag^+ или Au^+ (рис. 3).

Было показано, что клетки гено-сенсоров под регуляторным контролем промоторов гена *copA* реагируют на терагерцовое излучение индукцией синтеза GFP белка. При этом уровень индукции репортерного белка GFP у гено-сенсора *E. coli/pCopA-GFP* заметно ниже, чем у гено-сенсора *E. coli/pKatG-GFP*. Кинетика индукционного ответа, полученная после нетермического воздействия ТГц излучения, отличается от кинетики ответной реакции на ионы меди. Для облученных образцов характерен менее интенсивный флуоресцентный ответ, чем в случае положительного контроля, однако если сравнивать с характером кинетических кривых гено-сенсора *E. coli/pKatG-GFP*, развитие ответной реакции у гено-сенсора *E. coli/pCopA-GFP* не отсрочено во времени, что позволяет предположить либо прямую активацию гена *copA*, либо небольшое количество промежуточных звеньев в процессе активации.

Еще одна стрессовая система, протестированная относительно нетермического воздействия ТГц излучения, была разработана на основе промотора гена *emrR*, входящего в оперон *emrRAB*. Ген *emrR* в составе оперона *emrRAB* кодирует белок MarR, участвующий в мультилекарственной устойчивости *E. coli* (рис. 4).

Оперон *emrRAB* придает клетке устойчивость к таким структурно несхожим соединениям, как протонофоры (CCCP и TCS), 2-хлорофенилгидразин и некоторые ан-

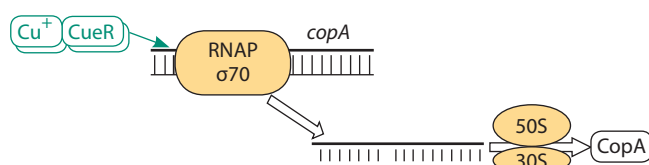


Fig. 3. Regulation of the *copA* gene (<http://ecocyc.org>).

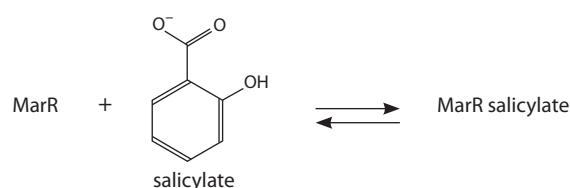


Fig. 4. Salicylate inactivation (<http://ecocyc.org>).

тибиотики (налидиксовая кислота, тиолактомицин). Ген *emrR* является негативным регулятором генов этого оперона (Lomovskaya et al., 1995). Для геносенсора *E. coli*/pEmrR-GFP после воздействия нетермическим ТГц излучением индукция флуоресцентного белка GFP не наблюдалась, для положительного контроля при индукции салициловой кислотой всегда было характерно стабильное выраженное увеличение флуоресценции.

Отсутствие индукции геносенсора *E. coli*/pEmrR-GFP указывает на то, что в реакции на ТГц излучение не участвуют стрессовые системы защиты от антибиотиков и протонофоров. Тот факт, что геносенсор *E. coli*/pEmrR-GFP (Sakamoto et al., 2015) участвует в стрессовой системе окислительного стресса, вызываемого супероксид-анионами, и в то же время не реагирует на ТГц излучение, свидетельствует в пользу того, что стрессовая система защиты от супероксид-анионов также не задействована в исследуемом процессе.

Промотор гена *copA* содержит сайт связывания транскрипционного фактора CueR, который, связываясь с ионами меди, осуществляет положительную регуляцию гена *copA* даже при умеренных концентрациях меди в аэробных условиях (Grass, Rensing, 2001; Rademacher, Masepohl, 2012). Также известно, что в промоторной области *copA* содержится гипотетический сайт связывания глобального транскрипционного фактора FruR (Fructose repressor), но непонятно, в роли активатора или репрессора он выступает. Примечательно, что транскрипционный фактор FruR тоже является активатором гена *crp* (Zhang et al., 2014), который вовлечен в генную сеть окислительного стресса и опосредованно (через OxyR) способен активировать транскрипцию гена *katG* (Demidova et al., 2016).

Известно, что FruR обеспечивает переключение между анаболизмом и катаболизмом сахаров и в случае недостатка сахаров активирует анаболические пути, а в случае их избытка – катаболические. Было показано, что в *E. coli* данный белок контролирует экспрессию большого числа генов, вовлеченных в различные процессы центрального метаболизма (Ramseier et al., 1995; Saier, 1996), в результате чего белок был переименован в Cra (catabolite repressor-activator) (Saier, Ramseier, 1996). В регуляторной области гена *emrR* сайты связывания FruR обнаружены не были.

Основная принятая на сегодняшний день гипотеза воздействия ТГц излучения на живые системы была предложена Г. Фройлихом в 1982 г. (Fröhlich, 1982). Его идея состоит в наличии резонансных частот для некоторых компонентов живых систем (например, комплексов белков) в миллиметровых и, особенно, субмиллиметровых диапазонах. Следует отметить не менее существенный приоритетный вклад отечественных ученых Н.Д. Девяткова и М.Б. Голанта в формирование представлений о воздействии излучения этого диапазона на живые системы (Девятков, Голант, 1982; Девятков и др., 1983а,б; Девятков, Бецкий, 1985). Была сформулирована гипотеза не только о прямом воздействии излучения на живые системы, но и о модификации степени гидратации органических молекул в воде (Девятков и др., 1982). Полученные нами результаты однозначно свидетельствуют о том, что нетермическое влияние ТГц излучения избирательно включает некоторые генные сети в клетках *E. coli*.

Таким образом, терагерцовое излучение, избирательно возбуждающее отдельные регуляторные элементы клетки, может оказаться мощным научным инструментом, пригодным для исследования внутриклеточных механизмов регуляции метаболизма путем направленного воздействия на отдельные элементы генных сетей.

Acknowledgments

This work was supported by the program «Study of the biologic hazard of the nonthermal action of terahertz radiation on bacterial and human cells», grant agreement 14.616.21.0053 of November 11, 2015, unique identifier RFMEFI61615X0053. Irradiation experiments were performed on a unique device, the Novosibirsk free-electron laser, at the Siberian Synchrotron and Terahertz Radiation Center (Budker Institute of Nuclear Physics, Novosibirsk).

Conflict of interest

The authors declare no conflict of interest.

References

- Altuvia S., Weinstein-Fischer D., Zhang A., Postow L., Storz G. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*. 1997;90(1):43-53.
- Altuvia S., Zhang A., Argamon L., Tiwari A., Storz G. The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J*. 1998;17(20):6069-6075.
- Andersen J., Sternberg C., Poulsen L.K., Bjorn S.P., Givskov M., Molin S. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* 1998;64(6):2240-2246.
- Arsène F., Tomoyasu T., Bukau B. The heat shock response of *Escherichia coli*. *Int. J. Food Microbiol.* 2000;55(1-3):3-9.
- Aslund F., Zheng M., Beckwith J., Storz G. Regulation of the OxyR transcription factor by hydrogen peroxide and the cellular thiol-disulfide status. *Proc. Natl. Acad. Sci. USA*. 1999;96(11):6161-6165.
- Belkin S., Smulski D.R., Vollmer A.C., Van Dyk T.K., Larossa R.A. Oxidative stress detection with *Escherichia coli* harboring a *katG*::lux fusion. *Appl. Environ. Microbiol.* 1996;62(7):2252-2256.
- Bird L.J., Coleman M.L., Newman D.K. Iron and copper act synergistically to delay anaerobic growth of bacteria. *Appl. Environ. Microbiol.* 2013;79(12):3619-3627.
- Bogomazova A.N., Vassina E.M., Goryachkovskaya T.N., Popik V.M., Sokolov A.S., Kolchanov N.A., Lagarkova M.A., Kiselev S.L., Peltek S.E. No DNA damage response and negligible genome-wide transcriptional changes in human embryonic stem cells exposed to terahertz radiation. *Sci. Rep.* 2015;5:7749. DOI 10.1038/srep07749.
- Borges-Walmsley M.I., Beauchamp J., Kelly S.M., Jumel K., Candlish D., Harding S.E., Price N.C., Walmsley A.R. Identification of oligomerization and drug-binding domains of the membrane fusion protein EmrA. *J. Biol. Chem.* 2003;278(15):12903-12912.
- Brown N.L., Stoyanov J.V., Kidd S.P., Hobman J.L. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* 2003;27(2-3): 145-163.
- Cabiscol E., Tamarit J., Ros J. Oxidative stress in bacteria and protein damage by reactive oxygen species. *Intern. Microbiol.* 2000;3(1):3-8.
- Chillappagari S., Seubert A., Trip H., Kuipers O.P., Marahiel M.A., Miethke M.J. Copper stress affects iron homeostasis by destabilizing iron-sulfur cluster formation in *Bacillus subtilis*. *J. Bacteriol.* 2010; 192(10):2512-2524.
- de Macario C.E., Macario A.J. Heat-shock response in Archaea. *Trends Biotechnol.* 1994;12(12):512-518.
- Demidova E.V., Goryachkovskaya T.N., Malup T.K., Bannikova S.V., Semenov A.I., Vinokurov N.A., Kolchanov N.A., Popik V.M., Peltek S.E. Studying the non-thermal effects of terahertz radiation

- on *E. coli*/pkatG-GFP biosensor cells. *Bioelectromagnetics*. 2013; 34(1):15-21.
- Demidova E.V., Goryachkovskaya T.N., Mescheryakova I.A., Malup T.K., Semenov A.I., Vinokurov N.A., Kolchanov N.A., Popik V.M., Peltek S.E. Impact of terahertz radiation on stress-sensitive genes of *E. coli* cell. *IEEE Transactions Terahertz Sci. Technol.* 2016;6(3):435-441. DOI 10.1109/THZ.2016.2532344.
- Devyatkov N.D., Betskiy O.V. Features of the interaction of low-intensity millimetric radiation with biologic objects. *Primenenie millimetrovogo izlucheniya nizkoy intensivnosti v biologii i meditsine* [Applications of low-intensity millimetric radiation in biology and medicine]. Moscow, 1985;6-20. (in Russian).
- Devyatkov N.D., Betskiy O.V., Ilyina S.A., Putvinskiy A.B. Influence of low-intensity millimetric radiation on the ionic permeability of erythrocyte membranes. *Effekty neteplovogo vozdeystviya millimetrovogo izlucheniya na biologicheskie ob'ekty* (pod red. N.D. Devyatkov) [Devyatkov N.D. (Ed.). Effects of the nonthermal action of millimetric radiation on biologic objects]. Moscow, IRE AN SSSR Publ. 1983a;78-96. (in Russian).
- Devyatkov N.D., Betskiy O.V., Zavizion V.A., Kudryashova V.A., Hurgin Yu.I. Absorption of electromagnetic radiation of the MM range of wavelengths and negative hydration in aqueous urea solutions. *Doklady AN SSSR = Proceedings of the Academy of Sciences of the USSR*. 1982;264(6):1409-1411. (in Russian).
- Devyatkov N.D., Golant M.B. On the informational essence of nonthermal and some energetic impacts of electromagnetic oscillations on a living organism. *Pisma v ZhTF = Letters to the Journal of Technical Physics*. 1982;8(1):39-41. (in Russian).
- Devyatkov N.D., Golant M.B., Tager A.S. A synchronization role of the action of weak electromagnetic signals of the millimetric wavelength range on living organisms. *Biofizika = Biophysics*. 1983b;28(5):895-896. (in Russian).
- Finney L.A., O'Halloran T.V. Transition metal speciation in the cell: insights from the chemistry of metal ion receptors. *Science*. 2003; 300(5621):931-936.
- Flint D.H., Tuminello J.F., Emptage M.H. The inactivation of Fe-S cluster containing hydro-lyases by superoxide. *J. Biol. Chem.* 1993; 268(30):22369-22376.
- Franke S., Grass G., Rensing C., Nies D.H. Molecular analysis of the copper-transporting efflux system CusCFBA of *Escherichia coli*. *J. Bacteriol.* 2003;185(13):3804-3812.
- Fröhlich H. What are non-thermal electric biological effects? *Bioelectromagnetics*. 1982;3(1):45-46.
- Fung D.K., Lau W.Y., Chan W.T., Yan A. Copper efflux is induced during anaerobic amino acid limitation in *Escherichia coli* to protect iron-sulfur cluster enzymes and biogenesis. *J. Bacteriol.* 2013; 195(20):4556-4568.
- González-Flecha B., Demple B. Metabolic sources of hydrogen peroxide in aerobically growing *Escherichia coli*. *J. Biol. Chem.* 1995; 270:13681-13687.
- González-Flecha B., Demple B. Genetic responses to free radicals. Homeostasis and gene control. *Ann. N.Y. Acad. Sci.* 2000;899:69-87.
- Grass G., Rensing C. Genes involved in copper homeostasis in *Escherichia coli*. *J. Bacteriol.* 2001;183(6):2145-2147.
- Grass G., Thakali K., Klebba P.E., Thieme D., Müller A., Wildner G.F., Rensing C. Linkage between catecholate siderophores and the multicopper oxidase CueO in *Escherichia coli*. *J. Bacteriol.* 2004;186(17):5826-5833.
- Grkovic S., Brown M.H., Skurray R.A. Regulation of bacterial drug export systems. *Microbiol. Mol. Biol. Rev.* 2002;66(4):671-701.
- Gutteridge J.M., Halliwell B. Comments on review of free radicals in biology and medicine (2nd ed.). *Free Radic. Biol. Med.* 1992; 12(1):93-95.
- Hakkila K., Maksimow M., Karp M., Virta M. Reporter genes *lucFF*, *luxCDABE*, *gfp*, and *dsred* have different characteristics in whole-cell bacterial sensors. *Anal. Biochem.* 2002;301(2):235-242.
- Helsel M.E., Franz K.J. Pharmacological activity of metal binding agents that alter copper bioavailability. *Dalton Trans.* 2015;44(19):8760-8770.
- Hengge-Aronis R. Interplay of global regulators and cell physiology in the general stress response of *Escherichia coli*. *Current Opinion Microbiol.* 1999;2(2):148-152.
- Hengge-Aronis R. Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* 2002;66(3):373-395.
- Henle E.S., Han Z., Nang N., Rai P., Luo Y. Sequence-specific DNA cleavage by Fe²⁺-mediated Fenton reactions has possible biological implications. *J. Biol. Chem.* 1999;274(2):962-971.
- Hidalgo E., Bollinger J.M., Jr., Bradley T.M., Walsh C.T., Demple B. Binuclear [2Fe-2S] clusters in the *Escherichia coli* SoxR protein and role of the metal centers in transcription. *J. Biol. Chem.* 1995;270(36):20908-20914.
- Hidalgo E., Demple B. An iron-sulfur center essential for transcriptional activation by the redox-sensing SoxR protein. *EMBO J.* 1994; 13(1):138-146.
- Imlay J.A., Linn S. DNA damage and oxygen radical toxicity. *Science*. 1988;240:1302-1309.
- Imlay K.R., Imlay J.A. Cloning and analysis of *sodC*, encoding the copper-zinc superoxide dismutase of *Escherichia coli*. *J. Bacteriol.* 1996;178(9):2564-2571.
- Jang S., Imlay J.A. Micromolar intracellular hydrogen peroxide disrupts metabolism by damaging iron-sulfur enzymes. *J. Biol. Chem.* 2007;282(2):929-937.
- Khlebodarova T.M. How cells protect themselves against stress? *Genetika = Genetics* (Moscow). 2002;38(4):437-452. (in Russian).
- Khlebodarova T.M., Tikunova N.V., Kachko A.V., Stepanenko I.L., Podkolodny N.L., Kolchanov N.A. Application of bioinformatics resources for genosensor design. *J. Bioinform. Comput. Biol.* 2007; 5(2B):931-938.
- Kullik I., Stevens J., Toledano M.B., Storz G. Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for DNA binding and multimerization. *J. Bacteriol.* 1995;177(5):1285-1291.
- Liochev S.I., Fridovich I. The role of O₂⁻ in the production of HO•: *in vitro* and *in vivo*. *Free Radic. Biol. Med.* 1994;16(1):29-33.
- Loewen P.C., Hu B., Strutinsky J., Sparling R. Regulation in the *rpoS* regulon of *Escherichia coli*. *Can. J. Microbiol.* 1998;44(8):707-717.
- Lomovskaya O., Lewis K., Matin A. EmrR is a negative regulator of the *Escherichia coli* multidrug resistance pump EmrAB. *J. Bacteriol.* 1995;177(9):2328-2334.
- Lushchak V.I. Environmentally induced oxidative stress in aquatic animals. *Aquat. Toxicol.* 2011;101(1):13-30.
- Macomber L., Imlay J.A. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc. Natl. Acad. Sci. USA*. 2009;106(20):8344-8349.
- Macomber L., Rensing C., Imlay J.A. Intracellular copper does not catalyze the formation of oxidative DNA damage in *Escherichia coli*. *J. Bacteriol.* 2007;189(5):1616-1626.
- Miller P. F., Sulavik M.C. Overlaps and parallels in the regulation of intrinsic multiple-antibiotic resistance in *Escherichia coli*. *Mol. Microbiol.* 1996;21(3):441-448.
- Morona R., Manning P.A., Reeves P. Identification and characterization of the TolC protein, an outer membrane protein from *Escherichia coli*. *J. Bacteriol.* 1983;153(2):693-699.
- Nies D.H. Microbial heavy-metal resistance. *Appl. Microbiol. Biotechnol.* 1999;51(6):730-750.
- Nunoshiba T., DeRojas-Walker T., Tannenbaum S.R., Demple B. Roles of nitric oxide in inducible resistance of *Escherichia coli* to activated murine macrophages. *Infect. Immun.* 1995;63(3):794-798.
- Outten F.W., Huffman D.L., Hale J.A., O'Halloran T.V. The independent cue and cus systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. *J. Biol. Chem.* 2001;276(33):30670-30677.
- Outten F.W., Outten C.E., Hale J., O'Halloran T.V. Transcriptional activation of an *Escherichia coli* copper efflux regulon by the chromosomal MerR homologue, cueR. *J. Biol. Chem.* 2000;275(40):31024-31029.

- Park S., Imlay J.A. High levels of intracellular cysteine promote oxidative DNA damage by driving the Fenton reaction. *J. Bacteriol.* 2003;185(6):1942-1950.
- Park S., You X., Imlay J.A. Substantial DNA damage from submicromolar intracellular hydrogen peroxide detected in Hpx-mutants in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* 2005;102(26):9317-9322.
- Pena M.M., Lee J., Thiele D.J. A delicate balance: homeostatic control of copper uptake and distribution. *J. Nutr.* 1999;129(7):1251-1260.
- Petersen C., Moller L.B. Control of copper homeostasis in *Escherichia coli* by a P-type ATPase, CopA, and a MerR-like transcriptional activator, CopR. *Gene.* 2000;261(2):289-298.
- Pomposiello P.J., Demple B. Redox-operated genetic switches: the SoxR and OxyR transcription factors. *TRENDS Biotechnol.* 2001;19(3):109-114.
- Putman M., van Veen H.W., Konings W.N. Molecular properties of bacterial multidrug transporters. *Microbiol. Mol. Biol. Rev.* 2000;64(4):672-693.
- Rademacher C., Masepohl B. Copper-responsive gene regulation in bacteria. *Microbiology.* 2012;158(10):2451-2464. DOI 10.1099/mic.0.058487-0.
- Ramseier T.M., Bledig S., Michotey V., Feghali R., Saier M.H. The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. *Mol. Microbiol.* 1995;16(6):1157-1169.
- Rensing C., Ghosh M., Rosen B. Families of soft-metal-ion-transporting ATPases. *J. Bacteriol.* 1999;181(9):5891-5897.
- Rowley D.A., Halliwell B. Superoxide-dependent formation of hydroxyl radicals from NADH and NADPH in the presence of iron salts. *FEBS Lett.* 1982;142(1):39-41.
- Saier M.H. Cyclic AMP-independent catabolite repression in bacteria. *FEMS Microbiol. Lett.* 1996;138(2-3):97-103.
- Saier M.H., Ramseier T.M. The catabolite repressor/activator (Cra) protein of enteric bacteria. *J. Bacteriol.* 1996;178(12):3411-3417.
- Sakamoto A., Terui Y., Yoshida T., Yamamoto T., Suzuki H., Yamamoto K., Ishihama A., Igarashi K., Kashiwagi K. Three members of polyamine modulator under oxidative stress conditions: two transcription factors (SoxR and EmrR) and a glutathione synthetic enzyme (GshA). *PLoS One.* 2015;10(4). DOI 10.1371/journal.pone.0124883.
- Seoane A.S., Levy S.B. Characterization of MarR, the repressor of the multiple antibiotic resistance (mar) operon in *Escherichia coli*. *J. Bacteriol.* 1995;177(12):3414-3419.
- Storz G., Tartaglia L.A., Ames B.N. Transcriptional regulator of oxidative stress-inducible genes: direct activation by oxidation. *Science.* 1990;248(4952):189-194.
- Tao K., Fujita N., Ishihama A. Involvement of the RNA polymerase alpha subunit C-terminal region in co-operative interaction and transcriptional activation with OxyR protein. *Mol. Microbiol.* 1993;7(6):859-864.
- Thieme D., Neubauer P., Nies D.H., Grass G. Sandwich hybridization assay for sensitive detection of dynamic changes in mRNA transcript levels in crude *Escherichia coli* cell extracts in response to copper ions. *Appl. Environ. Microbiol.* 2008;74(24):7463-7470.
- Thomas C., Mackey M.M., Diaz A.A., Cox D.P. Hydroxyl radical is produced via the Fenton reaction in submitochondrial particles under oxidative stress: implications for diseases associated with iron accumulation. *Redox Rep.* 2009;14(3):102-108.
- Tsaneva I.R., Weiss B. SoxR, a locus governing a superoxide response regulon in *Escherichia coli* K-12. *J. Bacteriol.* 1990;172(8):4197-4205.
- Woodmansee A.N., Imlay J.A. Reduced flavins promote oxidative DNA damage in nonrespiring *Escherichia coli* by delivering electrons to intracellular free iron. *J. Biol. Chem.* 2002;277(37):34055-34066.
- Wu J., Dunham W.R., Weiss B. Overproduction and physical characterization of SoxR, a [2Fe-2S] protein that governs an oxidative response regulon in *Escherichia coli*. *J. Biol. Chem.* 1995;270(17):10323-10327.
- Xiong A., Gottman A., Park C., Baetens M., Pandza S., Matin A. The EmrR protein represses the *Escherichia coli* *emrRAB* multidrug resistance operon by directly binding to its promoter region. *Antimicrob. Agents Chemother.* 2000;44(10):2905-2907.
- Yamamoto K., Ishihama A. Transcriptional response of *Escherichia coli* to external copper. *Mol. Microbiol.* 2005;56(1):215-227.
- Yoshida M., Kashiwagi K., Shigemasa A., Taniguchi S., Yamamoto K., Makinoshima H., Ishihama A., Igarashi K. A unifying model for the role of polyamines in bacterial cell growth, the polyamine modulator. *J. Biol. Chem.* 2004;279(44):46008-46013.
- Zhang A., Altuvia S., Tiwari A., Argaman L., Hengge-Aronis R., Storz G. The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J.* 1998;17(20):6061-6068.
- Zhang Z., Aboulwafa M., Saier M.H. Regulation of *crp* gene expression by the catabolite repressor/activator, Cra, in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 2014;24(3):135-141.
- Zheng M., Doan B., Schneider T.D., Storz G. OxyR and SoxRS regulation of fur. *J. Bacteriol.* 1999;181(15):4639-4643.
- Zheng M., Wang X., Templeton L.J., Smulski D.R., LaRossa R.A., Storz G. DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J. Bacteriol.* 2001;183:4562-4570.



Полиморфизм регуляторной области гена *LCT* в некоторых тюркоязычных популяциях Алтае-Саянского региона

И.В. Пилипенко¹, М.С. Пристязнюк², В.Ф. Кобзев¹, М.И. Воевода^{1, 2, 3}, А.С. Пилипенко^{1, 2, 4}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное бюджетное научное учреждение «Научно исследовательский институт терапии и профилактической медицины» Сибирского отделения Российской академии медицинских наук, Новосибирск, Россия

⁴ Федеральное государственное бюджетное научное учреждение «Институт археологии и этнографии Сибирского отделения Российской академии наук», Новосибирск, Россия

Сохранение активности фермента лактазы во взрослом возрасте (персистенция лактазы) является одним из важных адаптативно значимых признаков для популяций человека, потребляющих в пищу свежее молоко домашних животных. На молекулярно-генетическом уровне персистенция лактазы детерминирована наличием специфических аллельных вариантов ряда полиморфных позиций в цис-регуляторных элементах гена *LCT*, расположенного на хромосоме 2q21. Установление молекулярно-генетических причин персистенции лактазы сделало этот признак одной из удобных моделей для изучения механизмов адаптации популяций человека к условиям среды. Население многих регионов остается недостаточно исследованным в отношении генетической вариативности этого локуса. В данной работе приведены результаты анализа полиморфизма локуса, включающего в себя энхансерный элемент гена *LCT* и фланкирующие его районы, в двух тюркоязычных группах населения Южной Сибири – алтайских казахов и хакасов. Показано, что для тюркоязычного населения Алтае-Саянского региона из всех полиморфных вариантов, ассоциированных с персистенцией лактазы, наиболее характерным является «европейский» аллель LCT-13910T. Проникновение «европейского» аллеля LCT-13910T в генофонд населения Южной Сибири могло быть связано с миграционными волнами древних популяций животноводов Западной Евразии в Южную Сибирь в эпоху бронзы (III–II тысячелетия до н.э.). Снижение частоты аллеля LCT-13910T и суммарной частоты его носителей у тюркоязычных популяций Южной Сибири по сравнению с большинством европейских популяций и казахами Средней Азии может объясняться существенным вкладом в генофонд населения региона популяций восточно-евразийского происхождения, для которых данный аллельный вариант нехарактерен, и меньшей адаптивной значимостью для населения региона персистенции лактазы во взрослом возрасте, по сравнению с населением Европы. Редкие и уникальные однонуклеотидные полиморфизмы (ОНП) в исследуемом локусе, обнаруженные нами у алтайских казахов (LCT-13895G > C и LCT-13927C > G) и хакасов (LCT-14011C > T), являются потенциально значимыми для регуляции активности гена *LCT*, так как попадают в пределы энхансера, регулирующего активность его промотора.

Ключевые слова: персистенция лактазы; лактоза; гиполактазия; ген *LCT*; однонуклеотидные полиморфизмы; алтайские казахи; хакасы, Южная Сибирь; адаптивные признаки человека.

Polymorphism of the *LCT* gene regulatory region in Turkic-speaking populations of the Altay-Sayan region (southern Siberia)

I.V. Pilipenko¹, M.S. Pristiyazhnyuk², V.F. Kobzev¹, M.I. Voevoda^{1, 2, 3}, A.S. Pilipenko^{1, 2, 4}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Internal and Preventive Medicine SB RAMS, Novosibirsk, Russia

⁴ Institute of Archaeology and Ethnography SB RAS, Novosibirsk, Russia

Retention of lactase activity in adulthood (lactase persistence) is one of the most important adaptive traits for human populations that consume fresh milk from domestic animals. At a molecular-genetic level, lactase persistence is determined by the presence of specific alleles of polymorphic sites in cis-regulatory elements of the *LCT* gene located on chromosome 2q21. Ascertainment of the molecular-genetic causes of lactase persistence has made this trait one of the most convenient for studying mechanisms of human population adaptation to environmental conditions. But the populations of many regions remain insufficiently investigated in relation to the genetic variability of the *LCT* loci. This paper presents the results of polymorphism analysis of loci, including the enhancer element for the *LCT* gene and its flanking regions, in two Turkic-speaking populations from southern Siberia, Altaian Kazakhs and Khakasses. It was found that the “European” allele LCT-13910T is the most characteristic of the Turkic-speaking populations from Altai-Sayan regions among all the polymorphic variants associated with lactase persistence. The expansion of the “European” allele LCT-13910T to the gene pool of the populations in southern Siberia could be related to migration waves of ancient herders from western Eurasia during the Bronze Age (in III – II millennium BC). A decrease of the LCT-13910T allele frequency and the total frequency of its carriers in the Turkic-speaking populations of southern Siberia in comparison with the majority of European populations



and the Kazakhs from southern Central Asia can be attributed to: (1) a significant influence on the Altai-Sayan population's gene pool by Eastern Eurasian populations, for which the LCT-13910T allele is rare; (2) a lesser adaptive significance of lactase persistence for south Siberian populations, compared to the populations of Europe. Rare and unique SNPs in the locus under consideration that were found in the Altaian Kazakhs (LCT-13895G > C and LCT-13927C > G) and Khakasses (LCT-14011C > T) potentially play a role in regulation of LCT gene expression, because they are located within the enhancer, regulating activity of its promoter.

Key words: lactase persistence; lactose; hypolactasia; LCT gene; single nucleotide polymorphism; Altaian Kazakhs; Khakasses; Southern Siberia; human adaptive traits.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Пилипенко И.В., Пристяжнюк М.С., Кобзев В.Ф., Воевода М.И., Пилипенко А.С. Полиморфизм регуляторной области гена *LCT* в некоторых тюркоязычных популяциях Алтае-Саянского региона. Вавиловский журнал генетики и селекции. 2016;20(6):887-893. DOI 10.18699/VJ16.209

HOW TO CITE THIS ARTICLE:

Pilipenko I.V., Pristyazhnyuk M.S., Kobzev V.F., Voevoda M.I., Pilipenko A.S. Polymorphism of the *LCT* gene regulatory region in Turkic-speaking populations of the Altay-Sayan region (southern Siberia). Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):887-893. DOI 10.18699/VJ16.209

Способность усваивать молочный сахар (лактозу) во взрослом возрасте за счет сохранения активности фермента лактазы является одним из важных, адаптивно значимых признаков для многих популяций человека, потребляющих в пищу свежее молоко домашних животных. Лактаза – это фермент, ответственный за гидролиз лактозы, основного углевода молока, в глюкозу и галактозу. Поскольку молоко – единственный источник питания новорожденных, активность этого фермента становится основополагающим компонентом для раннего развития организма и выживаемости потомства. После прекращения грудного вскармливания активность лактазы, как правило, снижается. Эволюционные причины и молекулярные механизмы этого процесса до конца не изучены (Gerbault et al., 2011). У некоторых индивидов экспрессия фермента клетками эпителия кишечника сохраняется всю жизнь, что обеспечивает им нормальное переваривание свежего молока. Такие изменения в активности фермента детерминированы генетически и представляют собой два варианта физиологической нормы (Enattah et al., 2002; Gerbault et al., 2009).

Частота персистенции лактазы в популяциях человека изменяется от 15–54 % в Восточной и Южной Европе до 62–86 % в Центральной и Западной. Значительно выше она на Британских островах и в Скандинавии (89–96 %) (Ingram et al., 2009; Itan et al., 2010). В Индии частота персистенции лактазы выше в северной части (63 %) по сравнению с южной и восточной (23 %) (Swallow, Hollox, 2000). Немногочисленные сведения о популяциях Восточной Азии, опубликованные к настоящему времени, говорят о низкой частоте данного признака в этих регионах (~ 1 %) (Itan et al., 2010; Peng et al., 2012). В Африке распространение персистенции лактазы очень неоднородно: чаще всего она наблюдается у традиционных скотоводов – 64 % (Swallow, 2003; Ingram et al., 2007; Tishkoff et al., 2007; Gerbault et al., 2009). В исследованных

популяциях европейской части России этот показатель варьирует от 10 до 60 % (Khabarova et al., 2009, 2011, 2012; Боринская и др., 2009).

Установлено, что на молекулярно-генетическом уровне персистенция лактазы во взрослом возрасте детерминруется наличием специфических аллельных вариантов ряда полиморфных позиций в цис-регуляторных элементах (энхансерах) гена *LCT*, расположенного на хромосоме 2q21. Наиболее значимым для регуляции данного гена цис-элементом является энхансерная последовательность, расположенная на расстоянии ~14 т. п. н. выше старт-кодона гена *LCT*, в 13 интроне гена *MCM6*. Первым однонуклеотидным полиморфным сайтом, детерминирующим персистенцию лактазы в европейских популяциях, был ОНП, расположенный в позиции –13910 от старт-кодона гена *LCT* (rs4988235; 2:135851076 NC_000002.12) (Enattah et al., 2002). Гомозиготный вариант *C/C* по полиморфизму LCT-13910C > T ассоциирован с гиполактазией во взрослом возрасте, в то время как наличие аллеля LCT-13910T в гетерозиготном или гомозиготном состоянии приводит к персистенции лактазы (Ingram et al., 2009). Частота аллеля LCT-13910T в большинстве европейских популяций варьирует в пределах 30–90 % (Enattah et al., 2002, 2007). С помощью работ на клеточных линиях *in vitro* была продемонстрирована значимость участка, расположенного между позициями –13800 и –14017 от стартового кодона гена *LCT*, в регуляции активности его промотора (Olds, Sibley, 2003). Механизм действия данного энхансерного элемента обусловлен локализацией в нем сайтов связывания нескольких транскрипционных факторов, участвующих в транскрипции гена лактазы (Lewinsky et al., 2005).

В многочисленных ассоциативных исследованиях была доказана связь аллеля LCT-13910T с поддержанием уровня синтеза лактазы, помимо Европы, в популяциях Средней Азии, Индии, северо-западной и центральной части России (Enattah et al., 2008; Khabarova et al., 2009, 2011, 2012;

Heyer et al., 2011). Дальнейшие исследования позволили идентифицировать еще три дополнительных аллельных варианта ОНП, расположенных в непосредственной близости от LCT-13910C > T, которые были ассоциированы с персистенцией лактазы в популяциях Африки и Ближнего Востока: LCT-13907G > C (rs41525747; 2:135851073 NC_000002.12), LCT-13915A > C (rs41380347; 2:135851081) и LCT-14010C > G (rs145946881; 2:135851176) (Ingram et al., 2007; Tishkoff et al., 2007; Itan et al., 2010; Jones et al., 2013).

В то же время население многих регионов остается недостаточно исследованным в отношении генетической вариативности данного локуса. К таким популяциям можно отнести значительную часть коренного населения Южной Сибири, включая тюркоязычные популяции. При этом для ряда популяций сопредельных районов Средней Азии (в том числе для тюркоязычных казахов из северной части Узбекистана) была показана строгая корреляция между статусом полиморфной позиции LCT-13910C > T и наличием/отсутствием гиполактазии – такая же, как и для популяций Европы (Heyer et al., 2011).

Целью данной работы был анализ полиморфизма регуляторного элемента, расположенного в позиции от –14383 до –13506 от старт-кодона гена *LCT* в пределах 13 интрона гена *MCM6* (2:135851549–135850672 NC_000002.12), включающего в себя энхансерный элемент гена лактазы и фланкирующие его районы, в двух тюркоязычных группах коренного населения Южной Сибири – алтайских казахов и хакасов, для выявления частоты аллелей и генотипов распространенных ОНП, ранее ассоциированных с персистенцией лактазы в других популяциях (в первую очередь LCT-13910C > T), а также поиска новых однонуклеотидных полиморфизмов, потенциально ассоциированных с гиполактазией/персистенцией лактазы.

Материалы и методы

Материалом для исследования послужили выборки представителей коренного населения Алтае-Саянского региона: казахи из Республики Алтай (128 чел.) и хакасы из Республики Хакасия (64 чел.). Образцы крови были собраны ранее, во время экспедиционных обследований данных популяций. Письменное информированное согласие на участие в исследовании получено от всех представителей изучаемых популяций. Этническую принадлежность индивидов определяли с помощью специального анкетирования. Выборки представлены не связанными родством индивидами.

ДНК из крови выделяли с помощью фенол-хлороформной экстракции (Sambrook et al., 1989).

Скрининг однонуклеотидных полиморфизмов проводили методом прямого секвенирования ПЦР-продукта протяженного участка регуляторного района гена *LCT* размером 878 п. н. на хромосоме 2:135851549–135850672.

Для амплификации фрагмента ДНК (878 п. н.) были использованы праймеры F878 5'-agaggacttcattgtggagc-a-3', R878 5'-gggaagagggttaagactga-3'. Реакционная смесь объемом 50 мкл содержала: 75 мМ трисHCl (pH 8.8), 20 мМ (NH₄)₂SO₄, 0.01 % твин 20, 0.5 мкг тотальной ДНК, по 1 мкМ каждого из праймеров, 0.2 мМ каждого из dNTP, 2.5 мМ MgCl₂ и 0.6 ед. акт. *Taq* ДНК полимеразы («Сиб-

энзим», Новосибирск). Режим ПЦР: денатурация при 95 °C (3 мин), затем 10 циклов, включающих денатурацию при 95 °C (45 с), отжиг при 62 °C (45 с) и элонгацию при 72 °C (45 с), далее 20 циклов, включающих денатурацию при 95 °C (30 с), отжиг при 60 °C (30 с) и элонгацию при 72 °C (30 с).

Для секвенирования исследуемого фрагмента ДНК были подобраны внутренние праймеры F878seq 5'-cattgtggagcaatataaac-3' и R878seq 5'-ggctaagactgaaaagaggc-3'. Последовательности нуклеотидов определяли с использованием набора реактивов ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, США). Продукты секвенирующей реакции анализировали на автоматическом капиллярном секвенаторе ABI Prism 3130XL Genetic Analyser (Applied Biosystems, США) в центре коллективного пользования «Геномика» СО РАН (<http://sequest.niboch.nsc.ru>).

Соответствие частот генотипов равновесию Харди–Вайнберга оценивали с использованием критерия χ^2_{HW} и компьютерной программы CHNHW (Zaykin, Pudovkin, 1993). Межпопуляционное сравнение частот аллелей и носительства аллелей исследуемых ОНП проводили по критерию χ^2 с помощью программы SPSS 11.0. Различия считали статистически достоверными при уровне значимости $p < 0.05$.

Результаты

ОНП LCT-13910C > T. В исследованных популяциях алтайских казахов и хакасов с наибольшей частотой был выявлен полиморфизм в позиции LCT-13910C > T (rs4988235). В обеих выборках выявлены носители всех трех генотипов по данному ОНП (рис. 1). Распределение частот генотипов в выборке алтайских казахов соответствует равновесию Харди–Вайнберга ($\chi^2_{\text{HW}} = 0.001$), в выборке хакасов наблюдается отклонение от равновесия ($\chi^2_{\text{HW}} = 7.802$).

Обнаруженные частоты аллелей и генотипов по ОНП LCT-13910C > T (rs4988235) представлены в табл. 1. Частота аллеля LCT-13910T, ассоциированного с персистенцией лактазы в популяции алтайских казахов, составила 9.0 %, хакасов – 7.8 %. Суммарная доля носителей аллеля LCT-13910T (генотипы LCT-13910C/T, T/T) в этих популяциях составила 17.2 и 12.5 % соответственно. Достоверных различий между исследованными выборками как по частоте аллеля LCT-13910T ($p = 0.699$), так и по суммарной доле его носителей ($p = 0.399$) не выявлено (табл. 2). Таким образом, обе тюркоязычные популяции, населяющие различные части Алтае-Саянской горной системы, демонстрируют схожую представленность аллелей ОНП LCT-13910C > T (rs4988235) в генофонде. И частота аллеля ОНП LCT-13910T, и суммарная частота его носителей, выявленные у алтайских казахов и хакасов, ниже, чем в большинстве европейских популяций, а также статистически значимо ниже, чем в популяции казахов с севера Узбекистана (Heyer et al., 2011) (см. табл. 2).

Редкие ОНП локуса LCT-14383-13506. Помимо наиболее распространенного ОНП LCT-13910C > T (rs4988235), в исследованных популяциях были выявлены более редкие полиморфные варианты. В популяции алтайских казахов обнаружены ранее не описанные ОНП LCT-13895G > C

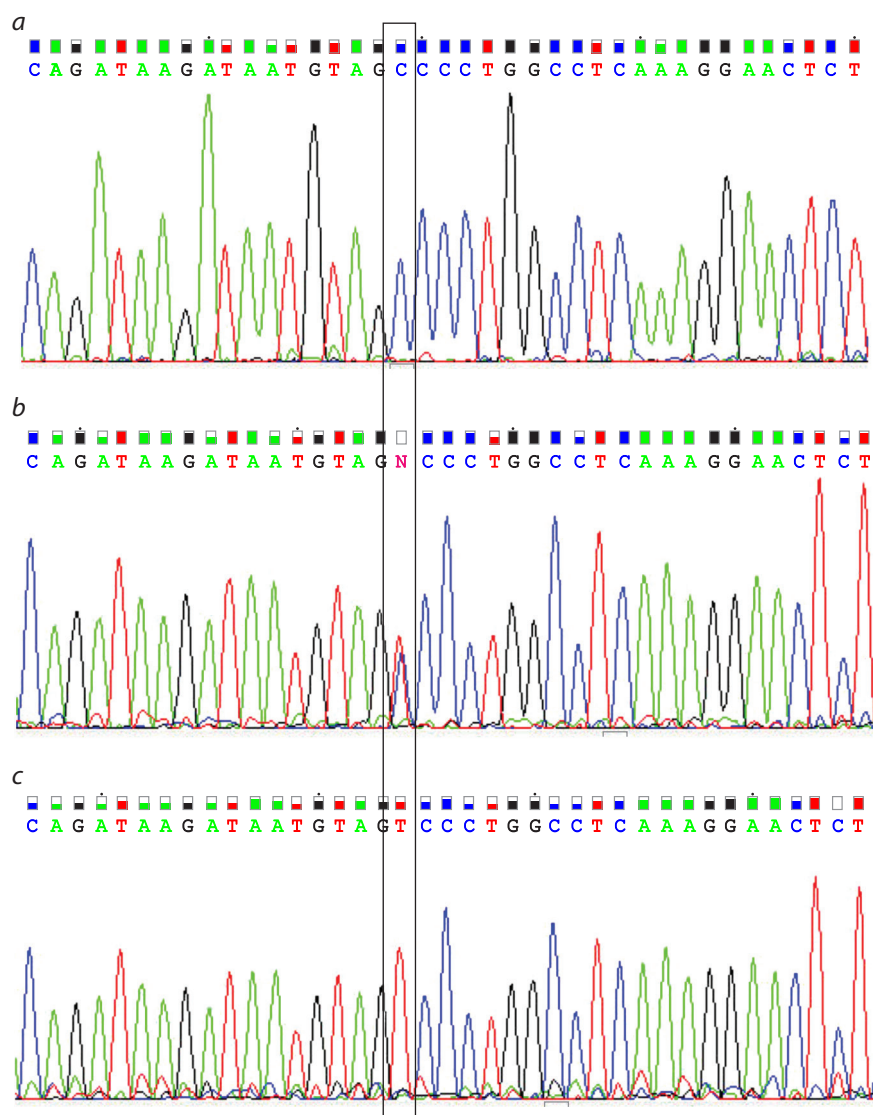


Fig. 1. Sequenograms of the cis-regulatory element of the *LCT* gene with different genotypes for the LCT-13910C>T SNP (rs4988235): (a) LCT-13910CC, (b) LCT-13910CT, (c) LCT-13910TT.

(2:135851061 NC_000002.12) и LCT-13927C>G (2:135851093 NC_000002.12) (по одному носителю ОНП в гетерозиготном состоянии) (рис. 2, а, б). При этом носитель генотипа LCT-13895GC являлся также гетерозиготой LCT-13910CT (см. рис. 2, а).

У одного представителя популяции хакасов был обнаружен в гетерозиготном состоянии ОНП LCT-14011C>T (2:135851177; rs4988233) (см. рис. 2, в), ранее выявленный с низкой частотой в популяциях различных регионов (Эстония, Индия, Пакистан, Бразилия) (Lember et al., 2006; Friedrich et al., 2012; Gallego et al., 2012).

Обсуждение

Способность усваивать лактозу во взрослом возрасте обеспечивает очевидное конкурентное преимущество носителям данного признака в популяциях, в традиционный рацион которых входит сырое молоко домашних животных. Установление достоверной взаимосвязи этого признака с полиморфизмами цис-регуляторных элементов гена лактазы *LCT* сделало персистенцию лактазы одной из наиболее удобных моделей для изучения механизмов адаптации популяций человека к условиям среды, в данном случае – к особенностям питания.

В наибольшей степени эта модель рассмотрена для популяций Европы, включая широкомасштабные исследования современных популяций из различных районов континента (Ingram et al., 2009) и анализ разновременных древних групп населения (Burger et al., 2007; Witas et al., 2015). В других регионах планеты, в том числе в южных регионах Сибири, полиморфизм регуляторных элементов гена лактазы изучен в меньшей степени.

Нами получены данные по полиморфизму цис-регуляторной области гена *LCT* для двух групп современного коренного населения Южной Сибири – алтайских казахов и хакасов. Наибольшая представленность в исследованных выборках ОНП LCT-13910C>T (rs4988235) была ожидаема, так как ранее подобный результат был получен для группы казахов из Узбекистана (Heyer et al., 2011). Важно, что для среднеазиатских популяций в упомянутой работе продемонстрирована достоверная связь аллелей по данному полиморфизму с персистенцией лактазы, аналогичная европейским популяциям (Heyer et al., 2011). Близкая частота аллеля LCT-13910T в обеих исследованных группах свидетельствует, что наши данные отражают характерную для тюркоязычного населения Южной Сибири ситуацию, а не являются случайным свойством выборок. Таким образом, для населения Южной Сибири, как и для населения Средней Азии, из всех полиморфных вариантов, ассоциированных с персистенцией лактазы во взрослом возрасте, наиболее характерным является «европейский» аллель LCT-13910T. Его частота в популяциях Южной Сибири существенно ниже, чем в популяциях Европы, а также достоверно снижена по сравнению с казахами Средней Азии (см. табл. 2).

Появление и распространение «европейского» аллеля LCT-13910T в генофонде населения Южной Сибири, на наш взгляд, могло быть связано с влиянием древних популяций из Западной Евразии, практиковавших разведение домашнего скота. Наиболее известными волнами миграции древних скотоводов в южные области Сибири являются миграция носителей афанасьевской культуры на тер-

Table 1. Allele and genotype frequencies of the LCT-13910C > T (rs4988235) polymorphism in the regulatory region of the *LCT* gene in the populations under study

Population	N	CC	CT	TT	Allele frequency, %		Frequencies of carriers of the LCT-13910T allele	χ^2_{HW}
					C	T		
Altaiian Kazakhs	128	106 (82.8 %)	21 (16.4 %)	1 (0.8 %)	91.0	9.0	22 (17.2 %)	0.001
Khakassians	64	56 (87.5 %)	6 (9.4 %)	2 (3.1 %)	92.2	7.8	8 (12.5 %)	7.802
Kazakhs from Uzbekistan (Heyer et al., 2011)	83	58 (69.9 %)	24 (28.9 %)	1 (1.2 %)	84.3	15.7	25 (30.1 %)	0.741

Table 2. Interpopulational differences in the LCT-13910T allele frequency and the overall frequency of LCT-13910T carriers as estimated by the Chi-square test

Parameter	Population	Kazakhs from Uzbekistan (Heyer et al., 2011)	Khakassians
Frequency of LCT-13910T	Altaiian Kazakhs	0.036	0.699
	Kazakhs from Uzbekistan (Heyer et al., 2011)	–	0.042
	Khakassians	–	–
LCT-13910T carrier frequency	Altaiian Kazakhs	0.027	0.399
	Kazakhs from Uzbekistan (Heyer et al., 2011)	–	0.011
	Khakassians	–	–

риторию Алтае-Саянской горной системы в III тысячелетии до н.э., предположительно с территории Восточной Европы (Вадецкая и др., 2014), а также расселение племен скотоводов-носителей андроновской (федоровской) культуры во II тысячелетии до н.э., предположительно с территории современного Казахстана (Molodin et al., 2012).

Более низкая частота варианта LCT-13910T в популяциях Южной Сибири может объясняться существенным вкладом в генофонд населения региона популяций восточно-евразийского происхождения, для которых данный аллельный вариант не характерен (Sun et al., 2007), и меньшей адаптивной значимостью для населения региона персистенции лактазы во взрослом возрасте, по сравнению с населением Европы. Так, для некоторых групп коренного населения Центральной Азии – казахов Северо-Западного Китая и особенно монголов из Внутренней Монголии, основу хозяйства которых составляет скотоводство, включая производство молока, была показана низкая распространенность персистенции лактазы во взрослом возрасте (Wang et al., 1984).

Интересно, что в исследованных нами популяциях Южной Сибири выявлено статистически значимое сни-

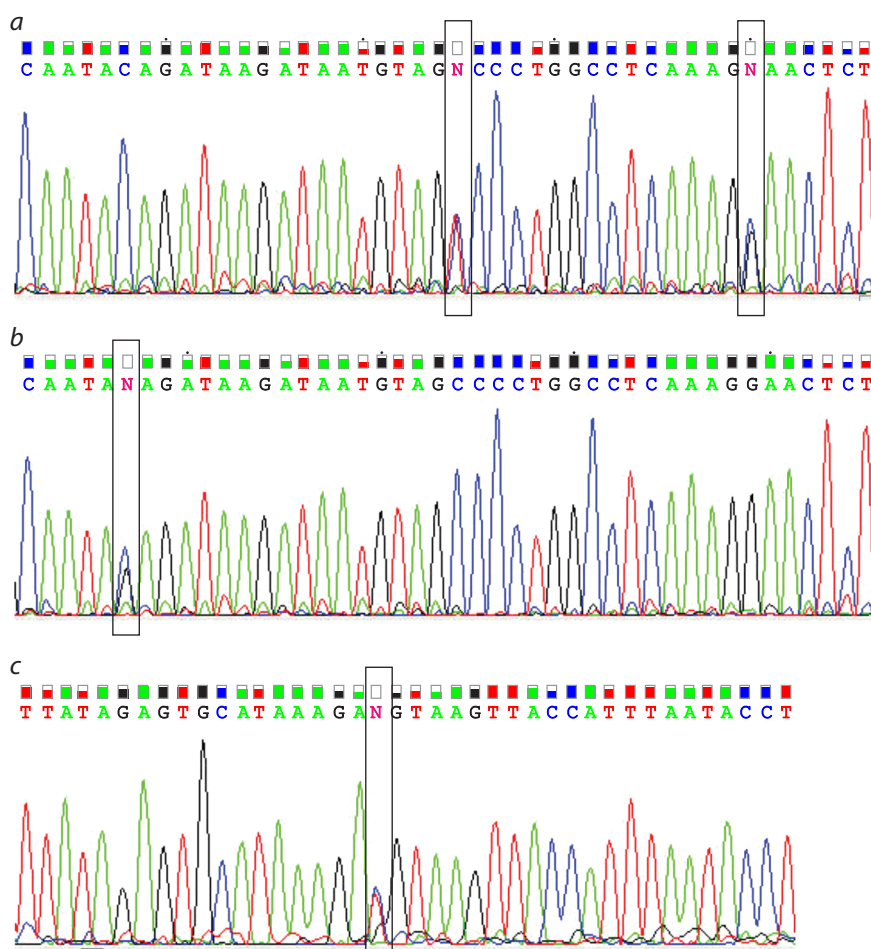


Fig. 2. Sequenograms of the cis-regulatory element of the *LCT* gene with different genotypes for SNPs: (a) LCT-13895GC (2:135851061 NC_000002.12) combined with LCT-13910CT (rs4988235), (b) LCT-13927CG (2:135851093 NC_000002.12), (c) LCT-14011CT (2:135851177, rs4988233).

–14383 (2:135851549 NC_000002.12) AGACGGAGACGATCACGTCATAGTTTATAGAGTGCATAAAGACCTGTAAGTTACCA
TTTAATACCTTTTCATTCAGG

LCT-14011C > T (2:135851177; rs4988233)

AAAAATGTACTTAGACCCCTACAATGTACTAGTAGGCCTCTGCGCTGGCAATACAGAGATAAGATAATGTAGCTCCTGGCCTC
AAAGGCAACTCTCCTCCTT

LCT-13927C > G (2:135851093);
LCT-13910C > T (2:135851076; rs4988235)
LCT-13895G > C (2:135851061)

AG GTTGCATTTGTAATGTTTGATTTTAGAATTGTTCTTTGAGCCCTGCATTCCACGAGGATAGGTCAGTGGGTATTA
AC.....–13506 (2:135850672)

Fig. 3. Locations of SNPs identified in Altaian Kazakhs and Khakassians within the sequenced fragment of the *LCT* gene regulatory region (intron 13 of the *MCM6* gene).

Previously predicted transcription factor binding sites are shown (Lewinsky et al., 2005).

жение частоты аллеля *LCT*-13910T и суммарной частоты его носителей по сравнению с казахами из Средней Азии. Это согласуется с более существенным влиянием коренных популяций восточных районов Центральной Азии (Монголия, Северный Китай и южные районы Восточной Сибири) на популяции Алтае-Саянской горной системы по сравнению с населением южных районов Средней Азии.

Другие ОНП в пределах энхансера промотора гена *LCT*, обнаруженные нами у алтайских казахов и хакасов, также заслуживают внимания и дальнейшего исследования, несмотря на их низкую частоту. Все ОНП попадают в пределы участка последовательности энхансера, значимость которого для регуляции активности промотора гена *LCT* была продемонстрирована *in vitro* (Olds, Sibley, 2003). Следует отметить, что ни один из выявленных нами ОНП не попадает в сайты связывания транскрипционных факторов, предсказанных ранее (рис. 3) (Lewinsky et al., 2005).

Данное исследование является лишь первым шагом к пониманию эволюционной истории персистенции лактазы на территории Южной Сибири. Для реконструкции более полной картины и проверки наших предположений необходим широкомасштабный анализ полиморфизма регуляторной области гена *LCT* как в группах современного коренного населения региона, так и в репрезентативных сериях образцов от разновременного древнего населения региона. Эти исследования осуществляются нашим коллективом в настоящее время.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research, project 16-06-00495, and State Budgeted Project 0324-2015-0004.

Conflict of interest

The authors declare no conflict of interest.

References

- Borinskaya S.A., Kozlov A.I., Yankovsky N.K. Genes and food traditions. Etnograficheskoe obozrenie = Ethnographic Review. 2009;3:117-137. (in Russian)
- Burger J., Kirchner M., Bramanti B.M., Haak W., Thomas M.G. Absence of the lactase-persistence-associated allele in early Neolithic

- Europeans. Proc. Natl. Acad. Sci. USA. 2007;104(10):3736-3741. DOI 10.1073/pnas.0607187104.
- Enattah N.S., Jensen T.G., Nielsen M., Lewinsky R., Kuokkanen M., Rasinpera H., El-Shanti H., Seo J.K., Alifrangis M., Khalil I.F., Nattah A., Ali A., Nattah S., Comas D., Mehdi S.Q., Groop L., Vestergaard E.M., Imtiaz F., Rashed M.S., Meyer B., Troelsen J., Peltonen L. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. Am. J. Hum. Genet. 2008;82(1):57-72. DOI 10.1016/j.ajhg.2007.09.012.
- Enattah N.S., Sahi T., Savilahti E., Terwilliger J.D., Peltonen L., Jarvela I. Identification of a variant associated with adult type hypolactasia. Nature Genet. 2002;30:233-237. DOI 10.1038/ng826.
- Enattah N.S., Trudeau A., Pimenoff V., Maiuri L., Auricchio S., Greco L., Rossi M., Lentze M., Seo J.K., Rahgozar S., Khalil I., Alifrangis M., Nattah S., Groop L., Shaat N., Kozlov A., Verschubskaya G., Comas D., Bulayeva K., Mehdi S.Q., Terwilliger J.D., Sahi T., Savilahti E., Perola M., Sajantila A., Jarvela I., Peltonen L. Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. Am. J. Hum. Genet. 2007;81(3):615-625. DOI 10.1086/520705.
- Friedrich D.C., Santos S.E.B., Ribeiro-dos-Santos A.K.S., Huts M.H. Several different lactase persistence associated alleles and high diversity of the lactase gene in the admixed Brazilian population. PLoS One. 2012;7(9):e46520. DOI 10.1371/journal.pone.0046520.
- Gallego R.I., Basu M.C., Liebert A., Crivellaro F., Chaubey G., Itan Y., Metspalu M., Easwarkanth M., Pitchappan R., Vilems R., Reich D., Singh L., Thangaraj K., Thomas M.G., Swallow D.M., Mirazon-Lahr M., Kivisild T. Herders of Indian and European cattle share their predominant allele for lactase persistence. Mol. Biol. Evol. 2012;29(1):249-260. DOI 10.1093/molbev/msr190.
- Gerbault P., Liebert A., Itan Y., Powell A., Currat M., Burger J., Swallow D.M., Thomas M.G. Evolution of lactase persistence: an example of human niche construction. Phil. Trans. R. Soc. B. 2011;366(1566):863-877. DOI 10.1098/rstb.2010.0268.
- Gerbault P., Moret C., Currat M., Sanchez-Mazas A. Impact of selection and demography on the diffusion of lactase persistence. PLoS One. 2009;4(7):e6369. DOI 10.1371/journal.pone.0006369.
- Heyer E., Brazier L., Segurel L., Hegay T., Austerlitz F., Quintana-Murci L., Georges M., Pasquet P., Veuille M. Lactase persistence in Central Asia: phenotype, genotype, and evolution. Hum. Biol. 2011;83(3):379-392. DOI 10.3378/027.083.0304.
- Ingram C.J., Elamin M.F., Mulcare C.A., Weale M.E., Tarekegn A., Raga T.O., Bekele E., Elamin F.M., Thomas M.G., Bradman N., Swallow D.M. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? Hum. Genet. 2007;120(6):779-788. DOI 10.1007/s00439-006-0291-1.

- Ingram C.J.E., Mulcare C.A., Itan Y., Thomas M.G., Swallow D.M. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* 2009;124:579-591. DOI 10.1007/s00439-008-0593-6.
- Itan Y., Jones B., Ingram C.J.E., Swallow D.M., Thomas M.G. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol. Biol.* 2010;10:36. DOI 10.1186/1471-2148-10-36.
- Jones B.L., Raga T.O., Liebert A., Zmarz P., Bekele E., Danielsen E.T., Olsen A.K., Bradman N., Troelsen J.T., Swallow D.M. Diversity of lactase persistence alleles in Ethiopia: signature of a soft selective sweep. *Am. J. Hum. Genet.* 2013;93(3):538-544. DOI 10.1016/j.ajhg.2013.07.008.
- Khabarova Y., Grigoryeva V., Tuomisto S., Karhunen P.J., Mattila K., Isokoski M. High prevalence of lactase non-persistence among indigenous nomadic Nenets, north-west Russia. *Int. J. Circumpolar Health.* 2012;71. DOI 10.3402/ijch.v71i0.17898.
- Khabarova Y., Torniaainen S.T., Nurmi H.A., Järvelä I.E., Isokoski M.K., Mattila K.J. Prevalence of lactase persistent/non-persistent genotypes and milk consumption in a young population in north-west Russia. *World. J. Gastroenterol.* 2009;15(15):1849-1853. DOI 10.3748/wjg.15.1849.
- Khabarova Y., Tornianen S., Tuomisto S., Järvelä I., Karhunen P., Isokoski M., Mattila K. Lactase non-persistent genotype influences milk consumption and gastrointestinal symptoms in Northern Russians. *BMC Gastroenterol.* 2011;11:124. DOI 10.1186/1471-230X-11-124.
- Lember M., Torniaainen S., Kull M., Kallikorm R., Saadla P., Rajasalu T., Komu H., Jarvela I. Lactase non-persistence and milk consumption in Estonia. *World. J. Gastroenterol.* 2006;12(45): 7329-7331. DOI 10.3748/wjg.v12.i45.7329.
- Lewinsky R.H., Jensen T.G., Moller J., Stensballe A., Olsen J., Troelsen J.T. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*. *Hum. Mol. Genet.* 2005;14(24):3945-3953. DOI 10.1093/hmg/ddi418.
- Molodin V.I., Pilipenko A.S., Romaschenko A.G., Zhuravlev A.A., Trapezov R.O., Chikisheva T.A., Pozdnyakov D.V. Human migrations in the southern region of the West Siberian Plain during the Bronze Age: Archaeological, palaeogenetic and anthropological data. *Population Dynamics in Pre- and Early History: New Approaches Using Stable Isotopes and Genetics.* Berlin, 2012:95-113.
- Olds L.C., Sibley E. Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a cis regulatory element. *Hum. Mol. Genet.* 2003;12(18):2333-2340. DOI 10.1093/hmg/ddg244.
- Peng M.S., He J.D., Zhu C.L., Wu S.F., Jin J.Q., Zhang Y.P. Lactase persistence may have an independent origin in Tibetan populations from Tibet, China. *J. Hum. Genet.* 2012;57(6):394-397. DOI 10.1038/jhg.2012.41.
- Sambrook J., Fritsch E.F., Maniatis T. *Molecular Cloning: A Laboratory Manual* (2nd ed.). N.Y.: Cold Spring Harbor Lab. Press, 1989.
- Sun H.M., Qiao Y.D., Chen F., Xu L.D., Bai J., Fu S.B. The lactase gene -13910T allele can not predict the lactase-persistence phenotype in north China. *Asia Pac. J. Clin. Nutr.* 2007;16(4):598-601.
- Swallow D.M. Genetics of lactase persistence and lactose intolerance. *Ann. Rev. Genet.* 2003;37:197-219. DOI 10.1146/annurev.genet.37.110801.143820.
- Swallow D.M., Hollox E.J. The genetic polymorphism of intestinal lactase activity in adult humans. The metabolic and molecular basis of inherited disease (Eds. C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle). New York: McGraw-Hill, 2000;1651-1662.
- Tishkoff S.A., Reed F.A., Ranciaro A., Voight B.F., Babbitt C.C., Silberman J.S., Powell K., Mortensen H.M., Hirbo J.B., Osman M., Ibrahim M., Omar S.A., Lema G., Nyambo T.B., Ghorji J., Bumpstead S., Pritchard J.K., Wray G.A., Deloukas P. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 2007;39(1):31-40. DOI 10.1038/ng1946.
- Vadetskaya E.B., Polyakov A.V., Stepanova N.F. *Svod pamyatnikov afanasevskoy kultury* [A compendium of Afanasevo cultural artefacts]. Barnaul, Azbuka Publ., 2014. (in Russian)
- Wang Y.G., Yan Y.S., Xu J.J., Du R.F., Flatz S.D., Kuhnau W., Flatz G. Prevalence of primary adult lactose malabsorption in three populations of northern China. *Hum. Genet.* 1984;67(1):103-106.
- Witas H.W., Ploszaj T., Jedrychowska-Danska K., Witas P.J., Maslowska A., Jerszynska B., Kozłowski T., Osipowicz G. Hunting for the LCT-13910*T allele between the Middle Neolithic and the Middle Ages suggests its absence in Dairying LBP people entering the Kuyavia region in the 8th millennium BP. *PLoS One.* 2015;10(4): e0122384. DOI 10.1371/journal.pone.0122384.
- Zaykin D.V., Pudovkin A.I. Two programs to estimate significance of χ^2 values using pseudoprobability tests. *J. Hered.* 1993;84:152.

Динамика экспрессии транскрипционных факторов REL, RELA и IRF1 в макрофагоподобной линии U937 после воздействия диоксина

Е.В. Кашина¹, Д.Ю. Ощепков¹, Е.В. Антонцева¹, М.Ю. Шаманина¹, Д.П. Фурман^{1, 2}, В.А. Мордвинов^{1, 3}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт молекулярной биологии и биофизики» Сибирского отделения Российской академии медицинских наук, Новосибирск, Россия

Арил-гидрокарбонный рецептор (AhR) – это активируемый лигандом транскрипционный фактор, который участвует в широком диапазоне критических клеточных событий в ответ на эндогенные сигналы или ксенобиотики. Одним из наиболее известных лигандов, имеющих максимальное сродство к AhR, является 2,3,7,8-тетрахлордибензо-пара-диоксин (ТХДД, диоксин). Среди диоксиновых ксенобиотиков ТХДД наиболее токсичен и вызывает широкий спектр биологических реакций, включая иммуносупрессию и рак. Комплекс лиганд:AhR:ARNT функционирует как транскрипционный фактор, связываясь со специфической последовательностью в регуляторной области генов-мишеней, называемой dioxin responsive element (DRE). Макрофаги являются ключевыми регуляторами врожденного иммунного ответа и, располагаясь во всех органах и тканях организма, одними из первых встречаются с ксенобиотиками, поэтому изучение влияния диоксина на макрофаги имеет большое значение. Распознавание потенциальных DRE в регуляторных районах генов, кодирующих транскрипционные факторы IRF1, REL, RELA, экспрессирующиеся в макрофагах, проводилось с помощью программного пакета SITECON. Ядерный экстракт и РНК были выделены из макрофагоподобных клеток линии U937, обработанных 10 нМ концентрацией ТХДД (или 0.1 % ДМСО в качестве контроля) в течение 1, 3 и 6 ч. Гель-ретардация с олигонуклеотидами, содержащими потенциально активные DRE, специфичные для промоторных районов *IRF1*, *REL*, *RELA* генов, контрольными олигонуклеотидами и антителами к AhR подтвердила, что AhR вовлечен в образование ДНК/белковых комплексов. Результаты количественной ПЦР-РВ демонстрируют достоверное повышение уровней экспрессии этих генов при воздействии 10 нМ концентрации на U937 макрофаги через 1 ч (характерное время максимальной транслокации комплекса диоксин:AhR:ARNT в ядро). Совокупность полученных результатов подтверждает функциональную активность выявленных DRE, расположенных в регуляторных районах генов *IRF1*, *REL*, *RELA* через Ah-рецепторный путь передачи сигналов.

Ключевые слова: макрофаг; диоксин; AhR; транскрипционные факторы REL, RELA, IRF1.

Dynamics of REL, RELA and IRF1 transcription factor expression in U937 macrophages after dioxin exposure

E.V. Kashina¹, D.Y. Oshchepkov¹, E.V. Antontseva¹, M.Y. Shamanina¹, D.P. Furman^{1, 2}, V.A. Mordvinov^{1, 3}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Molecular Biology and Biophysics SB RAMS, Novosibirsk, Russia

The aryl hydrocarbon receptor (AhR), a ligand-activated transcription factor, participates in a wide range of critical cellular events in response to endogenous signals or xenobiotic chemicals. 2,3,7,8-tetrachlorodibenzo-para-dioxin (TCDD) is one of the AhR ligands with a very high binding affinity for the AhR. TCDD is the most toxic among the dioxin xenobiotics and induces a broad spectrum of biological responses, including immunotoxicity and cancer. The complex ligand:AhR:ARNT functions as a transcription factor, binding to the dioxin responsive element (DRE) sequences in the regulatory regions of target genes. Macrophages are key regulators of the innate immune response, as well as one of the first types of cells which respond to chemical stress, so the study of the action of TCDD on these cells is important. Putative DREs were predicted using the SITECON software tool in the regulatory regions of the genes encoding transcription factors REL, RELA and IRF1 expressed in macrophages. Nuclear extract and total RNA were isolated from U937 macrophages treated with 10 nM TCDD (or 0.1 % DMSO as a control) for 1, 3 and 6 hours. The binding of the TCDD:AhR:ARNT transcription complex from the nuclear extract with double-stranded oligonucleotides containing the putative DREs was studied by the EMSA. Isolated RNA was used for the study of the TCDD-mediated alteration of gene expression levels using Real-time PCR with SYBR Green I. Obtained data demonstrate the functional activity of DREs in the *IRF1*, *REL*, *RELA* gene promoters via AhR signaling pathway.

Key words: macrophage; dioxin; AhR; transcription factors REL, RELA, IRF1.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Кашина Е.В., Ощепков Д.Ю., Антонцева Е.В., Шаманина М.Ю., Фурман Д.П., Мордвинов В.А. Динамика экспрессии транскрипционных факторов REL, RELA и IRF1 в макрофагоподобной линии U937 после воздействия диоксина. Вавиловский журнал генетики и селекции. 2016;20(6):894-898. DOI 10.18699/VJ16.210

HOW TO CITE THIS ARTICLE:

Kashina E.V., Oshchepkov D.Y., Antontseva E.V., Shamanina M.Y., Furman D.P., Mordvinov V.A. Dynamics of REL, RELA and IRF1 transcription factor expression in U937 macrophages after dioxin exposure. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):894-898. DOI 10.18699/VJ16.210

ORIGINAL ARTICLE

Received 03.11.2016

Accepted for publication 14.12.2016

© AUTHORS, 2016

Арил-гидрокарбоновый рецептор (AhR) – это транскрипционный фактор, являющийся членом bHLH-PAS семейства белков (Connor, Aylward, 2006). В отсутствие лиганда AhR располагается в цитоплазме клетки в комплексе с димером белков-шаперонов HSP90 и белками-кошаперонами P23 и XAP-2 (Fujii-Kuriyama, Kawajiri, 2010). К лигандам арил-гидрокарбонового рецептора относится большая группа химических веществ – как эндогенных (билирубин, метаболиты триптофана), так и экзогенных (полиароматические углеводороды, бифенилы) (Denison, Nagy, 2003). 2,3,7,8-Тетрахлордibenзо-пара-диоксин (далее ТХДД или диоксин) – это один из наиболее известных лигандов AhR, имеющий максимальное сродство к нему (Mulero-Navarro, Fernandez-Salguero, 2016). После связывания с лигандом цитоплазматический комплекс белков диссоциирует, и AhR, оставшись связанным с лигандом, транслоцируется в ядро, где взаимодействует с белком ARNT. Комплекс лиганд:AhR:ARNT взаимодействует со специфической последовательностью в регуляторной области генов-мишеней, называемой dioxin responsive element (DRE) (Connor, Aylward, 2006), что инициирует транскрипцию соответствующих генов. Коровой консенсусной последовательностью DRE является 5'-TNGCGTG-3' (Boutros et al., 2004). При взаимодействии комплекса диоксин:AhR:ARNT с DRE AhR контактирует с 5'-TNGC, а ARNT – с GTG-3' полусайтами консенсусной последовательности (Sun et al., 2004).

Известно, что воздействие диоксина на организм человека вызывает широкий спектр эффектов, выражающихся в поражении репродуктивной, нервной, эндокринной систем организма (Mandal, 2005). В последнее время все больше внимания уделяется его воздействию и на иммунную систему. Доказано, что ТХДД влияет на дифференцировку CD4+ и CD8+ T-клеток, дифференцировку дендритных клеток, а также уменьшает их количество, подавляет образование антител В-клетками, изменяет процессы активации полиморфно-ядерных клеток (Kerkvliet, 2012). Опубликованы результаты ряда экспериментальных работ и по воздействию диоксина на макрофаги, особый интерес к которым обусловлен их полифункциональностью (Biswas et al., 2012). Они участвуют в процессах фагоцитоза чужеродных веществ и бактерий, в кооперации с Т- и В-лимфоцитами реализуют и регулируют иммунный ответ, способны проявлять цитотоксическое действие на опухолевые клетки (Gordon, Taylor, 2005). Доказано, что воздействие ТХДД на макрофаги стимулирует экспрессию ряда цитокинов, например ИЛ-1 β , ИЛ-8, ФНО- α (Sciullo et al., 2009), а также некоторых транскрипционных факторов, таких как Baff и IRF3 (Esser et al., 2009).

Биоинформатический анализ, проведенный ранее нашей группой, показал, что список цитокинов и транскрипционных факторов, экспрессия которых, вероятно, способна модулироваться под воздействием диоксина, может быть значительно расширен (Furman et al., 2009). В результате экспериментальной проверки нами установлен двойной механизм регуляции экспрессии ИЛ-12 при воздействии диоксина (Oshchepkov et al., 2014). Сначала происходит прямая активация его транскрипции диоксинсодержащим комплексом арил-гидрокарбонового рецептора через DRE в промоторе, а в последующем, вследствие влияния транс-

крипционного фактора ATF3, происходит существенное снижение экспрессии ИЛ-12. Подобная модуляция продукции цитокинов макрофагом может наблюдаться и при воздействии других транскрипционных факторов. Поэтому представляется интересным проанализировать на наличие DRE и их функциональность ряд транскрипционных факторов, влияющих на регуляцию синтеза цитокинов макрофагом, и в особенности на синтез ИЛ-12 и продукцию интерферона гамма, таких как IRF1, REL и RELA.

Материалы и методы

В работе использовались следующие реактивы: [α -32P] ATP (6000 Ci/mM) (Costa Mesa, США); DNase I RNase-free (Fermentas, США); набор для проведения обратной транскрипции HighCapacity cDNA Archive Kit (Applied Biosystems, США); реакционная смесь M-436 для проведения ПЦР-РВ в присутствии SYBR Green I («Синтол», Россия); этилендиаминтетрауксусная кислота (EDTA) («Хеликон», Россия); бромистый этидий, бромфеноловый синий, 4-форбол-12-миристан-13-ацетат (ФМА), фенилметилсульфонилфторид (ФМСФ), пенициллин-стрептомицин и среда RPMI-1640 (Sigma-Aldrich, США); агароза (Hybaid-AGS, Германия); ДНК-маркер 100 п. н. (10 фрагментов от 100 до 1000 п. н.) («СибЭнзим», Россия); 2,3,7,8-ТХДД (чистота > 99.99 %) («НПО Экрос», Россия); диметилсульфоксид (ДМСО) (чистота > 99.99 %) и таблетированный препарат натрий-фосфатного буфера (Amresco, США); коктейль ингибиторов протеаз (Pierce, США); фетальная бычья сыворотка с проведенной тепловой декомплементацией (Thermo Fisher Scientific, США).

Поиск потенциальных сайтов связывания арил-гидрокарбонового рецептора. Поиск DRE в регуляторных районах генов проводился с помощью программного пакета SITECON (<http://www.mgs.bionet.nsc.ru/mgs/programs/sitecon/>; Oshchepkov et al., 2004), в соответствии с процедурой, описанной нами ранее (Furman et al., 2009). Для идентификации DRE параметры метода были выбраны таким образом, что ошибка первого рода (неопределение) была равна нулю, что означает безошибочное распознавание сайтов во всех последовательностях из обучающей выборки. Последовательности были извлечены из базы данных UCSC Genome Browser database (Meyer et al., 2013). Анализировались участки генов от –5000 до +600 относительно старта транскрипции.

Культивирование клеток. Исследования проводились на перевиваемой гистиомоноцитойдной клеточной линии человека U937, полученной из коллекции клеточных культур ГНЦ ВБ «Вектор». С целью получения макрофагального фенотипа клетки инкубировали с 4-форбол-12-миристан-13-ацетатом в концентрации 0.2 мкг/мл в течение 48 ч, затем клетки отмывали стерильным фосфатным буфером и на трое суток помещали в полную среду (RPMI 1640 (10 % FBS)), не содержащую форболового эфира. Полученные макрофагоподобные клетки (далее – U937 макрофаги) обрабатывались 10 нМ/л концентрацией 2,3,7,8-ТХДД (опыт) или 0.1 % (v/v) диметилсульфоксидом (контроль) в течение 1, 3 и 6 ч. По окончании времени воздействия клетки собирали с помощью скребка и использовали для получения суммарной РНК и ядерного экстракта.

Метод задержки ДНК-зонда в геле белками ядерного экстракта. Получение белкового экстракта ядер из U937 макрофагов проводили следующим методом: 10^7 клеток инкубировали с 1 мл лизирующего буфера (10 mM Hepes, 10 mM KCl, 1 mM ДТТ, 0.5 mM спермидина, 0.15 mM спермина, 0.1 mM ЭДТА, 0.1 mM ЭГТА, 0.5 mM ПМСФ, коктейль ингибиторов протеаз Pierce) в течение 15 мин на льду, затем добавляли 62 мкл 10 % NP-40 и центрифугировали в течение 5 мин при 4 °C, 400 g. Осадок ядер лизировали в буфере (20 mM Hepes, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM ЭДТА, 1 mM ДТТ, 25 % глицерин, коктейль ингибиторов протеаз Pierce) в течение 20 мин на льду, после чего центрифугировали 10 мин при 4 °C, 10000 g. Супернатант, содержащий суммарный белок ядер, отбирали в чистую пробирку, измеряли концентрацию на спектрофотометре по методу Бредфорда, фасовали на аликвоты и хранили при -72 °C. В качестве ДНК-зондов использовали двуцепочечные олигонуклеотиды, соответствующие районам расположения предсказанных сайтов связывания AhR. Введение метки в ДНК-зонд осуществляли с помощью достройки укороченных 3'-концов фрагментом Клёнова ДНК-полимеразы I. Реакцию вели 5 мин при комнатной температуре в 10 мкл реакционной смеси, содержащей 0.01 mM олигонуклеотида, 1 мкл $10\times$ буфера для мечения (500 mM трис-HCl pH = 8.0, 100 mM NaCl, 100 mM $MgCl_2$, 1 mM ДТТ, 2 mM dГТФ, 2 mM dТТФ, 2 mM dСТФ), 2 единицы активности фрагмента Клёнова, 10 мкКи (α -32P)dATФ. Для очистки олигонуклеотидов от не включившейся метки реакционную смесь наносили на ионообменную бумагу ДЕАЕ-81 и промывали 0,25 M KH_2PO_4 .

Анализ связывания фрагментов ДНК с белками ядерных экстрактов проводили по следующей методике: белковый ядерный экстракт инкубировали с обработанной ультразвуком ДНК спермы лосося (из расчета 1 мкг ДНК на 7 мкг суммарного белка) в течение 10 мин на льду для предотвращения неспецифического связывания белков с ДНК-зондом. После этого 4 мкг экстракта добавляли к пробам, содержащим 50 пМ радиоактивно меченого ДНК-зонда, и инкубировали при комнатной температуре в течение 15 мин. Электрофорез вели в 4.5 % ПААГ в $0.5\times$ TBE при 4 °C. Гель выдерживали в фиксирующем растворе, высушивали и экспонировали с рентгеновской пленкой.

Количественная ПЦР в реальном времени (ПЦР-РВ). Суммарная РНК выделялась с помощью набора Aurum Total RNA Mini Kit (производство Bio-Rad). Качество и количество выделенной РНК оценивались на приборе Agilent 2100 Bioanalyzer ($RIN \geq 7$ для всех проб), согласно инструкции производителя. ДНКазную обработку проводили набором DNase I, RNase-free фирмы Fermentas, согласно инструкции производителя. Для синтеза кДНК использовали набор High Capacity cDNA Archive Kit (Applied Biosystems). Уровень экспрессии генов *REL*, *RELA*, *IRF1* и *CYP1A1* определяли методом количественной ПЦР в реальном времени, используя набор реактивов «М-436» для ПЦР-РВ, в присутствии SYBR Green I («Синтол») на приборе ABI PRISM 7000. Дизайн праймеров для вышеперечисленных генов и гена *ACTB* (как эндогенного контроля) осуществлялся

с помощью программы Primer Express® Software v2.0 (Applied Biosystems). Последовательности праймеров: *CYP1A1 F* (5'-CCCAGCTCAGCTCAGTACCT-3'); *CYP1A1 R* (5'-GAGGCCAGAAGAACTCCGT-3'); *ACTB F* (5'-TGGCATTTGCCGACAGGAT-3'); *ACTB R* (5'-CGGAGTACTTGCCTCAGGA-3'); *REL F* (5'-GACCAGGATGGTCTTGAAGTCC-3'); *REL R* (5'-GACCAGGTCTTGGTCTCTCAGG-3'); *RELA F* (5'-TCAAGATCTGCCGAGTGACC-3'); *RELA R* (5'-CACCTCAATGTCCTCTTTCTGC-3'); *IRF1 F* (5'-GCACCAGTGATCTGTACAAGTTC-3'); *IRF1 R* (5'-TCATGATGTCCTCAGGTAATTTCC-3'). Условия проведения реакции соответствовали инструкции производителя набора для ПЦР-РВ и были одинаковы для всех пар праймеров.

Статистический анализ. Полученные в результате количественной ПЦР-РВ данные от нескольких независимых экспериментов, проведенных на разных клеточных пассажах, были подвергнуты статистическому анализу с использованием программы Statistica 6.0 (Statsoft, США), визуализированы с помощью Microsoft Office Excel 2007 и Microsoft Office PowerPoint 2007 (Microsoft, США). Достоверность различий между экспериментальными группами оценивали с помощью t-критерия Стьюдента с использованием программы Statistica 6.0 (Statsoft, США).

Результаты

Результат поиска сайтов связывания. Результаты распознавания потенциальных DRE в регуляторных районах генов транскрипционных факторов *IRF1*, *REL* и *RELA*, экспрессирующихся в макрофаге, суммированы в табл. 1.

Одним из показателей, позволяющих оценить выявленные сайты как потенциальные DRE, являлась их позиция вблизи старта транскрипции (Sun et al., 2004; Fujita et al., 2006). Как видно из приведенных данных, в каждом гене присутствуют DRE, расположенные не далее 500 п. о. от старта транскрипции, что делает их вероятными участниками регуляции этих генов при воздействии диоксина. Для выборочной экспериментальной проверки методом задержки в геле было выбрано по одному DRE для каждого из генов транскрипционных факторов *REL*, *RELA*, *IRF1*.

Метод задержки ДНК-зонда в геле белками ядерного экстракта. Для доказательства AhR-зависимого механизма активации генов транскрипционных факторов *IRF1*, *REL* и *RELA* в U937 макрофагах под действием 10 нМ/л концентрации диоксина использовали метод задержки ДНК-зонда в геле белками ядерного экстракта. В качестве дополнительного контроля был взят функциональный DRE из гена *CYP1A1*, кодирующего один из ферментов, осуществляющих биотрансформацию ксенобиотиков. Быстрое и длительное повышение экспрессии гена *CYP1A1* наблюдается во множестве типов клеток, включая макрофаги, в ответ на обработку диоксином (Sciullo et al., 2009). Появление ДНК-связывающей формы AhR в ядерном экстракте, выделенном из U937 макрофагов, выявляли с помощью добавления антител к арил-гидрокарбонному рецептору. После инкубации ядерного экстракта с антителами к AhR (+) происходит ослабление формирования ДНК-белкового комплекса K, что видно из попарного сравнения дорожек 1 и 2, 3 и 4, 5 и 6, 7 и 8 (рис. 1).

Table 1. DREs predicted in the genes encoding the REL, RELA and IRF1 transcription factors

Gene	Site position relative to the transcription initiation site	Site sequence
<i>IRF1</i>	<u>-145</u>	cggctGCGTGccgt
<i>REL</i>	<u>-207</u> <u>-346</u>	ggggaGCGTGcgcg tagcaGCGTGagaa
<i>RELA</i>	<u>-110</u> <u>-137</u> <u>-199</u>	gcgcgGCGTGcact gctgtGCGTGcagc ggctgGCGTGcccg
Consensus sequence		TNGCGTG

DREs taken into subsequent experimental validation are underlined.

Добавление антител способствует образованию нового комплекса С (суперсдвиг с антителами к AhR), появляющегося на дорожках 4, 6 и 8, что служит дополнительным подтверждением специфичности исследуемого комплекса. Полученные результаты свидетельствуют, что AhR вовлечен в образование ДНК/белковых комплексов с Р³² мечеными олигонуклеотидами, содержащими предсказанные DRE, расположенные в промоторных районах генов транскрипционных факторов *IRF1*, *REL* и *RELA*.

Результаты количественной ПЦР в реальном времени. Результаты ПЦР в реальном времени демонстрируют изменение уровней экспрессии исследуемых генов при воздействии 10 нМ/л концентрации диоксина (рис. 2, табл. 2) по сравнению с уровнями экспрессии в контроле (клетки, обработанные 0,1 % ДМСО). Наблюдается увеличение уровня экспрессии исследуемых генов и контрольного гена *CYP1A1* через 1 ч с момента воздействия (см. рис. 2), что соответствует литературным данным о времени максимальной транслокации транскрипционного комплекса диоксин:AhR:ARNT в ядро (Komura et al., 2001).

Однако уже через 3 ч уровень экспрессии мРНК всех трех генов транскрипционных факторов был сопоставим с уровнями экспрессии в контрольной группе (см. табл. 2).

Уровень экспрессии мРНК контрольного гена *CYP1A1* демонстрирует устойчивое нарастание в зависимости от времени воздействия диоксином, что согласуется с литературными данными (Sciullo et al., 2009).

Обсуждение

Гель-ретардация с олигонуклеотидами, содержащими потенциально активные DRE, специфичные для промоторных районов *IRF1*, *REL*, *RELA* генов, контрольным олигонуклеотидом и антителами к AhR подтвердила, что AhR вовлечен в образование ДНК/белковых комплексов.

Результаты количественной ПЦР-РВ демонстрируют достоверное повышение уровней экспрессии генов *IRF1*, *REL*, *RELA* в результате воздействия 10 нМ/л концентрации диоксина на U937 макрофаги по сравнению с контролем (0,1 % ДМСО) через 1 ч, что является характерным временем максимальной транслокации комплекса диоксин:AhR:ARNT в ядро (Komura et al., 2001). Таким

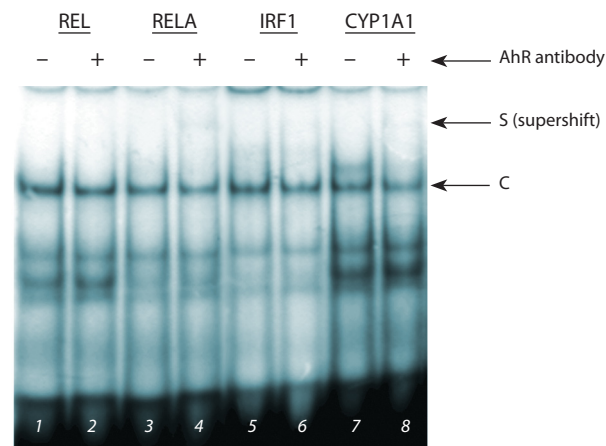


Fig. 1. The binding of nuclear extract proteins obtained from U937 macrophages to α -³²P-end-labeled oligonucleotides containing DREs.

The electrophoretic image of DNA-protein complexes formed by incubation of α -³²P-end-labeled oligonucleotides containing predicted DREs for the *IRF1*, *REL* and *RELA* genes and a control DRE from the *CYP1A1* gene enhancer with nuclear extract proteins obtained from cells treated with 10 nM dioxin for 24 h. C, the retention band corresponding to the dioxin:AhR:ARNT protein complex with the oligonucleotide containing a putative DRE.

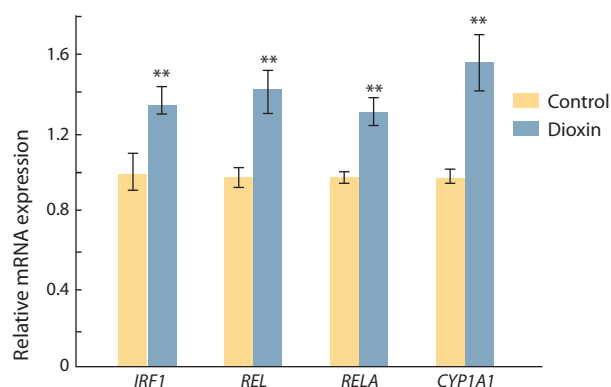


Fig. 2. Effects of 2,3,7,8-TCDD on the mRNA expression of *IRF1*, *REL*, *RELA* and *CYP1A1* genes in U937 macrophages after 1 hour of exposure.

Cells were treated with 10 nM 2,3,7,8-TCDD for 1 hour (dioxin) or 0.1 % DMSO (control). The levels of the mRNA expressions of the genes were measured by qPCR and normalized to the expression of *ACTB*. The Y-axis shows the mean values of the relative mRNA expression and the standard error of the mean.

**compared to the control, $p < 0.01$.

образом, совокупность полученных результатов подтверждает функциональную активность предсказанных DRE, расположенных в регуляторных районах генов *IRF1*, *REL*, *RELA*, и позволяет сделать вывод о том, что транскрипционная индукция этих генов опосредуется через Ah-рецепторный путь передачи сигнала.

Наблюдаемые через 3 и 6 ч после воздействия диоксина на U937 макрофаги уровни экспрессии мРНК генов *IRF1* и *RELA* не отличаются от экспрессии этих генов в контроле (см. табл. 2). Это позволяет предположить, что регуляторная сеть компенсирует воздействие диоксина на указанные транскрипционные факторы, однако отклик этой регуляторной сети возможен для других ее компонентов.

Table 2. Early effects of 10nM 2,3,7,8-TCDD on the mRNA expression of the *IRF1*, *REL*, *RELA*, and *CYP1A1* genes in U937 macrophages

Time, h	Relative mRNA expression levels							
	<i>IRF1</i>		<i>REL</i>		<i>RELA</i>		<i>CYP1A1</i>	
	C	D	C	D	C	D	C	D
3	1.002±0.13	1.11±0.26	0.960±0.19	0.99±0.17	0.96±0.12	1.06±0.19	0.95±0.13	3.59±0.45***
6	1.07±0.23	1.10±0.22	0.997±0.06	1.55±0.23**	0.99±0.06	1.14±0.27	0.99±0.1	6.03±0.85***

Cells were treated with 10 nM 2,3,7,8-TCDD for 3 and 6 hours (dioxin, D) or 0.1 % DMSO (control, C). The levels of the mRNA expression of the genes were measured by qPCR and normalized to the expression of the *ACTB* gene. The results are shown as mean values of relative mRNA expression and standard errors of the mean.

** $p < 0.01$, *** $p < 0.001$ as compared to the control.

Интересно, что уровень экспрессии мРНК транскрипционного фактора REL достоверно повышается по сравнению с контролем как через 1 ч, так и через 6 ч с момента воздействия диоксина (см. рис. 2, табл. 2). Таким образом, реализуется сходный с обнаруженным ранее для генов субъединиц ИЛ-12 двухстадийный ответ на воздействие диоксина. Известно, что этот транскрипционный фактор играет ключевую роль в патогенезе ревматоидного артрита и псориаза, в развитии фиброза кожи и миокарда. Его индукция характерна для ряда солидных опухолей (Hunter et al., 2016), поэтому необходимо проведение дополнительных исследований влияния диоксина на экспрессию гена *REL* в макрофагах в зависимости от дозы и времени воздействия.

Закключение

Совокупность полученных результатов подтверждает функциональную активность предсказанных DRE, расположенных в регуляторных районах генов *IRF1*, *REL*, *RELA*, через Ah-рецепторный путь передачи сигналов.

Acknowledgments

This work was supported by State Budgeted Project 0324-2015-0004. The qPCR experiments with U937 macrophages were supported by the Russian Science Foundation, project 14-32-00006. The computer-assisted analysis of the results was supported by the Russian Science Foundation, project 14-24-00123.

Conflict of interest

The authors declare no conflict of interest.

References

- Biswas S.K., Chittezhath M., Shalova I.N., Lim J.Y. Macrophage polarization and plasticity in health and disease. *Immunol. Res.* 2012; 53(1/3):11-24.
- Boutros P.C., Moffat I.D., Franc M.A., Tuomisto J., Okey A.B. Dioxin-responsive AHRE-II gene battery: identification by phylogenetic footprinting. *J. Biochem. Biophys. Res. Commun.* 2004;321(3):707-715.
- Connor K.T., Aylward L.L. Human response to dioxin: aryl hydrocarbon receptor (AhR) molecular structure, function, and dose-response data for enzyme induction indicate an impaired human AhR. *J. Toxicol. Environ. Health B. Crit. Rev.* 2006;9(2):147-171.
- Denison M.S., Nagy S.R. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.* 2003;43:309-334.
- Esser C., Rannug A., Stockinger B. The aryl hydrocarbon receptor in immunity. *Trends Immunol.* 2009;30(9):447-454. DOI 10.1016/j.it.2009.06.005.
- Fujii-Kuriyama Y., Kawajiri K. Molecular mechanisms of the physiological functions of the aryl hydrocarbon (dioxin) receptor, a multifunctional regulator that senses and responds to environmental stimuli. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* 2010;86(1):40-53.
- Fujita H., Samejima H., Kitagawa N., Mitsuhashi T., Washio T., Yonemoto J., Tomita M., Takahashi T., Kosaki K. Genome-wide screening of dioxin-responsive genes in fetal brain: bioinformatic and experimental approaches. *Cong. Anom.* 2006;46:135-143.
- Furman D.P., Oshchepkova E.A., Oshchepkov D.Yu., Shamanina M.Yu., Mordvinov V.A. Promoters of the genes encoding the transcription factors regulating the cytokine gene expression in macrophages contain putative binding sites for aryl hydrocarbon receptor. *Comput. Biol. Chem.* 2009;33(6):465-468.
- Gordon S., Taylor P.R. Monocyte and macrophage heterogeneity. *Nat. Rev. Immunol.* 2005;5(12):953-964.
- Hunter J.E., Leslie J., Perkins N.D. c-Rel and its many roles in cancer: an old story with new twists. *Br. J. Cancer.* 2016;114(1):1-6. DOI 10.1038/bjc.2015.410.
- Kerkvliet N.I. TCDD: an environmental immunotoxicant reveals a novel pathway of immunoregulation – a 30-year odyssey. *Toxicol. Pathol.* 2012;40(2):138-142.
- Komura K., Hayashi S., Makino I., Poellinger L., Tanaka H. Aryl hydrocarbon receptor/dioxin receptor in human monocytes and macrophages. *Mol. Cell Biochem.* 2001;226(1/2):107-118.
- Mandal P.K. Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology. *J. Comp. Physiol. B.* 2005;175:221-230.
- Meyer L.R., Zweig A.S., Hinrichs A.S., Karolchik D., Kuhn R.M., Wong M., Sloan C.A., Rosenbloom K.R., Roe G., Rhead B., Raney B.J., Pohl A., Malladi V.S., Li C.H., Lee B.T., Learned K., Kirkup V., Hsu F., Heitner S., Harte R.A., Haussler M., Gurdvadoo L., Goldman M., Giardine B.M., Fujita P.A., Dreszer T.R., Diekhans M., Cline M.S., Clawson H., Barber G.P., Haussler D., Kent W.J. The UCSC Genome Browser database: extensions and updates 2013. *Nucl. Acids Res.* 2013;41:64-69.
- Mulero-Navarro S., Fernandez-Salguero P.M. New trends in Aryl hydrocarbon receptor biology. *Front. Cell Dev. Biol.* 2016;4(45). DOI 10.3389/fcell.2016.00045.
- Oshchepkov D.Y., Kashina E.V., Antontseva E.V., Oshchepkova E.A., Mordvinov V.A., Furman D.P. Dynamics of IL-12 cytokine expression in human macrophages after dioxin exposure. *Rus. J. Genetics: Appl. Res.* 2014;4(6):568-574.
- Oshchepkov D.Y., Vityaev E.E., Grigorovich D.A., Ignatieva E.V., Khlebodarova T.M. SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition. *Nucl. Acids Res.* 2004;32:208-212.
- Sciullo E.M., Dong B., Vogel C.F., Matsumura F. Characterization of the pattern of the nongenomic signaling pathway through which TCDD-induces early inflammatory responses in U937 human macrophages. *Chemosphere.* 2009;74(11):1531-1537.
- Sun Y.V., Boverhof D.R., Burgoon L.D., Fielden M.R., Zacharewski T.R. Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences. *Nucl. Acids Res.* 2004;32:4512-4523.



Симбиотическая бактерия *Wolbachia* в популяциях вредителя хвойных лесов *Dendrolimus superans sibiricus* Tschetverikov, 1908 (Lepidoptera: Lasiocampidae)

М.А. Юдина^{1, 2}, В.В. Дубатолов³, Р.А. Быков¹, Ю.Ю. Илинский^{1, 2}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное бюджетное учреждение науки Институт систематики и экологии животных Сибирского отделения Российской академии наук, Новосибирск, Россия

Сибирский шелкопряд – крайне опасный вредитель хвойных деревьев, в особенности лиственницы и различных видов сосен. Вспышки численности сибирского шелкопряда приводят к дефолиации и гибели лесов на обширных площадях в азиатской части Российской Федерации. Многие биологические агенты, такие как вирусы, патогенные микроорганизмы и паразитоиды, сдерживают рост популяции вредителя. В данной работе рассмотрен непатогенный симбиотический микроорганизм *Wolbachia*, который передается от особи к особи трансовариально и может оказывать серьезное влияние на биологию вида-хозяина. Потенциально *Wolbachia* может как сдерживать, так и стимулировать рост популяции вредителя, что и определяет ключевой интерес исследования. Две выборки сибирского шелкопряда, собранные в 2014 и 2016 гг. в Хабаровском крае, были исследованы на присутствие *Wolbachia*. Обнаружена высокая частота инфицированных особей сибирского шелкопряда. Доля носителей эндосимбионта в 2014 г. составила 100 % и в 2016 г. – 90 %. Кроме того, в исследованной популяции сибирского шелкопряда на основании протокола мультилокусного генотипирования по генам *ftsZ* и *fbpA* обнаружено присутствие по меньшей мере двух штаммов *Wolbachia*, характеризующихся аллельными вариантами *ftsZ*-36, *fbpA*-4 и *ftsZ*-22, *fbpA*-9. В статье обсуждается возможная роль *Wolbachia* в симбиотической ассоциации с сибирским шелкопрядом и направления дальнейших исследований.

Ключевые слова: сибирский шелкопряд; *Dendrolimus superans sibiricus*; *Wolbachia*; симбиоз; популяция; ПЦР.

Wolbachia infection in populations of the coniferous forest pest *Dendrolimus superans sibiricus* Tschetverikov, 1908 (Lepidoptera: Lasiocampidae)

М.А. Yudina^{1, 2}, V.V. Dubatolov³, R.A. Bykov¹, Yu.Yu. Ilinsky^{1, 2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia

Siberian silk moth (*Dendrolimus superans sibiricus*) is a very dangerous pest of coniferous trees, in particular, larch and various pine species. Outbreaks of this pest lead to defoliation and forest destruction in a vast area of the Asian part of Russia. Many biological agents, such as viruses, pathogenic microorganisms and parasitoids, prevent the growth of Siberian silk moth population. Here we consider non-pathogen symbiotic *Wolbachia* bacteria, which are transovarially transmitted between specimens from mother to offspring. This symbiont has an ability to affect biology of its host. In theory, *Wolbachia* can prevent the growth of population size or induce it, which determines the focus of interest in *Wolbachia*-host investigation. Two samples from a Siberian silk moth population collected in 2014 and 2016 in Khabarovsk area were studied for *Wolbachia* infection. We found a high *Wolbachia* prevalence in the population of Siberian silk moth, in particular, the sample of 2014 was totally infected and the sample of 2016 had 90 % infected specimens. There were at least two distinct *Wolbachia* strains revealed by analysis of two loci from the MLST protocol, namely *ftsZ*-36, *fbpA*-4 and *ftsZ*-22, *fbpA*-9. In this study, a possible role of *Wolbachia* in the symbiotic association with Siberian silk moth and general ways of investigation of this symbiosis are discussed.

Key words: Siberian silk moth; *Dendrolimus superans sibiricus*; *Wolbachia*; symbiosis; population; PCR.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Юдина М.А., Дубатолов В.В., Быков Р.А., Илинский Ю.Ю. Симбиотическая бактерия *Wolbachia* в популяциях вредителя хвойных лесов *Dendrolimus superans sibiricus* Tschetverikov, 1908 (Lepidoptera: Lasiocampidae). Вавиловский журнал генетики и селекции. 2016;20(6):899-903. DOI 10.18699/VJ16.208

HOW TO CITE THIS ARTICLE:

Yudina M.A., Dubatolov V.V., Bykov R.A., Ilinsky Yu.Yu. *Wolbachia* infection in populations of the coniferous forest pest *Dendrolimus superans sibiricus* Tschetverikov, 1908 (Lepidoptera: Lasiocampidae). Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):899-903. DOI 10.18699/VJ16.208

ORIGINAL ARTICLE

Received 03.11.2016

Accepted for publication 16.12.2016

© AUTHORS, 2016



e-mail: paulee@bionet.nsc.ru

Сибирский шелкопряд (*Dendrolimus superans sibiricus* Tschetverikov, 1908) – аборигенный обитатель хвойных лесов Северной Азии. Ареал этого вида простирается по лесной зоне от берегов Тихого океана до востока европейской части РФ (Gninenko, Orlinskii, 2002; Золотухин, 2015). Южная граница ареала проходит с запада примерно по 55-й параллели до рек Иртыш и Обь, отсюда вид распространен по всей лесной и лесостепной зоне Северо-Восточного Казахстана, Северной Монголии, северо-западных и северо-восточных провинций Китая, включая территорию Джунгарского Алатау, Внутренней Монголии (Nei Mongol), Хэбэя (Hebei) и Шаньдуна (Shandong), а также на Корейском полуострове. Номинативный подвид *Dendrolimus superans superans* (Butler, 1881) распространен на юге Сахалина (на севере острова обитает *D. s. sibiricus*), Южных Курилах и в Японии до о. Хонсю включительно (Золотухин, 2015). Недавно сибирскому подвиду был присвоен видовой статус (Mikkola, Stahls, 2008) на основании молекулярных исследований ядерных и митохондриальных генов. Тем не менее из-за отсутствия каких-либо морфологических различий В.В. Золотухин (2015) считает *D. s. sibiricus* и *D. s. superans* не более чем территориально разобщенными подвидами одного вида.

Кормовой базой сибирского шелкопряда обычно служат различные виды деревьев семейства Pinaceae, прежде всего лиственница, которая наиболее сильно повреждается насекомым. Также гусеницы питаются хвоей разных видов сосен (включая *Pinus sibirica* и *P. koraiensis*), пихт и елей (Рожков, 1963). По вредоносному воздействию на массивы лиственных и кедровых лесов Сибири сибирский шелкопряд занимает первое место. Вспышки численности этого вида приводят к дефолиации хвойных лесов и нередко к их гибели на обширных площадях, что наносит колоссальный экономический урон. По данным современных учетов, на территории России площадь усохших в результате размножения сибирского шелкопряда лесов составляет более 10 млн га (Рожков, 1965; Кондаков 2002; Колтунов, Ермаков, 2013). Обычно гусеницы сибирского шелкопряда зимуют два раза, но на севере ареала – три раза, на юге – один раз (Рожков, 1963), а номинативный подвид в Японии дает два поколения за год (Золотухин, 2015).

Вспышки численности сибирского шелкопряда могут быть непродолжительными, всего один-три сезона, после чего наступает сильная депрессия популяции вследствие стремительно возросшего груза насекомых-паразитов, вирусов и патогенных микроорганизмов. Однако даже одного сезона активного действия сибирского шелкопряда может быть достаточно для гибели леса, поскольку ослабленные деревья в последующие годы становятся восприимчивыми к пагубному влиянию стволовых вредителей, которые окончательно губят растение и приводят древесину в полную непригодность для хозяйственного использования.

Возрастание численности сибирского шелкопряда связано со стечением разных факторов. Например, весенние пожары лесной подстилки уничтожают одного из главных регуляторов численности сибирского шелкопряда – яйцееда рода *Telenomus* (Гродницкий, 2004). Иногда после

теплых сезонов гусеницы успевают развиваться за один год вместо двух лет (Гродницкий, 2004), что приводит к увеличению эффективного размера популяции. Засушливые сезоны снижают нагрузку грибного энтомопатогена *Beauveria bassiana* на сибирского шелкопряда, поскольку для гриба высокая влажность является оптимальным условием развития (Huaifeng et al., 1998).

В настоящей работе мы обратились к вопросу о присутствии непатогенного микроорганизма *Wolbachia* в популяциях сибирского шелкопряда. *Wolbachia* – бактериальный внутриклеточный симбионт членистоногих и нематод, который является факультативным компонентом генома макросимбионта, поскольку передается от особи к особи строго по материнской линии параллельно с митохондриями. Этот симбионт глубоко интегрирован в биологию своих хозяев. Некоторым видам-хозяевам *Wolbachia* обеспечивает метаболическую комплементацию, тем самым повышая конкурентоспособность инфицированных особей. *Wolbachia* влияет на репродуктивную биологию макросимбионтов, в частности может как увеличивать, так и сокращать репродуктивный потенциал особей в популяции. На популяционном уровне присутствие *Wolbachia* может оказывать влияние на генетическое разнообразие вида (Werren et al., 2008; Zabal-Aguirre et al., 2010).

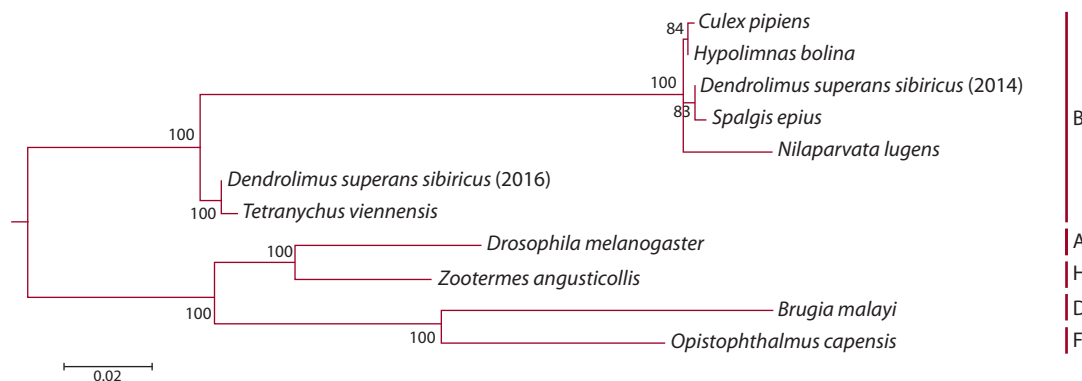
Материалы и методы

Коллекция насекомых. Насекомые собраны в 2014 ($n = 9$) и 2016 гг. ($n = 32$) на территории Большехежирского заповедника (с. Бычиха, 48°17.9' с. ш., 134°49.3' в. д. и ниже течение руч. Соснинский, 48°16.2' с. ш., 134°46.3' в. д.) в окрестностях г. Хабаровска и хранились на энтомологических матрасиках при комнатной температуре до выделения ДНК. Вся выборка, за исключением одной особи, представлена самцами.

Выделение ДНК и экспериментальные процедуры. Из каждой особи выделяли ДНК отдельно из брюшка и ноги. Это было сделано для того, чтобы, во-первых, оценить сохранность ДНК в музейных образцах и таким образом дать рекомендации по выделению ДНК, и, во-вторых, сопоставить представленность *Wolbachia* в соматических и репродуктивных тканях, т. е. определить релевантность использования конечностей для выявления *Wolbachia* у сибирского шелкопряда.

Процедура выделения ДНК включала инкубацию образца в экстрагирующем буфере и последующую ее очистку (Илинский и др., 2013). Пригодность полученной тотальной ДНК проверяли посредством ПЦР с праймерами к гену насекомых 28S *rRNA* (Choudhury, Werren, 2006). Статус инфицированности устанавливали амплификацией с *Wolbachia*-специфичными праймерами к бактериальному гену 16S *rRNA* (Werren, Windsor, 2000). Нарботку ампликонов генов мультилокусного типирования и их секвенирование проводили согласно рекомендациям базы данных PubMLST (pubmlst.org/wolbachia) и (Baldo et al., 2006).

Для реконструкции филогенетических взаимоотношений выявленных в данной работе штаммов *Wolbachia* использовали конкатенированные последовательности локусов *ftsZ* и *fbpA* *Wolbachia* (864 п. н.). Кроме того, из базы данных PubMLST были извлечены последователь-



Phylogenetic relationships among *Wolbachia* strains reconstructed by the maximum likelihood method based on concatenated sequences of the *ftsZ* and *fbpA* loci (864 bp.).

Bootstrap values exceeding 75 are indicated above the corresponding branches. *Wolbachia* host species and supergroups are shown.

ности аллелей этих генов от других штаммов *Wolbachia*, наиболее близких к обнаруженным нами изолятам симбионта. Нуклеотидные последовательности выбранных генов были выравнены друг относительно друга с помощью алгоритма Muscle (Edgar, 2004) на онлайн-сервисе <http://www.ebi.ac.uk/Tools/msa/muscle/>. Филогенетический анализ проводили с использованием метода максимального правдоподобия и с применением Тамура-трехпараметрической модели нуклеотидных замен, входящих в пакет программ MEGA 6 (Tamura et al., 2013). Достоверность ветвления оценивалась посредством бутстреп-анализа, который состоял из 1000 итераций.

Результаты

Качество ДНК, выделенной из музейных образцов сибирского шелкопряда, как правило, было приемлемо для поиска симбиотической бактерии *Wolbachia* методом ПЦР. Ампликоны ядерного гена 28S *rRNA* были получены для 76 из 82 образцов ДНК, из них 36 образцов были выделены из брюшка и 40 – из ног. Для одной особи *D. s. sibiricus* из сборов 2016 г. не удалось получить положительный сигнал на 28S *rRNA* ни для ДНК, выделенной из ноги, ни для ДНК, выделенной из брюшка. Этот образец был исключен из анализа. Таким образом, выделение ДНК из ног более предпочтительно по сравнению с выделением ДНК из брюшка, что может быть связано с большей сохранностью ДНК в ногах (из-за их быстрого высыхания), а также с тем, что в брюшке содержится жировое тело, которое может быть источником загрязнений препарата ДНК, в том числе таких, которые способны ингибировать ПЦР.

Эндосимбиотическая бактерия *Wolbachia* обнаружена в популяционных сборах как 2014, так и 2016 г. В большинстве случаев результаты поиска эндосимбионта *Wolbachia* для препаратов ДНК, полученных из ног и брюшка, совпадали. Для тех образцов, где было найдено расхождение, нами проведена дополнительная проверка статуса инфицированности, а именно ПЦР с праймерами к локусам бактерии *fbpA* и *ftsZ*. Так, у 10 особей (27 %) бактерия выявлена только в экстрактах ДНК из брюшка, а у одной особи (2,7 %) – только в ноге. На отсутствие *Wolbachia* в соматических тканях и на мозаичное в целом

распределение симбионта в тканях организма указывалось и ранее (Dobson et al., 1999), поэтому отсутствие *Wolbachia* в ногах было вполне ожидаемым. Однако случай, когда не удалось выявить ДНК бактерии в брюшке, требует отдельных пояснений. Данный образец ДНК выделен из самца, а, как известно, самцы не участвуют в передаче *Wolbachia* следующему поколению. Вероятно, распределение бактерий в тканях организма самцов и самок одного вида различается, и репродуктивные органы этой конкретной особи не содержали или утратили *Wolbachia*. Еще одним объяснением данного факта может служить крайне низкая (в силу плохой сохранности внутренних органов образца) доля ДНК из репродуктивных тканей, что не позволило обнаружить *Wolbachia*. В целом можно заключить, что для работы с ДНК насекомого, выделяемой из воздушно-сухих музейных образцов, предпочтительно брать ногу, а для работы с ДНК матерински-наследуемых симбионтов желательно использовать брюшко, при этом образцы ДНК, полученные из ног, могут быть использованы как дополнительный материал для верификации результатов работы.

Частота инфицированности в выборке сибирского шелкопряда в 2014 г. составила 100 %, доверительный интервал для анализируемой популяции был 66.4–100 % ($p = 0.95$). В 2016 г. эти значения составили 90.3 ± 5.3 %, доверительный интервал 74.2–98.0 % ($p = 0.95$). Таким образом, частота встречаемости *Wolbachia* в популяции *D. s. sibiricus* Большехецирского заповедника находится на высоком уровне и составляет не менее 66 %.

Нами проведено типирование симбиотической бактерии *Wolbachia* для двух образцов, один из которых принадлежал выборке 2014 г., а второй – 2016 г. В результате обнаружено, что в популяциях сибирского шелкопряда присутствуют как минимум два генетических варианта *Wolbachia* (см. рисунок).

Выявленные аллельные варианты, согласно базе данных PubMLST, входят в состав нескольких гаплотипических профилей *Wolbachia*, которые принадлежат супергруппе В *Wolbachia*. Отметим, что штаммы, относимые к этой супергруппе, наиболее часто встречаются среди чешуекрылых насекомых (Salunke et al., 2012).

Обсуждение

Эндосимбионт *Wolbachia* широко распространен у дневных чешуекрылых (Ahmed et al., 2015, 2016). Сведения по другим семействам фрагментарны, большинство из них, особенно из группы Microlepidoptera, вовсе не изучены в отношении симбиоза с *Wolbachia*. Среди представителей семейства коконопрядов (Lasiocampidae) к настоящему времени известен только один инфицированный вид – *Dendrolimus spectabilis* (Tagami, Miura, 2004). Таким образом, в нашей работе впервые показано присутствие симбионта *Wolbachia* в аборигенных популяциях сибирского шелкопряда.

Выявленный нами высокий уровень инфицированности популяции *D. s. sibiricus*, по всей видимости, может указывать на то, что инфицированным особям благоприятствует отбор и/или что *Wolbachia* осуществляет манипуляцию репродукцией вида-хозяина, тем самым увеличивая численность инфицированных особей в популяции.

Известно, что *Wolbachia* может выступать облигатным симбионтом. Так, филарийные нематоды не способны нормально развиваться и размножаться без *Wolbachia* (Sung-gradit, Nuchprayoon, 2010; McCall et al., 2014). У членистоногих облигатность *Wolbachia* описана для паразитической осы *Asobara tabida* (Dedeine et al., 2001; Kremer et al., 2009), сеноеда *Liposcelis tricolor* (Dong et al., 2007) и клопа *Cimex lectularius* (Hosokawa et al., 2010). У *Drosophila melanogaster* симбионт увеличивает продолжительность жизни мух в стрессовых условиях (Harcombe, Hoffmann, 2004; Вайсман и др., 2009), восстанавливает фертильность мутантных мух (Starr, Cline, 2002) и определяет ряд других эффектов, повышающих приспособленность. Инфицированные *Wolbachia* самки комара *Aedes albopictus* оказываются более плодовитыми и имеют большую продолжительность жизни, чем неинфицированные особи (Dobson et al., 2002).

Репродуктивные манипуляции выражаются в форме таких феноменов, как цитоплазматическая несовместимость, андроцид, феминизация или телитокный партеногенез, и все указанные явления известны для чешуекрылых насекомых (Salunkhe et al., 2014). Последние три феномена характеризуются сдвигом соотношения полов в сторону преобладания самок. Мы не можем сделать заключение о соотношении полов для *D. s. sibiricus*, поскольку самцы этого вида летают значительно лучше, чем самки, вследствие чего самцы преобладают в сборах. Однако такую проверку возможно осуществить в будущем, учитывая, например, соотношения полов при лабораторном выведении бабочек из коконов.

В настоящее время мы не располагаем какими-либо сведениями о роли *Wolbachia* в популяции сибирского шелкопряда. Однако в свете накопленных данных о биологии других симбиотических ассоциаций вольбахия–хозяин представляется возможным обсудить гипотетические сценарии ассоциации вольбахия–сибирский шелкопряд и указать направления дальнейшего изучения этого симбиоза. Например, в случае, если для сибирского шелкопряда характерно проявление высокого уровня цитоплазматической несовместимости, крайне важное значение имеет мониторинг частот конкретных генетических вариантов *Wolbachia*. Сопоставимые частоты двух и более вариантов

Wolbachia в популяции в этом случае будут указывать на явление двунаправленной несовместимости, что выражается для вида в сокращении репродуктивного потенциала популяции. Сдвиг частот в сторону одного из генетических вариантов симбионта снижает вероятность несовместимых скрещиваний, что может привести к популяционному взрыву. Исключительно важными аспектами вовлеченности *Wolbachia* в биологию сибирского шелкопряда могут оказаться мутуалистические эффекты симбионта. Сибирский шелкопряд – аборигенный вид, который постоянно находится под грузом разнообразных паразитоидов и инфекций. Тем не менее регулярно в разных частях ареала его численность стремительно возрастает (Колтунов, Ермаков, 2013). В связи с этим необходимо учитывать, что матерински наследуемые симбионты, включая *Wolbachia*, в некоторых случаях способны защищать своих хозяев от вирусных инфекций (Teixeira et al., 2008; Osborne et al., 2009) и даже паразитоидов (van Nouhuys et al., 2016).

Род *Dendrolimus* включает около 30 палеарктических и северных индомалайских видов (Золотухин, 2015). Некоторые из них, например *D. pini* (Linnaeus, 1758), также являются вредителями лесных массивов. Информация о распространенности *Wolbachia* среди этих видов может способствовать изучению биологии симбиотической ассоциации *D. s. sibiricus*–*Wolbachia* и помочь в решении проблемы феномена взрыва численности популяций сибирского шелкопряда.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research, project 16-04-00980, and State Budgeted Project 0324-2015-0004.

Conflict of interest

The authors declare no conflict of interest.

References

- Ahmed M.Z., Araujo-Jnr E.V., Welch J.J., Kawahara A.Y. *Wolbachia* in butterflies and moths: geographic structure in infection frequency. *Frontiers Zoology*. 2015;12(1):1.
- Ahmed M.Z., Breinholt J.W., Kawahara A.Y. Evidence for common horizontal transmission of *Wolbachia* among butterflies and moths. *BMC Evol. Biol.* 2016;16(1):1.
- Baldo L., Dunning Hotopp J.C., Jolley K.A., Bordenstein S.R., Biber S.A., Choudhury R.R., Hayashi C., Maiden M.C., Tettelin H., Werren J.H. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl. Environ. Microbiol.* 2006;72(11):7098-7110.
- Choudhury R., Werren J.H. Unpublished primers. 2006. Available at <http://troi.cc.rochester.edu/~wob/FIBR/downloads.html#protocols>.
- Dedeine F., Vavre F., Fleury F., Loppin B., Hochberg M.E., Bouletreau M. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proc. Natl. Acad. Sci.* 2001;98(11):6247-6252.
- Dobson S.L., Bourtzis K., Braig H.R., Jones B.F., Zhou W., Rousset F., O'Neill S.L. *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biology*. 1999; 29(2):153-160.
- Dobson S.L., Marsland E.J., Rattanadechakul W. Mutualistic *Wolbachia* infection in *Aedes albopictus*: Accelerating cytoplasmic drive. *Genetics*. 2002;160:1087-1094.
- Dong P., Wang J.-J., Hu F., Jia F.-X. Influence of *Wolbachia* infection on the fitness of the stored-product pest *Liposcelis tricolor* (Psocoptera: Liposcelidae). *J. Econ. Entomol.* 2007;100(4):1476-1481.

- Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 2004;32(5):1792-1797.
- Gninenko Y.I., Orlinskii A.D. *Dendrolimus sibiricus* in the coniferous forests of European Russia at the beginning of the twenty-first century. *Eppo Bull.* 2002;32(3):481-483.
- Grodnitskiy D.L. Siberian silk moth and the fate of fir taiga. *Priroda = Nature.* 2004;11:49-56. (in Russian)
- Harcombe W., Hoffmann A.A. *Wolbachia* effects in *Drosophila melanogaster*: in search of fitness benefits. *J. Invertebrate Pathology.* 2004;87:45-50.
- Hosokawa T., Koga R., Kikuchi Y., Meng X.-Y., Fukatsu T. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc. Natl. Acad. Sci.* 2010;107(2):769-774.
- Huafeng L.F., Meizhen L.Z., Cui H. Pathogenic effect of *Beauveria bassiana* infected on *Dendrolimus punctatus* under different temperature and humidity. *Chin. J. Appl. Ecol.* 1998;9:195-200.
- Ilinskiy Yu.Yu., Yudina M.A., Kalmykova Ye.A., Bykov R.A., Vysochina N.P., Vinarskaya N.P., Zakharov I.K. *Wolbachia* infection among fleas (Siphonaptera: Insecta) of Sverdlovsk region and Khabarovsk territory. *Ekologicheskaya genetika = Ecological genetics.* 2013;11(1):32-35. (in Russian)
- Koltunov E.V., Erdakov L.N. Features of the long-term cyclic occurrence of outbreaks of different Siberian moth (*Dendrolimus superans sibiricus* Tschetv.) populations in Siberia. *Sovremennye problemy nauki i obrazovaniya = Modern Problems of Science and Education.* 2013;6. (in Russian)
- Kondakov Yu.P. Massovye razmnzheniya sibirskogo shelkopryada v lesah Krasnojarskogo kraja [Population booms of Siberian silk moth in the forests of the Krasnoyarsk Krai]. *Entomologicheskie issledovaniya v Sibiri [Entomological Studies in Siberia].* Krasnoyarsk, 2002;2:25-74. (in Russian)
- Kremer N., Voronin D., Charif D., Mavingui P., Mollereau B., Vavre F. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. *PLoS Pathog.* 2009;5(10):e1000630.
- McCall J.W., Kramer L., Genchi C., Guerrero J., Dzimiński M.T., Mansour A., McCall S.D., Carson B. Effects of doxycycline on heartworm embryogenesis, transmission, circulating microfilaria, and adult worms in microfilaremic dogs. *Vet. Parasitol.* 2014;206(1): 5-13.
- Mikkola K., Stahls G. Morphological and molecular taxonomy of *Dendrolimus sibiricus* Chetverikov stat. rev. and allied lappet moths (Lepidoptera: Lasiocampidae), with description of a new species. *Entomol. Fennica.* 2008;19(2):65.
- Osborne S.E., Leong Y.S., O'Neill S.L., Johnson K.N. Variation in antiviral protection mediated by different *Wolbachia* strains in *Drosophila simulans*. *PLoS Pathog.* 2009;5(11):1-9.
- Rozhkov A.S. Sibirskiy shelkopryad [Siberian silk moth]. Moscow, AN SSSR Publ., 1963. (in Russian)
- Rozhkov A.S. Massovoe razmnzhenie sibirskogo shelkopryada i mery borby s nim [Population booms of Siberian silk moth and methods for controlling them]. Moscow, Nauka Publ., 1965. (in Russian)
- Salunke B.K., Salunke R.C., Dhotre D.P., Walujkar S.A., Khandagale A.B., Chaudhari R., Chandode R.K., Ghatge H.V., Patole M.S., Werren J.H., Shouche Y.S. Determination of *Wolbachia* diversity in butterflies from Western Ghats, India, by a multigene approach. *Appl. Environ. Microbiol.* 2012;78:4458-4467.
- Salunke R.C., Narkhede K.P., Shouche Y.S. Distribution and evolutionary impact of *Wolbachia* on butterfly hosts. *Indian J. Microbiol.* 2014;54(3):249-254.
- Starr D.J., Cline T.W. A host parasite interaction rescues *Drosophila* oogenesis defects. *Nature.* 2002;418:76-79.
- Sungratit S., Nuchprayoon S. *Wolbachia* of arthropods and filarial nematodes: biology and applications. *Chula Med. J.* 2010;54:605-621.
- Tagami Y., Miura K. Distribution and prevalence of *Wolbachia* in Japanese populations of Lepidoptera. *Insect Mol. Biol.* 2004;13(4): 359-364.
- Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 2013;30:2725-2729.
- Teixeira L., Ferreira A., Ashburner M. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* 2008;6(12):2753-2763.
- van Nouhuys S., Kohonen M., Duploup A. *Wolbachia* increases the susceptibility of a parasitoid wasp to hyperparasitism. *J. Exp. Biol.* 2016;219(19):2984-2990.
- Weisman N.Ya., Ilinskiy Yu.Yu., Golubovskiy M.D. Population genetic analysis of *Drosophila melanogaster* lifespan: similar effects of *Wolbachia* endosymbiont and tumor suppressor lgl under temperature stress. *Zhurnal obshchey biologii = Journal of General Biology.* 2009;70(5):425-434. (in Russian)
- Werren J.H., Windsor D. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc. Roy. Soc. Lond.* 2000;267: 1277-1285.
- Werren J.H., Baldo L., Clark M.E. *Wolbachia*: master manipulators of invertebrate biology. *Nature Rev. Microbiol.* 2008;6(10):741-751.
- Zabal-Aguirre M., Arroyo F., Bella J.L. Distribution of *Wolbachia* infection in *Chorthippus parallelus* populations within and beyond a Pyrenean hybrid zone. *Heredity.* 2010;104(2):174-184.
- Zolotukhin V.V. Kokonopryady (Lepidoptera: Lasiocampidae) fauny Rossii i sopredelnykh territoriy [Lasiocampidae (Lepidoptera) of Russia and adjacent territories]. Ulyanovsk, Korporatsiya tekhnologii prodvizheniya Publ., 2015. (in Russian)

База данных генов, повышающих устойчивость пшеницы и родственных ей злаков к патогенным грибам

О.Г. Смирнова , А.В. Кочетов

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия
Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Растения подвергаются воздействию огромного количества патогенных грибов. В последние годы идентифицировано большое количество генов, принимающих участие в формировании защитного ответа у пшеницы, и проведено секвенирование некоторых генов устойчивости. Секвенирование аллельных вариантов генов устойчивости у пшеницы и ее сородичей необходимо для понимания молекулярных механизмов формирования защитного ответа. В настоящей работе описывается информационный ресурс, предназначенный для сбора и хранения информации о генах, обеспечивающих устойчивость пшеницы и родственных ей злаков к заболеваниям, вызванным грибными патогенами. База PRG (Pathogenesis-Related Genes) содержит структурированную информацию о регуляции активности генов устойчивости. В базе представлены нуклеотидные последовательности генов устойчивости, а также данные об отдельных однонуклеотидных полиморфизмах генов, ассоциированных с различными уровнями устойчивости к заболеваниям. Собрана информация о продукте гена и его функциях. Суммированы данные о патогене и заболевании, хромосомной локализации гена устойчивости, даны ссылки на карточки в сопутствующих базах данных нуклеотидных последовательностей (GenBank) и белков (UniProt). Приводится важная информация об изменении экспрессии гена в ответ на действие патогенов, гормонов, а также на изменения условий окружающей среды. Информация вносится в базу в результате аннотирования научных публикаций. В настоящее время в базе данных содержится информация о 66 генах, представленных 75 аллельными вариантами. База PRG разработана на платформе SRS (Sequence Retrieval System). Средства SRS позволяют проводить множественный поиск по полям базы, что необходимо для эффективного построения запросов. Система SRS автоматически генерирует Web-интерфейс с поисковыми таблицами и визуализирует результаты поиска, которые могут быть сохранены в текстовом формате. Собранная информация и удобный интерфейс позволят широкому кругу исследователей расширить представления о разнообразии генов, принимающих участие в формировании защитного ответа, способствуя тем самым эффективному созданию форм растений с повышенной устойчивостью к заболеваниям. База PRG доступна по адресу <http://srs6.bionet.nsc.ru/srs6bin/cgi-bin/wgetz?-page+top+-newld>.

Ключевые слова: база данных; информационный ресурс; гены устойчивости; патогенные грибы; зерновые культуры.

A database on genes increasing resistance of wheat and relative species against pathogenic fungi

O.G. Smirnova , A.V. Kochetov

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia

Fungi belong to the major plant pathogens and investigation of plant resistance genes is a quite important task. During the last years many wheat resistance genes were identified. However, the sequencing of the *Triticum aestivum* L. genome is still going on and the nucleotide sequences of most resistance genes are not yet known. In addition, the study of allelic variants of resistance genes is important for better understanding of the molecular mechanisms of their action. In this paper we present an information resource for accumulation of data on sequenced genes of wheat and its relatives providing resistance against diseases caused by fungal pathogens. The database (Pathogenesis-Related Genes, PRG) contains information on gene chromosomal localizations and functional activities, nucleotide sequences and single nucleotide polymorphisms associated with their effects. PRG provides data on the proteins encoded, pathogens and diseases, as well as on the resistance gene expression patterns in response to pathogen inoculations, exposure to hormones and various external stimuli. It also has cross-references with related entries from the databases on nucleotide sequences (GenBank) and proteins (UniProt). Information entered into the database is a result of the annotation of scientific publications and manual curation. Currently PRG compiles data on 75 allelic variants of 66 resistance genes. The PRG database was developed on the basis of the SRS (Sequence Retrieval System) platform. This system allows the use of complex queries and visualization tools and automatically generates www-interface with the information in table or text formats. PRG may be useful for researchers studying plant biology or breeding new plant cultivars resistant to fungal diseases. It is available at the address: <http://srs6.bionet.nsc.ru/srs6bin/cgi-bin/wgetz?-page+top+-newld>.

Key words: databases; resistance gene; pathogenic fungus; crops.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Смирнова О.Г., Кочетов А.В. База данных генов, повышающих устойчивость пшеницы и родственных ей злаков к патогенным грибам. Вавиловский журнал генетики и селекции. 2016;20(6):904-908. DOI 10.18699/VJ16.203

HOW TO CITE THIS ARTICLE:

Smirnova O.G., Kochetov A.V. A database on genes increasing resistance of wheat and relative species against pathogenic fungi. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):904-908. DOI 10.18699/VJ16.203

REVIEW

Received 07.10.2016 г.

Accepted for publication 29.11.2016 г.

© AUTHORS, 2016

 e-mail: planta@bionet.nsc.ru

Увеличение производства зерна мягкой пшеницы *Triticum aestivum* L. может осуществляться не только за счет роста урожайности, но и за счет снижения потерь, связанных с заболеваниями, вызываемыми грибными патогенами. Потери урожая при поражении листовой ржавчиной могут достигать 30–50 % (McIntosh et al., 1995). Создание устойчивых сортов является сложной задачей. Механизмы защиты растений связаны с функционированием генов устойчивости (R-генов – resistance genes) и их аналогов, обеспечивающих в первую очередь узнавание молекул патогена по присутствию в составе R-белков консервативных аминокислотных доменов. Среди изученных R-генов наиболее многочисленный класс представляют гены устойчивости, кодирующие белки, содержащие NBS (Nucleotide Binding Site) и LRR (Leucine-Rich Repeat) домены (Sanseverino et al., 2013). Тем не менее существенные компоненты защитной системы растений остаются еще не изученными. Более того, быстрая эволюционная изменчивость патогенов в некоторых случаях опережает усилия специалистов по защите растений и приводит к неэффективности расоспецифических генов устойчивости (Singh et al., 2005). Отдельный интерес представляют гены, формирующие неспецифическую устойчивость, поскольку она обеспечивает длительный период эффективности сразу к нескольким патогенам, что делает ее предпочтительной для селекционеров и исследователей. У пшеницы подробно охарактеризованы всего четыре гена *Lr34*, *Lr46*, *Lr67* и *Lr27*, определяющих замедленное развитие сразу нескольких заболеваний, два из которых, *Lr34* и *Lr67*, были секвенированы (Rosewarne et al., 2006; Mago et al., 2011; Spielmeyer et al., 2013).

К настоящему времени описано более 70 генов *Lr* (leaf rust), контролирующих устойчивость пшеницы к листовой ржавчине и расположенных в разных хромосомах (McIntosh et al., 2013), и только четыре из них секвенировано. В связи с отсутствием данных о первичной структуре преобладающего большинства генов устойчивости к листовой ржавчине существует ряд проблем, связанных с маркированием этих генов. На сегодняшний день наиболее распространены маркеры, полученные исходя из информации о локализации генов на хромосомах. Эффективность таких маркеров снижается из-за возможных процессов рекомбинации между маркером и геном. Поэтому наиболее функциональными являются маркеры, разработанные на основании секвенирования соответствующих нуклеотидных последовательностей ДНК. Таким образом, для эффективной защиты пшеницы от патогенных грибов необходимо вести поиск новых генов устойчивости, используя данные по генам устойчивости у наиболее изученных в этом направлении видов растений и доступные сведения по структуре геномных нуклеотидных последовательностей пшеницы.

Компьютерные базы данных активно применяются для информационного обеспечения экспериментальных исследований, а также для поиска новых знаний с помощью методов компьютерной системной биологии. Разработано большое число целевых баз данных, без использования которых планирование экспериментов и интерпретация полученных результатов затруднительны. В области изучения генов устойчивости растений создана база Plant

Resistance Genes (<http://prgdb.org>), в которой приведены нуклеотидные последовательности 112 известных и более чем 100 000 предполагаемых R-генов, присутствующих у 233 видов растений (Sanseverino et al., 2013). Также следует отметить базу данных PathoPlant (Bülow et al., 2007), где аннотирована информация о генах *Arabidopsis thaliana*, экспрессия которых усиливается в ответ на инокуляцию различными патогенами.

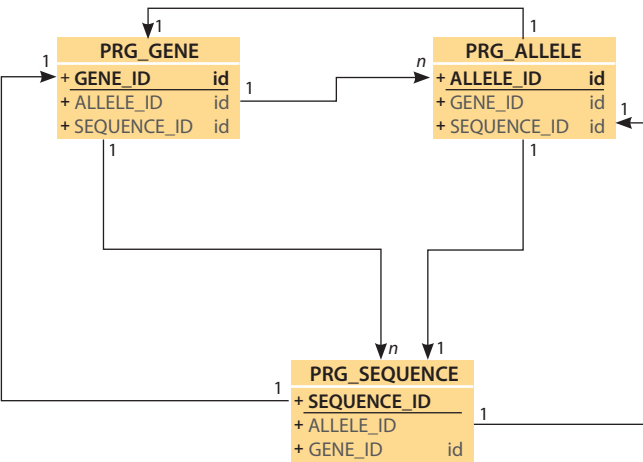
Нами создана целевая база данных генов (и аллелей генов) пшеницы, контролирующих устойчивость к фитопатогенным грибам. Было решено сконцентрироваться на аннотации генов устойчивости пшеницы и родственных ей злаков, поскольку пшеница является важной зерновой культурой и многие гены устойчивости были интрогрессированы в мягкую пшеницу из дикорастущих видов. В базу также внесена информация о генах, экспрессия которых изменяется после действия патогена. Продукты этих генов принимают участие в передаче сигнала и формировании защитного ответа, возникающего в клетке после контакта с патогеном. Концентрация такой информации в базе данных позволяет использовать ее в качестве базы знаний для анализа механизмов формирования защитного ответа пшеницы к действию патогенных грибов. В базу вносятся информация, полученная при анализе научных публикаций. Структура базы данных PRG, созданной на платформе SRS (Sequence Retrieval System), позволяет пользователю составлять сложные запросы и получать комплексную информацию, необходимую для разработки собственных маркеров при картировании аллелей генов устойчивости разной специфичности. Такая информация актуальна для планирования селекционно-генетических экспериментов при создании устойчивых сортов пшеницы в связи с появлением новых рас патогенных грибов и необходимостью постоянной работы по обновлению существующих сортов сельскохозяйственных растений.

Логическая структура и формат базы данных PRG

База данных PRG построена на платформе SRS и включает в себя три составляющие: таблицу генов PRG_GENE, таблицу аллелей PRG_ALLELE и таблицу последовательностей PRG_SEQUENCE, взаимосвязи между которыми устанавливаются согласно схеме (см. рисунок).

Таблица генов PRG_GENE включает информацию о названии организма, гена, локализации гена на хромосоме, патогене, заболевании и состоит из 14 полей. Таблица аллелей PRG_ALLELE содержит информацию о чувствительности к патогену, связанной с аллельным состоянием гена, и состоит из 12 полей. Таблица последовательностей PRG_SEQUENCE содержит информацию о последовательностях аллельных вариантов генов, описанных в таблице PRG_ALLELE, и состоит из 7 полей. Подробная информация о составе полей базы и их содержании приведена в табл. 1–3. Примеры заполнения полей представлены в табл. 4 и 5.

Все поля в таблицах PRG_GENE и PRG_ALLELE, за исключением поля END, являются индексируемыми, т. е. пользователь может осуществлять поиск информации по этим полям.



Schematic representation of the PRG database. Cross-indexed fields of the PRG_GENE, PRG_ALLELE, and PRG_SEQUENCE tables are shown.

Table 1. Fields of the PRG_GENE table

Field name	Description
GENE_ID	Gene identifier
ALLELE_ID	Cross-reference to ALLELE_ID in the PRG_ALLELE table
SEQUENCE_ID	Cross-reference to SEQUENCE_ID in the PRG_SEQUENCE table
DISEASE	Disease
PATHOGEN	Causative agent of the disease
HOST_SPECIES	The common and Latin names of the species
GENE	Gene names
MAPPING	Gene mapping on the chromosome
PRODUCT	Gene product
COMMENT	Comment
KEYWORDS	Keywords
REFERENCE	Bibliographic reference
PUBMED	Cross-reference to PUBMED
END	End of the entry

Технологии реализации
и организация ведения базы PRG

База PRG управляется средствами системы SRS 6.1.3, которая развернута на сервере баз данных под управлением CentOS Linux 6.6. Система SRS специально разработана для формализованного описания биологических данных (Zdobnov et al., 2002). Средства SRS позволяют проводить множественный поиск по полям базы, что необходимо для эффективного построения запросов пользователями. Система SRS автоматически генерирует Web-интерфейс с поисковыми таблицами и визуализирует результаты поиска, которые могут быть сохранены в текстовом формате. Средства SRS позволяют индексировать большинство информационных полей и эффективно осуществлять

Table 2. Fields of the PRG_ALLELE table

Field name	Description
ALLELE_ID	Allele identifier
SEQUENCE_ID	Cross-reference to SEQUENCE_ID in the PRG_SEQUENCE table
GENE_ID	Cross-reference to GENE_ID in the PRG_GENE table
CULTIVAR	Variety or mutant strain
RESISTANCE	Disease resistance
GENBANK	Cross-reference to GenBank
UNIPROT	Cross-reference to UniProt
COMMENT	Comment
KEYWORDS	Keywords
REFERENCE	Bibliographic reference
PUBMED	Cross-reference to PUBMED
END	End of the entry

Table 3. Fields in the PRG_SEQUENCE section

Field name	Description
SEQUENCE_ID	Sequence identifier
ALLELE_ID	Cross-reference to ALLELE_ID in the PRG_ALLELE table
GENE_ID	Cross-reference to GENE_ID in the PRG_GENE table
LOCALIZATION	Gene positions according to the GenBank notation
COMMENT	Comment
SEQUENCE	Sequence
END	End of the entry

перекрестную связь полей в таблицах баз данных, что необходимо для построения эффективных пользовательских запросов и свободной навигации между полями и записями в различных таблицах. Эта система также автоматически генерирует Web-интерфейс для обеспечения поиска и визуализации информации в базе данных (формы запроса, визуализация данных, гиперссылки на документы в базе и интернет-ресурсы, настройки способа визуализации). Эксперт-аннотатор заполняет содержимое полей в записи базы данных на основании анализа научной литературы. Данные представляются для каждой таблицы отдельно в виде текстового файла, как показано в табл. 1–3. Файлы с информацией передаются администратору системы, который вводит их согласно инструкции системы SRS. Структура базы данных позволяет расширять список аннотируемых генов по мере появления новых литературных данных.

Пользовательский интерфейс

Модуль интерфейса является программным компонентом и служит для взаимодействия пользователя с базой данных. Модуль интерфейса для базы данных PRG обеспе-

Table 4. An exemplary entry in the PRG_GENE table

Field name	Description
GENE_ID	TaLr34
ALLELE_ID	TaLr34-CS TaLr34-Renan
SEQUENCE_ID	TaLr34-CS_S TaLr34-Renan_S
DISEASE	Leaf rust, stripe rust, powdery mildew, stem rust
PATHOGEN	<i>Puccinia triticina</i> , <i>Puccinia striiformis</i> (tritici), <i>Blumeria graminis</i> (tritici), <i>Puccinia graminis</i>
HOST_SPECIES	<i>Triticum aestivum</i> (bread wheat)
GENE	Lr34, Lr34 allele D, Lr34-D, Yr18, Sr57, Pm38, Sb1, Ltn1
MAPPING	Chromosome 7DS
PRODUCT	LR34 protein, ABC transporter, ATF-binding cassette transporter, adenosine triphosphate-binding cassette transporter
COMMENT	Alleles of <i>Lr34</i> conferring resistance or susceptibility differ by three genetic polymorphisms. The <i>Lr34</i> gene, which functions in the adult plant, stimulates senescence-like processes in the flag leaf tips and edges. The <i>Lr34</i> gene, then called <i>Yr18</i> , also confers moderate resistance to stripe or yellow rust (<i>Puccinia striiformis</i>) and powdery mildew (<i>Blumeria graminis</i>), where it is called <i>Pm38</i> . The locus is further associated with tolerance to stem rust (<i>Puccinia graminis</i>) and barley yellow dwarf virus, Bdv1. Flag leaves of many wheat cultivars containing <i>Lr34</i> in certain environments develop a necrotic leaf tip. This morphological marker is referred to as leaf tip necrosis
KEYWORDS	Multiple resistance, leaf rust, stripe rust, powdery mildew, chromosome 7DS, Lr34
REFERENCE	Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science. 2009, 323(5919):1360-3
PUBMED	19229000
END	

Table 5. An exemplary entry in the PRG_ALLELE table

Field name	Description
ALLELE_ID	TaLr34-CS
SEQUENCE_ID	TaLr34-CS_S
GENE_ID	TaLr34
CULTIVAR	Chinese Spring
RESISTANCE	Resistant (leaf rust)
GENBANK	FJ436983
UNIPROT	COJSA8
COMMENT	Allelic variation in <i>Lr34</i> between the resistant form in cv. Chinese Spring and the susceptible form in cv. Renan involves three polymorphisms. The first one is an A/T single-nucleotide polymorphism (SNP) in intron 4, the second one is a 3-bp ins/del in exon 11, and the third one is a C/T SNP in exon 12, where an amino acid is altered. No difference within 2 kb of the putative <i>Lr34</i> promoter region was detected between the two alleles
KEYWORDS	Chinese Spring, resistant, Lr34
REFERENCE	Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science. 2009, 323(5919):1360-3
PUBMED	19229000
END	

чивает поиск и визуализацию информации о генах устойчивости растений. Разработаны словари для поисковых полей базы данных по генам устойчивости к патогенным грибам. Пользователю предоставляется возможность просматривать списки видов растений, генов и их продуктов, сортов и мутантных линий, устойчивых и чувствительных

генотипов, заболеваний и вызывающих их патогенных организмов, а также ключевых слов, которые содержатся в базе. Кроме этого, пользователь имеет возможность осуществлять поиск по большому числу поисковых полей таблиц PRG_GENE и PRG_ALLELE. Поиск проводится отдельно для таблиц PRG_GENE и PRG_ALLELE.

Поиск в таблице PRG_ALLELE позволяет найти сорта, устойчивые или чувствительные к определенным заболеваниям (поиск по полю RESISTANCE). Затем пользователь имеет возможность получить нуклеотидную последовательность аллельного варианта гена, обеспечивающего устойчивость, кликнув на идентификатор последовательности в поле SEQUENCE_ID таблицы PRG_ALLELE.

Поиск в таблице PRG_GENE позволяет находить гены по следующим критериям:

- названию (поиск по полю GENE);
- расположению в определенной хромосоме (поиск по полям MAPPING и KEYWORDS);
- принадлежности определенному организму (поиск по полю HOST_SPECIES);
- участию в формировании защитного ответа при действии определенного патогена (поиск по полю PATHOGEN);
- участию в формировании защитного ответа при определенном заболевании (поиск по полю DISEASE).

Если найденный сигнал удовлетворяет требованиям, пользователь может получить дополнительную информацию об уровне экспрессии гена из поля COMMENT и нуклеотидную последовательность из поля SEQUENCE_ID таблицы PRG_GENE. В таблице PRG_GENE существует возможность проводить комплексный поиск по нескольким полям. Для этого необходимо перейти на страницу сложного запроса, нажав кнопку «Search» на странице с типовым описанием структуры полей таблицы PRG_GENE. На этой странице меню «Combine searches with» задает логическую операцию для выполнения совместных запросов по нескольким полям. Например, для поиска генов, устойчивых к действию *Puccinia triticina* и локализованных в пятой хромосоме пшеницы, необходимо установить значение этого меню «AND», в левом столбце формы запроса выбрать поле «PATHOGEN», напротив в правом столбце ввести его название – *Puccinia triticina*. В следующей строке формы в левом столбце выбрать «MAPPING», а в правом ввести часть названия хромосомной локализации со звездочкой, чтобы сделать расширенный поиск (*5), и выполнить запрос. В результате выполнения запроса система выдает список генов, расположенных в 5 хромосоме, которые принимают участие в защитной реакции на действие патогена *P. triticina*, вызывающего листовую ржавчину, например PRG_GENE:TaLr1.

Заключение

Разработана база данных PRG, предназначенная для накопления информации, необходимой для развития представлений о механизмах формирования защитного ответа у растений и поиска новых генов устойчивости. База данных PRG содержит структурированную информацию

о генах устойчивости и защитного ответа. Текущая версия базы данных генов устойчивости к фитопатогенам состоит из трех таблиц (генов, аллелей, нуклеотидных последовательностей) и содержит 216 уникальных записей. Формат представления данных обеспечивает быстрый поиск последовательностей генов, принимающих участие в формировании защитного ответа на действие патогенных грибов у пшеницы и родственных ей злаков, что необходимо для создания дополнительных маркеров, обеспечивающих поиск новых генов устойчивости у растений.

Acknowledgments

This work was supported by the Ministry of Education and Science of the Russian Federation, grant agreement 14.604.21.0107 of August 7, 2014, unique identifier RFMEFI60414X0107.

Conflict of interest

The authors declare no conflict of interest.

References

- Bülow L., Schindler M., Hehl R. PathoPlant®: a platform for microarray expression data to analyze co-regulated genes involved in plant defense responses. *Nucl. Acids Res.* 2007;35(1):D841-45. DOI 10.1093/nar/gkl835.
- Mago R., Tabe L., McIntosh R.A., Pretorius Z., Kota R., Paux E., Wicker T., Breen J., Lagudah E.S., Ellis J.G., Spielmeier W. A multiple resistance locus on chromosome arm 3BS in wheat confers resistance to stem rust (*Sr2*), leaf rust (*Lr27*) and powdery mildew. *Theor. Appl. Genetics.* 2011;123(4):615-623. DOI 10.1007/s00122-011-1611-y.
- McIntosh R.A., Wellings C.R., Park R.F. Wheat Rusts: An Atlas of Resistance Genes. Australia, Victoria: CSIRO Publ., 1995.
- McIntosh R.A., Yamazaki Y., Dubcovsky J., Rogers J., Morris C., Somers D.J., Appels R., Devos K.M. Catalogue of gene symbols for wheat. 2013. <http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp> (Accessed 19 Sept. 2016).
- Rosewarne G.M., Singh R.P., Huerta-Espino J., William H.M., Bouchet S., Cloutier S., McFadden H., Lagudah E.S. Leaf tip necrosis, molecular markers and β 1-proteasome subunits associated with the slow rusting resistance genes *Lr46/Yr29*. *Theor. Appl. Genetics.* 2006;112(3):500-508. DOI 10.1007/s00122-005-0153-6.
- Sanseverino W., Hermoso A., D'Alessandro R., Vlasova A., Andolfo G., Frusciante L., Lowy E., Roma G., Ercolano M.R. PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucl. Acids Res.* 2013;41(D1):D1167-1171. DOI 10.1093/nar/gks1183.
- Singh R.P., Huerta-Espino J., William H.M. Genetics and breeding for durable resistance to leaf and stripe rusts in wheat. *Turk. J. Agric. For.* 2005;29:121-127.
- Spielmeier W., Mago R., Wellings C., Ayliffe M. *Lr67* and *Lr34* rust resistance genes have much in common—they confer broad spectrum resistance to multiple pathogens in wheat. *BMC Plant Biology.* 2013;13:96. DOI 10.1186/1471-2229-13-96.
- Zdobnov E.M., Lopez R., Apweiler R., Eizold T. The EBI SRS server – recent developments. *Bioinformatics.* 2002;18(2):368-373.



Методы маркирования клеток для изучения судьбы клеточных поколений

А.М. Юнусова¹✉, Н.Р. Баттулин^{1, 2}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

В процессе развития многоклеточного организма из одной тотипотентной зиготы образуется огромное количество клеток (триллионы для человека) с разной специализацией. Во взрослом состоянии многие ткани постоянно самообновляются: старые клетки погибают, а из популяции стволовых клеток образуются новые. Изучение судьбы отдельных клеток, их происхождения и родственных связей дает ответы на многие вопросы, связанные как с нормальным развитием, так и с патогенезом. Прямые наблюдения за развивающимися эмбрионами позволили установить судьбу бластомеров асидии *Styela partita*, а также происхождение всех клеток червя *Caenorhabditis elegans*. Исследователям повезло с объектом: в первом случае происходит естественное «маркирование» бластомеров, во втором, поскольку тело червя прозрачно, можно следить за каждой клеткой нематоды. В большинстве же случаев определение клеточных линий и идентификация субпопуляций стволовых клеток представляют большую трудность для исследователей. Поэтому для отслеживания судьбы отдельных клеток стали разрабатываться методы маркирования, основанные на внесении в клетки специфических меток, которые наследуются в ходе делений. Поскольку все потомки исходной клетки несут одинаковые метки, их можно легко отличить от потомков других клеток. В обзоре обсуждаются методы маркирования клеток с помощью красителей и генетических конструкций, обеспечивающих синтез белков-репортеров, по наличию которых можно установить родство клеток. Особое внимание уделено методам, основанным на внесении наследуемой метки в геном клеток (генетическое маркирование), в том числе вирусному маркированию и клеточному баркодированию. Один из разделов обзора посвящен маркированию клеток с помощью системы CRISPR/Cas – популярного инструмента геномной инженерии.

Ключевые слова: отслеживание судьбы клеточных поколений; методы маркирования клеток; клеточное баркодирование; CRISPR/Cas-маркирование.

Cell-marking techniques for cell lineage tracing

A.M. Yunusova¹✉, N.R. Battulin^{1, 2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

A zygote, the only totipotent cell of the developing organism, transforms into a complex, multicellular entity with billions (for humans) of highly specialized cells and tissues. Most adult tissues are maintained by a combination of highly dynamic processes of senescence, apoptosis and rejuvenation with new cells constantly arising from dispersed depots of stem cells. Studying individual cell fates and their intertwined relations thus aids in understanding ontogenetic development as well as pathogenic processes in the body. Direct observations of developing embryos uncovered the fate of single blastomeres of ascidia and nematode. In both cases, research benefited from the simplicity of these objects, because, in ascidia, each blastomere has unique signatures naturally, and, in nematode, transparency of a worm's 959-cell body allows every cell to be traced through development individually. In most cases, however, studying cellular lineages and identification of stem cells' subpopulations are a true challenge for investigators. To trace the cell's fate, novel methods were invented that introduce special tags into cells, the tags that would be inherited during cell divisions. Every descendent of a marked cell bears the same tag and can easily be distinguished from unrelated cellular neighbors. This review focuses on modern methods for cell tracing with dyes and genetic constructs encoding protein reporters that mark cell lineages. Special focus is on genome-integrated tags (genetic labeling), such as viral and cellular barcoding. One chapter of the review describes novel advancements in the field of CRISPR/Cas9-based cellular barcoding.

Key words: cell lineage tracing; cell-marking techniques; cellular barcoding; CRISPR/Cas-barcoding.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Юнусова А.М., Баттулин Н.Р. Методы маркирования клеток для изучения судьбы клеточных поколений. Вавиловский журнал генетики и селекции. 2016;20(6):909-917. DOI 10.18699/VJ16.211

HOW TO CITE THIS ARTICLE:

Yunusova A.M., Battulin N.R. Cell-marking techniques for cell lineage tracing. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):909-917. DOI 10.18699/VJ16.211

REVIEW

Received 31.10.2016

Accepted for publication 08.12.2016

© AUTHORS, 2016



e-mail: yunusova-nastya92@mail.ru

Один из важнейших постулатов клеточной теории гласит: *omnis cellula ex cellula*. Это утверждение, сформулированное Р. Вирховым в 1858 г., означает, что всякая клетка происходит от другой клетки (Mazzarello, 1999). Многоклеточный организм со всем разнообразием клеточных типов берет начало от одной тотипотентной клетки – зиготы. Когда и каким образом происходит принятие клетками решения о дифференцировке в тот или иной тип? Все ли клетки-предшественники в равной мере участвуют в образовании тканей и органов? Являются ли процессы дифференцировки стохастическими (т. е. случайными) или они строго предопределены? Во всех этих вопросах центральным предметом исследования является судьба отдельной клетки в многоклеточном организме. Методы, позволяющие определять потомков отдельных клеток и изучать их судьбу, получили название «methods of lineage tracing», что можно перевести как «методы отслеживания клеточных поколений». В биологии развития эти методы широко используются для составления карт презумптивных зачатков (Kretzschmar, Watt, 2012). Помимо этого, методы отслеживания судьбы клеточных поколений нашли применение и для изучения гетерогенных популяций стволовых и раковых клеток (Buckingham, Meilham, 2011).

Изучение судьбы клеточных поколений. Историческая справка

Прямое наблюдение – исторически первый способ изучения судьбы отдельных клеток. В конце 19 в. Э. Уилсон, наблюдая под микроскопом за поведением клеток развивающегося эмбриона многощетинкового червя, ввел понятие «клеточная линия» (Wilson, 1892). Позднее, в 1905 г., Э. Конклин установил судьбу отдельных бластомеров эмбриона асидии *Styela partita*. В этом случае определение родства между клетками было возможным в силу того, что цитоплазма яйца этого вида асидий различается по цвету: от темно-желтой до темно-серой и прозрачной. При дроблении яйца различные виды цитоплазмы попадают в разные бластомеры, обеспечивая их естественное маркирование, что облегчает наблюдение за делением клеток (Conklin, 1905). Метод прямого наблюдения был также успешно применен и для изучения развития червя *Caenorhabditis elegans* (Sulston, Horvitz, 1977; Sulston et al., 1983). С помощью микроскопии Номарского (интерференционно-контрастная микроскопия) Дж. Салстон смог проследить происхождение всех клеток этой нематоды, пользуясь тем, что ее тело прозрачно. Оказалось, что судьба каждой клетки червя строго детерминирована, а совокупное число клеток, составляющих взрослый организм, постоянно и составляет 959 клеток. Преимуществами прямого наблюдения являются его неинвазивность, а также возможность длительного и непрерывного анализа (Schroeder, 2011). С развитием методов микроскопии в режиме реального времени в этой области удалось добиться большого прогресса. Так, например, исследователи смогли реконструировать раннее развитие эмбрионов плодовой мушки *Drosophila melanogaster* и рыбки *Danio rerio* (Keller et al., 2008; Amat et al., 2014).

Нужно отметить, что в работах Д. Салстона и Э. Конклина не было необходимости метить клетки. Однако в

большинстве случаев исследователь не имеет возможности проследить за судьбой отдельных клеток без внесения специальных меток, позволяющих отличить потомков исследуемой клетки от потомков других клеток. Стали активно развиваться способы маркирования на основе введения в клетку (с помощью инъекций или электропорации) прижизненных красителей и радиоактивных изотопов (Vogt, 1929; Rosenquist, 1970). Во избежание распространения красителя по межклеточным контактам в соседние клетки были разработаны модификации метода с использованием жирорастворимых цианинов, встраивающихся в мембрану клетки, а также флуорохромов, конъюгированных с большими молекулами, такими как декстран, стерически затрудняющими переход молекул между клетками (Axelrod, 1979; Serbedzija et al., 1989; Eagleson, Harris, 1990). Во многих исследованиях также была использована пероксидаза хрена, большой фермент, не способный пройти через щелевые контакты и образующий в результате ферментативной реакции люминесцирующий продукт (Weisblat et al., 1980; Balakier, Pedersen, 1982; Lawson et al., 1986). Например, инъецировав пероксидазу в один из бластомеров двухклеточного эмбриона мыши, исследователи смогли проследить расположение меченых клеток до стадии морулы (Balakier, Pedersen, 1982). Среди ограничений данного метода – разбавление метки при делении клеток, что делает невозможным длительное наблюдение.

Позднее были разработаны способы маркирования клеток, основанные на внесении в их геном репортерных конструкций, как правило, кодирующих флуоресцентные белки или люциферазу, присутствие которых в клетке легко детектируется (Holt et al., 1990; Chalfie et al., 1994). Поскольку такие конструкции стабильно интегрируются в геном и наследуются клеточными поколениями в ходе

Интенсивные испытания ядерного оружия в годы холодной войны (1950–1960) привели к накоплению большого количества радиоуглерода ($C-14$) в атмосфере и, как следствие, к массовому маркированию клеток населения земного шара радиоактивными изотопами. После запрета ядерных испытаний в 1963 г. уровень $C-14$ стал экспоненциально снижаться (De Vries, 1958; Nydal, Lovseth, 1965; Levin, Kromer, 2004). Поскольку при репликации происходит включение этого изотопа в новую цепь ДНК, последняя может выступать в роли своеобразной капсулы времени, храня запись о том, когда произошло событие синтеза ДНК или, иными словами, рождения новой клетки. Группа исследователей из Швеции и США, проведя анализ количества $C-14$ в разных тканях и органах, смогла установить время жизни некоторых типов клеток. Так, например, оказалось, что возраст кортикальных нейронов в затылочной части мозга совпадает с возрастом организма, что свидетельствует об отсутствии процессов нейрогенеза в этой области мозга после рождения (Spalding et al., 2005). Этим же методом было установлено, что у человека к 50 годам 55 % кардиомиоцитов составляют клетки, образовавшиеся во время эмбриогенеза, тогда как другие 45 % образовались в ходе перинатального развития (Bergmann et al., 2009).

делений, решается проблема с разбавлением метки. На основе флуоресцентных белков была разработана технология Brainbow (по аналогии с rainbow), с помощью которой были маркированы нейроны головного мозга мыши (Livet et al., 2007; Cai et al., 2013). Эта методика основана на введении в геном организма конструкции из нескольких последовательно расположенных генов разных флуоресцентных белков (рис. 1). Эти гены разделены lox-сайтами, и только один ген в конструкции, расположенный первым и взаимодействующий с регуляторной областью, экспрессируется. Под воздействием Cre-рекомбиназы происходят вырезание и инверсия генов флуоресцентных белков, вследствие чего любой другой ген получает возможность взаимодействовать с регуляторной областью и экспрессироваться. Учитывая тот факт, что в геном интегрируется несколько копий такой конструкции, а рекомбинационные события случайны, каждая клетка характеризуется уникальным паттерном экспрессии флуоресцентных белков. Например, первая конструкция Brainbow-1.0 кодировала три флуоресцентных белка и при наличии трех копий в геноме обеспечивала 10 разных комбинаций, а при восьми копиях исследователи смогли детектировать от 90 до 160 псевдоцветов (Livet et al., 2007). Технология Brainbow является важным инструментом для изучения коннектома, поскольку окрашивание каждого нейрона и его отростков в определенный цвет позволяет визуализировать взаимодействия между нейронами (Lichtman et al., 2008). Конструкция Flybow была использована для визуализации нейронов и их связей в мозге плодовой мушки *Drosophila melanogaster* (Hadjieconomou et al., 2011). Подобную конструкцию, названную Skinbow, использовали для изучения регенерации кожи рыбки *Danio rerio* (Chen et al., 2016).

Генетическое маркирование

В этом разделе описаны способы маркирования клеток, основанные на внесении в геном клеток специальных меток – от хромосомных перестроек до ДНК-баркодов.

Хромосомные перестройки как метки

Первый эксперимент с использованием хромосомных перестроек как меток был проведен в начале 60-х годов прошлого века Э. Беккером под руководством Мак-Каллоха и Тилла, признанных исследователей стволовых клеток. Цель эксперимента заключалась в отслеживании судьбы гемопоэтических клеток при трансплантации. Ранее уже было показано, что клетки костного мозга, трансплантированные мышам, получившим летальную дозу радиации, образуют на поверхности селезенки скопления клеток эритроидного и гранулоцитарного ряда (Till, McCulloch, 1961). Однако было неизвестно, являются ли эти клеточные скопления клонами – потомками всего одной клетки, обладающей потенциалом к образованию эритроцитов, и гранулоцитов. Э. Беккер провел свой знаменитый эксперимент, суть которого заключалась в следующем: клетки костного мозга были облучены сублетальной дозой радиации, вызывавшей хромосомные перестройки, а затем трансплантированы облученному реципиенту. Таким образом, трансплантированные клетки несли уникальные хромосомные перестройки, которые можно наблюдать

под микроскопом. Оказалось, что каждая колония состоит преимущественно из клеток, несущих одну и ту же хромосомную мутацию, т. е. являющихся клонами. Такие мультипотентные стволовые клетки впоследствии были названы колониеобразующими единицами (Becker et al., 1963). Последующие работы с использованием данного метода подтвердили этот результат и показали, что трансплантированные клетки обнаруживаются не только в костном мозге реципиента, но и в тимусе и лимфатических сосудах. Таким образом, с помощью этого метода было прослежено и происхождение колоний клеток, и их локализация в организме (Wu et al., 1967, 1968). Очевидно, однако, что такое сильное воздействие, как радиация, вызывающая хромосомные перестройки, в значительной мере изменяет судьбу клеток, поэтому поведение облученных клеток нельзя считать идентичным поведению нативных клеток в естественных условиях.

Сайты интеграции ретровирусов как метки

Ретровирусы представляют собой семейство РНК-содержащих вирусов, способных интегрироваться в геном клетки при заражении. Благодаря этому свойству векторные конструкции на основе дефективных (неспособных к репликации) ретровирусов часто используются генными инженерами в качестве доставщиков генетического материала в геном клеток. Поскольку встройка происходит случайным образом, каждая инфицированная клетка характеризуется уникальным сайтом интеграции провирусной последовательности. Паттерн интеграции наследуется дочерними клетками, и, следовательно, может использоваться как метка. В первых работах с использованием этой методики определение сайта интеграции провируса осуществлялось с помощью Саузерн-блоттинга (Capel et al., 1990; Jordan, Lemischka, 1990). Впервые было убедительно доказано, что стволовые клетки крови мыши являются мультипотентными и, кроме того, способны достаточно продолжительное время поддерживать самообновление, т. е. не дифференцироваться (Dick et al., 1985; Mazurier et al., 2004). Позднее были предложены более чувствительные методы для определения сайтов интеграции на основе ПЦР (Laukkanen et al., 2005; Keller et al., 2008; Maetzig et al., 2011) и микрочипового анализа (Schepers et al., 2008; van Heijst et al., 2009). Помимо изучения особенностей дифференцировки гемопоэтических клеток, методика подобного рода была использована и для изучения процессов нейрогенеза млекопитающих *in vivo*. Было показано, что стволовые клетки гиппокампа мультипотентны – они способны образовывать и нейроны, и олигодендроциты, и астроглию (Palmer et al., 1997). Также с помощью ретровирусного маркирования внутри эмбриональных стволовых клеток человека удалось выделить две субпопуляции, участвующие в самообновлении, но дающие разный вклад в дифференцировку *in vivo* (формирование тератом в иммунодепрессивной мыши) и *in vitro* (дифференцировка в культуре) (Stewart et al., 2010).

ДНК-баркоды как метки

Необходимо отметить, что термин «баркодирование» знаком отечественным специалистам преимущественно в контексте молекулярного баркодирования – метода, ис-

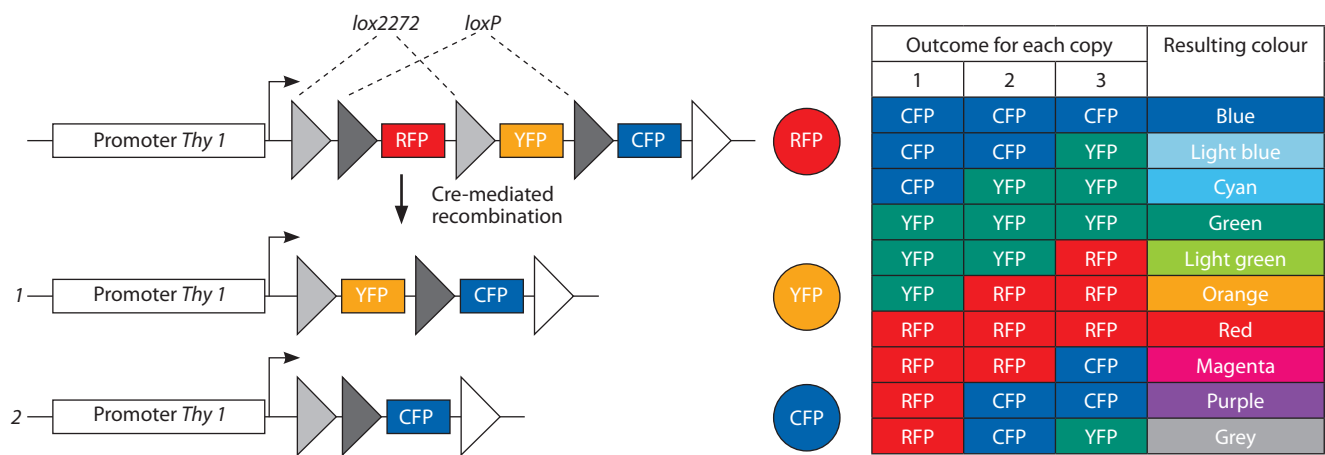


Fig. 1. Brainbow-1.0, the simplest Brainbow construction (Livet et al., 2007).
Cre-mediated recombination switches expression to either (1) YFP or (2) M-CFP. Two equally possible outcomes of the recombination event are shown. The *Thy1* promoter drives the expression of the constructs in neuronal cells. The table shows a list of pseudocolors that are assigned to every possible combination of three fluorophores (based on three copies of the construct in the genome).

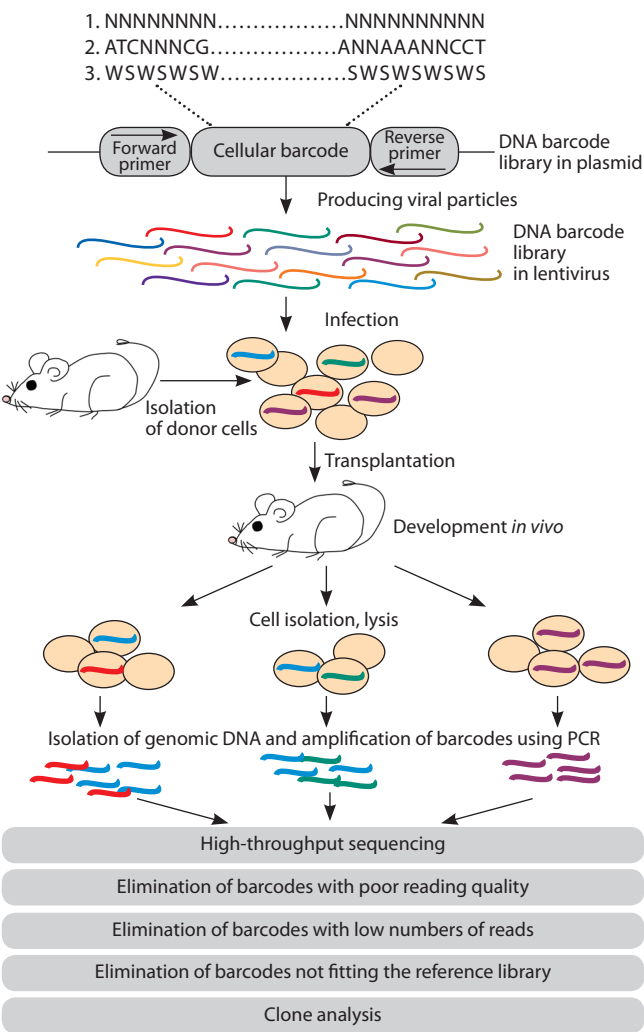


Fig. 2. Methods for cellular barcoding (Lu et al., 2011, with modifications).
The picture illustrates several variants of DNA-barcodes: (1) completely degenerate barcode; (2) partially degenerate barcode, in which random and constant nucleotides are combined; (3) GC content-balanced barcode (W stands for “weak” (AT) nucleotides and S for “strong” (GC) nucleotides).

пользуемого для устранения ошибок ПЦР и секвенирования при анализе данных высокопроизводительного секвенирования. Однако к маркированию клеточных поколений эта процедура не имеет отношения. Для этой цели был разработан метод клеточного баркодирования (cellular barcoding) (Scheepers et al., 2008; van Heijst et al., 2009). Основной принцип клеточного баркодирования заключается во внесении в геном клетки специальной метки – ДНК-баркода. ДНК-баркод представляет собой искусственно синтезированную вырожденную (или полувыврожденную) последовательность нуклеотидов, которая интегрируется в геном с помощью вирусов и наследуется в ходе делений следующими клеточными поколениями (рис. 2). Метод ДНК-баркодирования логично вытекает из описанного выше ретровирусного маркирования. Однако, в отличие от последнего, для того чтобы определить клональную принадлежность клеток, достаточно секвенировать только последовательность ДНК-баркода, а не анализировать сайты интеграции провирусов. Определение ДНК-баркодов проводится, как правило, с помощью массового параллельного секвенирования (Gerrits et al., 2010; Lu et al., 2011; Cheung et al., 2013; Gerlach et al., 2013; Naik et al., 2013; Nguyen et al., 2014a, b). Отличительная особенность данного метода заключается в возможности анализировать судьбу сотен клеток одновременно. Поскольку технология клеточного баркодирования является довольно новой и перспективной, а, кроме того, недостаточно освещенной в русскоязычной литературе, мы хотим более подробно описать некоторые особенности и ограничения этой методики.

Исследователям, собирающимся использовать клеточное баркодирование, необходимо обратить внимание на следующие экспериментальные параметры: 1) размер библиотеки баркодированных вирусов, 2) трансдукцию клеток баркодированными вирусами, а также 3) технические повторы.

1. Размер (разнообразие) библиотеки ДНК-баркодов. Поскольку внесение ДНК-баркода в геном клетки осуществляется с помощью ленти- или ретровирусных кон-

струкций, под размером библиотеки подразумевается количество уникальных ДНК-баркодов в лентивирусной библиотеке (т.е. число вирусных частиц с уникальными баркодами). Теоретически разнообразие ДНК-баркодов обуславливается длиной вырожденной части самого баркода. Так, например, для ДНК-баркода, содержащего 20 вырожденных позиций, количество уникальных комбинаций составляет более чем 10^{12} . Тем не менее на практике это разнообразие сильно ограничено в силу того, что получение библиотеки баркодированных лентивирусов включает в себя стадии клонирования ДНК-баркода в плазмидный вектор и сборку вирусных частиц. Эффективность этих стадий невысока и не позволяет сохранить все теоретическое разнообразие ДНК-баркодов. При этом чем больше клеток нужно маркировать, тем больше должен быть размер библиотеки ДНК-баркодов. В противном случае велика вероятность того, что неродственные клетки могут быть помечены одинаковым вариантом ДНК-баркода. В такой ситуации они интерпретируются как потомки одной и той же клетки. Для выбора оптимального соотношения количества уникальных ДНК-баркодов к количеству клеток создатели клеточного баркодирования предложили следующее эмпирическое правило: число маркируемых клеток должно составлять не более 10 % от количества уникальных вариантов баркодов (Naik et al., 2014). Для оценки разнообразия ДНК-баркодов рекомендуется проведение высокопроизводительного секвенирования ДНК-баркодов на разных этапах создания библиотеки, начиная от плазмидных векторов и заканчивая клетками, зараженными баркодированными вирусами. На основе полученных данных создается референсная библиотека, содержащая все доступные варианты ДНК-баркодов. Такая библиотека позволяет надежно отличить «настоящие» ДНК-баркоды от ошибок ПЦР и секвенирования. В обзоре Л.В. Быстрых с коллегами (Bystrykh et al., 2012) приведен пример, в котором при секвенировании пяти колоний клеток исследователи получили 1503 баркода, причем только шесть из них представляли пять начальных клеток (одна из клеток несла два баркода), тогда как остальные баркоды являлись ошибками ПЦР и секвенирования и были исключены из анализа как «шум». В другом обзоре, посвященном клеточному баркодированию, указано, что для включения ДНК-баркода в референсную библиотеку необходимо, чтобы количество прочтений этого баркода при секвенировании составляло не менее 0.00001 % от общего количества прочтений (Naik et al., 2014). Таким образом, чем больше разнообразие ДНК-баркодов в библиотеке, тем больше должна быть глубина секвенирования. Поэтому анализ состава ДНК-баркодов с помощью секвенирования возможен только при относительно небольшом размере библиотеки. В тех случаях, когда секвенирование полной библиотеки невозможно по ряду причин, исследователи прибегают к математическим методам, позволяющим по конечному разнообразию ДНК-баркодов в отдельных клетках/колониях клеток приблизительно оценить «изначальное» разнообразие библиотек, например с помощью симуляции Монте-Карло (Lu et al., 2011; Wu et al., 2014).

2. Трансдукция клеток баркодированными вирусами. Термин «трансдукция» означает внесение в геном клеток

генетического материала с помощью вирусов. Известно, что при трансдукции количество вирусных интеграций (multiplicity of infection) на геном клетки подчиняется распределению Пуассона. Значение MOI, равное 1, говорит о том, что в среднем на одну клетку приходится одна вирусная интеграция, при этом клетка инфицирована одной вирусной частицей с вероятностью 36.7 %, двумя – 18.39 %, тремя – 6.13 %, а вероятность того, что в клетку не попала ни одна вирусная частица, – 36.7 %. Таким образом, чем выше значение MOI, тем большая доля клеток несет одну и более вирусных интеграций. В контексте клеточного баркодирования это означает, что часть клеток несет два и больше ДНК-баркодов, что также затрудняет интерпретацию клеточных взаимоотношений. Так, например, одна клетка с тремя ДНК-баркодами будет воспринята как три клетки с индивидуальными ДНК-баркодами и одинаковой судьбой. Для того чтобы избежать этой переоценки, трансдукцию клеток баркодированными вирусами проводят с низкими значениями MOI (0.1–0.5).

3. Технические повторы. Для того чтобы учесть разного рода систематические ошибки, чрезвычайно важно проводить технические реплики. Баркодированные клетки делятся на две части (при этом предполагается, что дочерние клетки распределяются по ним равномерно), а их ДНК-баркоды амплифицируются и секвенируются независимо. При качественном анализе ДНК-баркоды реплик должны хорошо коррелировать (с коэффициентом корреляции Пирсона около 0.8).

Метод клеточного баркодирования интенсивно используется для изучения процессов кроветворения. С помощью этого метода было показано отличие клонального состава гемопоэтических стволовых клеток (ГСК) молодых и старых животных. Так, из пула «старых» ГСК при трансплантации образуется много небольших клональных линий, тогда как из пула «молодых» ГСК – преимущественно несколько больших колоний (Verovskaya et al., 2013). Изучено влияние провоспалительных цитокинов на миграцию клеток при трансплантации и, как следствие, на эффективность их приживления (Verovskaya et al., 2014). Кроме того, проведена реконструкция линии натуральных киллеров на макак-резусах. Оказалось, что несмотря на общность функций натуральных киллеров и Т-лимфоцитов, развиваются эти клеточные линии из разных предшественников (Wu et al., 2014). Метод клеточного баркодирования также используется для изучения гетерогенных раковых опухолей (Nguyen et al., 2014a; Bhang et al., 2015). С его помощью было установлено, что большинство резистентных к терапии клонов являются потомками небольших субпопуляций клеток, уже присутствующих в опухоли и селективно избегающих терапии (Bhang et al., 2015).

Маркирование клеток с помощью системы CRISPR/Cas9

CRISPR/Cas – система адаптивного приобретенного иммунитета архей и бактерий, направленная на уничтожение попавшей в клетку чужеродной ДНК (как правило, вирусной) (Barrangou et al., 2007). С помощью молекул РНК, закодированных в локусе CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), Cas-белки распоз-

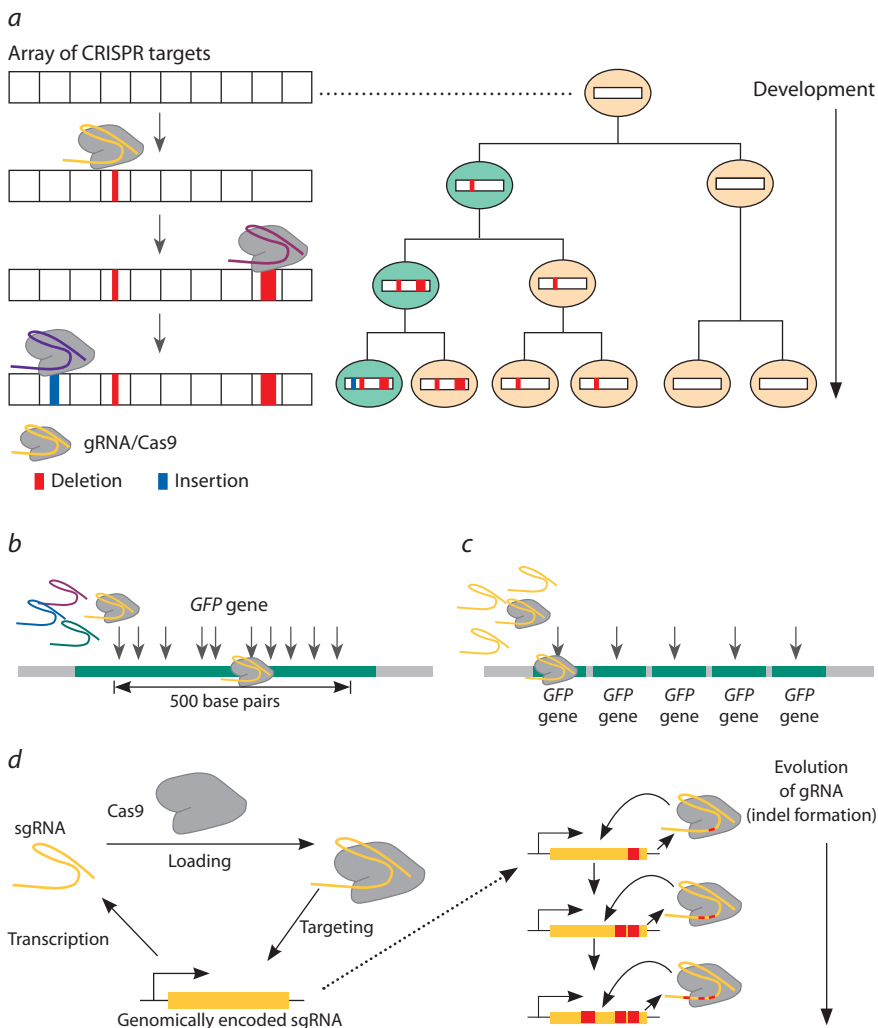


Fig. 3. Approaches for cellular barcoding based on CRISPR/Cas9: (a) GESTALT - method suggested by McKenna et al. (2016) (see in text). The barcode is represented by ten consecutive Cas9 target sites, integrated in the *Danio rerio* genome. Cas9-induced DNA damage leads to unique small insertions or deletions (indel mutations) after repair. As a result, related cells will have similar repair patterns at the break sites. (b) Method from (Schmidt et al., 2016) (see in text). Ten close Cas9 target sites are used as a DNA barcode again, but they are scattered over a 500-bp region in the *GFP* gene. (c) The Scartrace approach from (Junker et al., 2016) (see in text). A single locus with 4–5 copies of the *GFP* gene (each copy having a Cas9 recognition site) serves as a barcode. (d) The mSCRIBE – method proposed in (Perli et al., 2016) and (Kalhor et al., 2016) (see in text). In this approach, self-targeting “evolving” gRNA is chosen as a DNA barcode.

нают участки геномов вирусов и вносят двуцепочечные разрывы, вызывая их деградацию. Эта способность Cas-нуклеаз нашла широкое применение в геномной инженерии (Doudna, Charpentier, 2014; Hsu et al., 2014; Смирнов и др., 2016). С помощью синтетической химерной направляющей РНК (далее gRNA) исследователь может запрограммировать нуклеазу на целевой участок ДНК, комплементарный gRNA (Mali et al., 2013). В месте узнавания Cas-нуклеаза вносит двуцепочечный разрыв, который репарируется клеткой за счет неомологического соединения концов (non-homologous end joining, NHEJ). При этом под действием экзонуклеаз и полимераз происходит потеря, реже вставка, нескольких нуклеотидов (индел-мутации). Таким образом, при лигировании двуцепочечного разрыва, внесенного Cas-нуклеазой, образуется своего рода уникальный «шрам» в ДНК, который может быть использован как метка. Группа исследователей из США решила воспользоваться таким «геномным вандализмом» для маркирования клеток и определения их родства (McKenna et

al., 2016). Разработанный ими способ был назван GESTALT (genome editing of synthetic arrays for lineage tracing) и применен для изучения развития рыбки *Danio rerio*. В качестве баркода, интегрированного в геном клетки, исследователи использовали конкатенерную кассету с 10 последовательно расположенными сайтами узнавания и разрезания для Cas9-нуклеазы (рис. 3, a). Однажды внося разрыв в один из 10 сайтов, который репарируется с образованием мутаций, Cas9-нуклеаза уже не сможет его опознать. И эта метка в виде индел-мутации будет передана дочерним клеткам в ходе деления. Подобно тому как определяется филогенетическое родство между организмами, происхождение клеток и их родство внутри взрослого организма можно определить по количеству накопленных индел-мутаций. Клетки, разошедшиеся в эмбриогенезе позднее, будут иметь сходные баркоды. Исследователи ввели в одноклеточный эмбрион *Danio rerio*, содержащий кассету с 10 мишенями, 10 соответствующих gRNA (по одной на каждый целевой сайт) и Cas9-нуклеазу. Проанализировав баркоды различных органов выросших рыбок, исследователи смогли реконструировать «фамильные» деревья клеточных линий, которые можно найти на сайте <http://gestalt.gs.washington.edu>. Кроме того, было показано, что большинство клеток, составляющих органы взрослой особи рыбки, берут начало всего от нескольких клеток-предшественников. В среднем более половины клеток всех изученных органов были представлены не более чем семью аллелями баркода, за исключением головного мозга, 90 % клеток которого содержали 25 аллелей, и крови, где 98 % клеток содержали только пять аллелей баркода. Преимуществом этой методики является то, что все 10 сайтов могут быть просеквенированы за один раз.

Другая группа исследователей использовала сходный метод, но на другом объекте – круглом черве *Caenorhabditis elegans* (Schmidt et al., 2016). Предлагаемая авторами стратегия маркирования основана на внесении мутаций с помощью Cas9-нуклеазы и 10 направляющих gRNA в 10 участках гена зеленого флуоресцентного белка (green fluorescent protein – GFP) транс-

генной линии *Caenorhabditis elegans* EG6173 (рис. 3, б). Напомним, что для этого маленького червячка, состоящего из 959 клеток, происхождение всех клеточных линий было определено методом прямого наблюдения еще в 1970-х годах (Sulston, Horvitz, 1977). Сравнение двух древ клеточных линий показало достаточно высокий уровень корреляции. Помимо экспериментального доказательства концепции метода, авторы, используя компьютерное моделирование, протестировали ряд дополнительных параметров, влияющих на эффективность маркирования клеток. Среди них ситуация, когда нуклеаза разрезает второй сайт до того, как был репарирован первый, что может привести к удалению участка ДНК между этими двумя сайтами, т. е. к потере части баркода. Поскольку эффективность работы системы CRISPR/Cas9 высока, такие ситуации должны происходить нередко. Даже несмотря на это, маркирование клеток с помощью Cas9 нуклеазы позволяет достаточно точно установить происхождение клеток (Schmidt et al., 2016).

Совместная группа исследователей из Германии и Нидерландов предложила еще один вариант использования CRISPR/Cas системы для маркирования клеток (Junker et al., 2016) (препринт статьи выложен на ресурсе [biorxiv.org](https://www.biorxiv.org)). Технология маркирования, названная Scar-trace, также базируется на образовании индел-мутаций при репарации двуцепочечных разрывов ДНК. С помощью инъекций Cas9-нуклеазы и gРНК против гена *GFP* исследователи вносили мутации в одноклеточные эмбрионы *Danio rerio*, содержащие locus с несколькими интеграциями гена *GFP* (рис. 3, в). Помимо изучения происхождения клеточных линий, авторам удалось исследовать и процесс регенерации хвостового плавника. Оказалось, что пространственное расположение клеточных клонов исходного неповрежденного и регенерировавшего плавника совпадает, что свидетельствует об одинаковой программе развития и регенерации.

Принципиально другой подход с использованием программируемой нуклеазы Cas9 для маркирования клеток был реализован группой исследователей Массачусетского технологического института под руководством Т. Лу (Perli et al., 2016). Разработанная ими технология mSCRIBE (mammalian Synthetic Cellular Recorders Integrating Biological Events) основана на «самонаправленной» gRNA, которая способствует привлечению Cas9-нуклеазы в свой собственный locus, что приводит к возникновению в нем индел-мутаций (рис. 3, г). Впоследствии с этого локуса считывается немного видоизмененная gRNA, которая опять будет направлена на свой locus. Цикл повторяется много раз, при этом накапливается информация о каждом акте изменений. Эта информация может быть использована как для изучения судьбы клеточных поколений, так и для записи времени и силы воздействия различных стимулов на клетку. В последнем случае необходимо поместить такой «эволюционирующий» баркод под промотор, активирующийся под воздействием этого стимула. В роли стимула могут выступать обработка клеток антибиотиками, активация сигнальных путей, связанных с воспалением, и т. д.

Еще одна работа с использованием нуклеаз была проведена коллективом исследователей из лаборатории

Дж. Черча (Kallhor et al., 2016). Авторы также использовали принцип «эволюционирующего» баркода на основе самонаправленного локуса gRNA (в статье – homing gRNA) (рис. 3, з). В идеале в каждом таком локусе должно образовываться большое количество разнообразных вариантов мутаций, причем с равной вероятностью. Проанализировав варианты индел-мутаций и мисматчи, полученные в эксперименте, исследователи оценили информационную емкость одного локуса из 21 нуклеотидов в 5 бит. Этого количества достаточно для маркирования 32 клеток (2^5). Тогда как для маркирования всех нейронов в мозге мыши (примерно 75 млн клеток) требуется как минимум шесть самонаправленных локусов ($(2^5)^6$), а для маркирования всех клеток взрослой мыши (примерно 12 млрд клеток) необходимо семь локусов ($(2^5)^7$).

Заключение

Методы маркирования клеток прошли большой путь развития: от прямых наблюдений за развивающимися эмбрионами до баркодирования с помощью системы CRISPR/Cas9. По стратегии эти методы можно разделить на два типа. Первый тип подразумевает маркирование большого количества клеток, а затем изучение распределения и вклада их потомков в разные ткани и органы. К ним относятся методы маркирования с помощью репортерных белков, сайтов интеграций вирусов, а также метод клеточного баркодирования. Все они позволяют лишь установить родство клеток, но не открывают тайну их происхождения. Второй тип методов представлен исключительно маркированием с помощью системы CRISPR/Cas9 и имеет принципиальное отличие. В этом случае маркируется всего одна клетка, потомки которой наследуют изменяющийся, эволюционирующий баркод, что позволяет определить целые «фамильные» древа клеток. Система CRISPR/Cas9 в настоящее время заняла место главного инструмента генной инженерии, поэтому много работ, направленных на маркирование клеток с ее помощью, отражает потребность ученых в развитии методов отслеживания клеточных поколений. Можно ожидать, что в скором времени ученые смогут установить судьбу каждой клетки при развитии такого сложного организма, как человек.

Acknowledgments

This work was supported by State Budgeted Project 0324-2015-0004.

Conflict of interest

The authors declare no conflict of interest.

References

- Amat F., Lemon W., Mossing D.P., McDole K., Wan Y., Branson K., Myers E.W., Keller P.J. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. *Nat. Methods*. 2014;11:951-958. DOI 10.1038/nmeth.3036.
- Axelrod D. Carbocyanine dye orientation in red cell membrane studied by microscopic fluorescence polarization. *Biophys. J.* 1979;26:557-573. DOI 10.1016/S0006-3495(79)85271-6.
- Balakier H., Pedersen R.A. Allocation of cells to inner cell mass and trophectoderm lineages in preimplantation mouse embryos. *Dev. Biol.* 1982;90:352-362. DOI 10.1016/0012-1606(82)90384-0.

- Barrangou R., Fremaux C., Deveau H., Richards M., Moineau S., Romero D.A., Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007;315(5819):1709-1712. DOI 10.1126/science.1138140.
- Becker A.J., McCulloch E.A., Till J.E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature*. 1963;197:452-454. DOI 10.1038/197452a0.
- Bergmann O., Bhardwaj R.D., Bernard S., Zdunek S., Barnabé-Heider F., Walsh S., Zupicich J., Alkass K., Buchholz B.A., Druid H., Jovinge S., Frisén J. Evidence for cardiomyocyte renewal in humans. *Science*. 2009;324:98-102. DOI 10.1126/science.1164680.
- Bhang H.C., Ruddy D.A., Krishnamurthy Radhakrishna V., Causi J.X., Zhao R., Hims M.M., Singh A.P., Kao I., Rakiec D., Shaw P., Balak M., Raza A., Ackley E., Keen N., Schlabach M.R., Palmer M., Leary R.J., Chiang D.Y., Sellers W.R., Michor F., Cooke V.G., Korn J.M., Stegmeier F. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* 2015; 21:440-448. DOI 10.1038/nm.3841.
- Buckingham M.E., Meilhac S.M. Tracing cells for tracking cell lineage and clonal behavior. *Dev. Cell*. 2011;21(3):394-409. DOI 10.1016/j.devcel.2011.07.019.
- Bystrykh L.V., Verovskaya E., Zwart E., Broekhuis M., de Haan G. Counting stem cells: methodological constraints. *Nat. Methods*. 2012;9:567-574. DOI 10.1038/nmeth.2043.
- Cai D., Cohen K.B., Luo T., Lichtman J.W., Sanes J.R. Improved tools for the Brainbow toolbox. *Nat. Methods*. 2013;10(6):540-547. DOI 10.1038/nmeth.2450.
- Capel B., Hawley R.G., Mintz B. Long- and short-lived murine hematopoietic stem cell clones individually identified with retroviral integration markers. *Blood*. 1990;75:2267-2270.
- Chalfie M., Tu Y., Euskirchen G., Ward W.W., Prasher D.C. Green fluorescent protein as a marker for gene expression. *Science*. 1994; 263:802-805. DOI 10.1126/science.8303295.
- Chen C.H., Puliafito A., Cox B.D., Primo L., Fang Y., Di Talia S., Poss K.D. Multicolor cell barcoding technology for long-term surveillance of epithelial regeneration in zebrafish. *Dev. Cell*. 2016;36: 668-680. DOI 10.1016/j.devcel.2016.02.017.
- Cheung A.M.S., Nguyen L.V., Carles A., Beer P., Miller P.H., Knapp D.J.H.F., Dhillon K., Hirst M., Eaves C.J. Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. *Blood*. 2013;122:3129-3137. DOI 10.1182/blood-2013-06-508432.
- Conklin E.G. The organization and cell lineage of the ascidian egg. *J. Acad. Natl. Sci.* 1905;13:1-119. DOI 10.1007/s13398-014-0173-7.2.
- De Vries H. Atomic bomb effect: variation of radiocarbon in plants, shells, and snails in the past 4 years. *Science*. 1958;128:250-251. DOI 10.1126/science.128.3318.250.
- Dick J.E., Magli M.C., Huszar D., Phillips R.A., Bernstein A. Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/W^v mice. *Cell*. 1985;42:71-79. DOI 10.1016/S0092-8674(85)80102-1.
- Doudna J.A., Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014;346(6213):1258096. DOI 10.1126/science.1258096.
- Eagleson G.W., Harris W.A. Mapping of the presumptive brain regions in the neural plate of *Xenopus laevis*. *J. Neurobiol.* 1990;21:427-440. DOI 10.1002/neu.480210305.
- Gerlach C., Rohr J.C., Perié L., van Rooij N., van Heijst J.W.J., Velds A., Urbanus J., Naik S.H., Jacobs H., Beltman J.B., de Boer R.J., Schumacher T.N.M. Heterogeneous differentiation patterns of individual CD8⁺ T cells. *Science*. 2013;340:635-639. DOI 10.1126/science.1235487.
- Gerrits A., Dykstra B., Kalmykova O.J., Klauke K., Verovskaya E., Broekhuis M.J.C., De Haan G., Bystrykh L.V. Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood*. 2010;115:2610-2618. DOI 10.1182/blood-2009-06-229757.
- Hadjicoumou D., Rotkopf S., Alexandre C., Bell D.M., Dickson B.J., Salecker I. Flybow: genetic multicolor cell labeling for neural circuit analysis in *Drosophila melanogaster*. *Nat. Methods*. 2011;8:260-266. DOI 10.1038/nmeth.1567.
- Holt C.E., Garlick N., Cornel E. Lipofection of cDNAs in the embryonic vertebrate central nervous system. *Neuron*. 1990;4:203-214. DOI 10.1016/0896-6273(90)90095-W.
- Hsu P.D., Lander E.S., Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;157:1262-1278. DOI 10.1016/j.cell.2014.05.010.
- Jordan C.T., Lemischka I.R. Clonal and systemic analysis of long-term hematopoiesis in the mouse. *Gene Dev.* 1990;4(2):220-232. DOI 10.1101/gad.4.2.220.
- Junker J.P., Spanjaard B., Peterson-Maduro J., Alemany A., Hu B., Florescu M., van Oudenaarden A. Massively parallel whole-organism lineage tracing using CRISPR/Cas9 induced genetic scars. *BioRxiv*. 2016. DOI 10.1101/056499.
- Kalhor R., Mali P., Church G.M. Rapidly evolving homing CRISPR barcodes. *Nat. Methods*. 2016. DOI 10.1038/nmeth.4108.
- Keller P.J., Schmidt A.D., Wittbrodt J., Stelzer E.H.K. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science*. 2008;322:1065-1069. DOI 10.1126/science.1162493.
- Kretzschmar K., Watt F.M. Lineage tracing. *Cell*. 2012;148:33-45. DOI 10.1016/j.cell.2012.01.002.
- Laukkanen M.O., Kuramoto K., Calmels B., Takatoku M., Von Kalle C., Donahue R.E., Dunbar C.E. Low-dose total body irradiation causes clonal fluctuation of primate hematopoietic stem and progenitor cells. *Blood*. 2005;105:1010-1015. DOI 10.1182/blood-2004-04-1498.
- Lawson K.A., Meneses J.J., Pedersen R.A. Cell fate and cell lineage in the endoderm of the presomite mouse embryo, studied with an intracellular tracer. *Dev. Biol.* 1986;115:325-339. DOI 10.1016/0012-1606(86)90253-8.
- Levin I., Kromer B. The tropospheric ¹⁴CO₂ level in mid-latitudes of the Northern Hemisphere (1959–2003). *Radiocarbon*. 2004;46:1261-1272. DOI 10.2458/azu_js_rc.46.4181.
- Lichtman J.W., Livet J., Sanes J.R. A technicolour approach to the connectome. *Nat. Rev. Neurosci.* 2008;9:417-422. DOI 10.1038/nrn2391.
- Livet J., Weissman T.A., Kang H., Draft R.W., Lu J., Bennis R.A., Sanes J.R., Lichtman J.W. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*. 2007;450:56-62. DOI 10.1038/nature06293.
- Lu R., Neff N.F., Quake S.R., Weissman I.L. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* 2011;29:928-933. DOI 10.1038/nbt.1977.
- Maetzig T., Brugman M.H., Bartels S., Heinz N., Kustikova O.S., Modlich U., Li Z., Galla M., Schiedlmeier B., Schambach A., Baum C. Polyclonal fluctuation of lentiviral vector-transduced and expanded murine hematopoietic stem cells. *Blood*. 2011;117:3053-3064. DOI 10.1182/blood-2010-08-303222.
- Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M. RNA-guided human genome engineering via Cas9. *Science*. 2013;339:823-826. DOI 10.1126/science.1232033.
- Mazurier F., Gan O.I., McKenzie J.L., Doedens M., Dick J.E. Lentivector-mediated clonal tracking reveals intrinsic heterogeneity in the human hematopoietic stem cell compartment and culture-induced stem cell impairment. *Blood*. 2004;103:545-552. DOI 10.1182/blood-2003-05-1558.
- Mazzarello P. A unifying concept: the history of cell theory. *Nat. Cell Biol.* 1999;E13-E15. DOI 10.1038/8964.
- McKenna A., Findlay G.M., Gagnon J.A., Horwitz M.S., Schier A.F., Shendure J. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*. 2016;353:aaf7907. DOI 10.1126/science.aaf7907.
- Naik S.H., Perié L., Swart E., Gerlach C., van Rooij N., de Boer R.J., Schumacher T.N. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature*. 2013;496:229-232. DOI 10.1038/nature12013.

- Naik S.H., Schumacher T.N., Perié L. Cellular barcoding: A technical appraisal. *Exp. Hematol.* 2014. DOI 10.1016/j.exphem.2014.05.003.
- Nguyen L.V., Cox C.L., Eirew P., Knapp D.J., Pellacani D., Kannan N., Carles A., Moksa M., Balani S., Shah S., Hirst M., Aparicio S., Eaves C.J. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nat. Commun.* 2014a;5:5871. DOI 10.1038/ncomms6871.
- Nguyen L.V., Makarem M., Carles A., Moksa M., Kannan N., Pandoh P., Eirew P., Osako T., Kardel M., Cheung A.M.S., Kennedy W., Tse K., Zeng T., Zhao Y., Humphries R.K., Aparicio S., Eaves C.J., Hirst M. Clonal analysis via barcoding reveals diverse growth and differentiation of transplanted mouse and human mammary stem cells. *Cell Stem Cell.* 2014b;1-11. DOI 10.1016/j.stem.2013.12.011.
- Nydal R., Lovseth K. Distribution of radiocarbon from nuclear tests. *Nature.* 1965;206:1029-1031.
- Palmer T.D., Takahashi J., Gage F.H. The adult rat hippocampus contains primordial neural stem cells. *Mol. Cell Neurosci.* 1997;8:389-404. DOI 10.1006/mcne.1996.0595.
- Perli S.S.D., Cui C.H., Lu T.K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science.* 2016;353:aag0511. DOI 10.1126/science.aag0511.
- Rosenquist G.C. Location and movements of cardiogenic cells in the chick embryo: The heart-forming portion of the primitive streak. *Dev. Biol.* 1970;22:461-475. DOI 10.1016/0012-1606(70)90163-6.
- Schepers K., Swart E., van Heijst J.W.J., Gerlach C., Castrucci M., Sie D., Heimerikx M., Velds A., Kerkhoven R.M., Arens R., Schumacher T.N.M. Dissecting T cell lineage relationships by cellular barcoding. *J. Exp. Med.* 2008;205:2309-2318. DOI 10.1084/jem.20072462.
- Schmidt S.T., Zimmermann S.M., Wang J., Kim S.K., Quake S.R. Cell lineage tracing using nuclease barcoding. *arXiv.* 2016;1606.00786.
- Schroeder T. Long-term single-cell imaging of mammalian stem cells. *Nat. Methods.* 2011;8:S30-S35. DOI 10.1038/nmeth.1577.
- Serbedzija G.N., Bronner-Fraser M., Fraser S.E. A vital dye analysis of the timing and pathways of avian trunk neural crest cell migration. *Development.* 1989;106:809-816.
- Smirnov A.V., Yunusova A.M., Lukyanchikova V.A., Battulin N.R. CRISPR/Cas9: a universal tool for genomic engineering. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2016;20(4):493-510. DOI 10.18699/VJ16.175. (in Russian)
- Spalding K.L., Bhardwaj R.D., Buchholz B.A., Druid H., Frisén J. Retrospective birth dating of cells in humans. *Cell.* 2005;122:133-143. DOI 10.1016/j.cell.2005.04.028.
- Stewart M.H., Bendall S.C., Levadoux-Martin M., Bhatia M. Clonal tracking of hESCs reveals differential contribution to functional assays. *Nat. Methods.* 2010;7:917-922. DOI 10.1038/nmeth.1519.
- Sulston J.E., Horvitz H.R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* 1977;56:110-156. DOI 10.1016/0012-1606(77)90158-0.
- Sulston J., Schierenberg E., White J., Thomson J. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* 1983;100:64-119. DOI 10.1016/0012-1606(83)90201-4.
- Till J.E., McCulloch E.A. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.* 1961;175:145-149. DOI 10.1667/RRXX28.1.
- Van Heijst J.W.J., Gerlach C., Swart E., Sie D., Nunes-Alves C., Kerkhoven R.M., Arens R., Correia-Neves M., Schepers K., Schumacher T.N.M. Recruitment of antigen-specific CD8+ T cells in response to infection is markedly efficient. *Science.* 2009;325:1265-1269. DOI 10.1126/science.1175455.
- Verovskaya E., Broekhuis M.J.C., Zwart E., Ritsema M., van Os R., de Haan G., Bystrykh L.V. Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. *Blood.* 2013;122:523-532. DOI 10.1182/blood-2013-01-481135.
- Verovskaya E., Broekhuis M.J.C., Zwart E., Weersing E., Ritsema M., Bosman L.J., Poele T. van, Haan G. de, Bystrykh L.V. Asymmetry in skeletal distribution of mouse hematopoietic stem cell clones and their equilibration by mobilizing cytokines. *J. Exp. Med.* 2014;211:487-497. DOI 10.1084/jem.20131804.
- Vogt W. Gestaltungsanalyse am amphibienkeim mit örtlicher vitalfärbung. II. Teil gastrulation und mesodermbildung bei urodelen und anuren. *Wilhelm Roux Arch. Entwicklungsmech. Org.* 1929;120:384-706.
- Weisblat D.A., Zackson S.L., Blair S.S., Young J.D. Cell lineage analysis by intracellular injection of fluorescent tracers. *Science.* 1980;209:1538-1541.
- Wilson E.B. The cell-lineage of *Nereis*. A contribution to the cytogeny of the annelid body. *J. Morphol.* 1892;6(3):361-480. DOI 10.1002/jmor.1050060301.
- Wu A.M., Till J.E., Siminovich L., McCulloch E.A. A cytological study of the capacity for differentiation of normal hemopoietic colony-forming cells. *J. Cell Physiol.* 1967;69:177-184. DOI 10.1002/jcp.1040690208.
- Wu A.M., Till J.E., Siminovich L., McCulloch E.A. Cytological evidence for a relationship between normal hemotopoietic colony-forming cells and cells of the lymphoid system. *J. Exp. Med.* 1968;127:455-464. DOI 10.1084/jem.127.3.455.
- Wu C., Li B., Lu R., Koelle S.J., Yang Y., Jares A., Krouse A.E., Metzger M., Liang F., Lor K., Wu C.O., Donahue R.E., Chen I.S.Y., Weissman I., Dunbar C.E. Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell.* 2014;14:486-499. DOI 10.1016/j.stem.2014.01.020.

Морфофизиологические эффекты инсерционного мутагенеза гена контактин 5 (*Cntn5*) у трансгенных мышей

А.В. Смирнов¹, Н.А. Феофанова^{1,3}, Г.В. Концевая¹, М.В. Анисимова¹, И.И. Ковригин², И.А. Серова¹, М.П. Мошкин^{1,4},
Л.А. Герлинская¹, Н.Р. Баттулин^{1,2}✉

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный
исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт фундаментальной и клинической
иммунологии», Новосибирск, Россия

⁴ Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский
Томский государственный университет», Томск, Россия

Технологии трансгенеза активно применяются в самых разных областях биологических исследований. Наиболее активно используемая методика получения трансгенных животных с помощью инъекции ДНК в пронуклеус зиготы может сопровождаться нарушением функции генов в месте интеграции трансгенной конструкции. В данной работе описаны морфофизиологические эффекты интеграции трансгенной конструкции, обеспечивающей наработку гранулоцит-макрофаг колониестимулирующего фактора человека в молоко в линии мышей (GM9), встройка конструкции у которых произошла в интрон гена контактин 5 (*Cntn5*). Было показано, что инсерция конструкции не привела к нокауту *Cntn5*. Однако у трансгенных животных изменяется транскрипция гена в сердце и почках по сравнению с животными дикого типа, а также меняется спектр транскриптов, что может свидетельствовать о нарушении регуляции сплайсинга гена *Cntn5*. Из литературных данных известно, что полиморфизмы *Cntn5* ассоциированы со склонностью к ожирению и с предрасположенностью к артериальной гипертензии. Мы исследовали основные параметры жирового обмена и сердечно-сосудистой деятельности у мышей, гомозиготных по инсерции трансгена в ген *Cntn5*, гетерозиготных животных и животных дикого типа. Проведенное фенотипирование показывает, что гомозиготные мыши имеют меньшую, чем особи дикого типа, массу тела. Причем весовая разница между генотипами определяется не столько отставанием мутантов в развитии скелета и мышц, образующих в сумме тощую массу, сколько существенно меньшим накоплением жира. У исследуемых линий были обнаружены статистически значимые различия по параметрам, характеризующим интенсивность кровообращения. Гомозиготные мыши превосходят особей дикого типа по значениям артериального давления, частоте сердечных сокращений и скорости кровотока в сосудах хвоста. Следует отметить, что гетерозиготные особи имеют промежуточные значения между мутантами и диким типом по всем измеренным морфофункциональным параметрам.

Ключевые слова: контактин 5; трансгенез; жировой обмен; инсерционный мутагенез.

Morphophysiological alterations caused by insertional mutagenesis of contactin 5 (*Cntn5*) gene in transgenic mice

A.V. Smirnov¹, N.A. Feofanova^{1,3}, G.V. Kontsevaya¹,
M.V. Anisimova¹, I.I. Kovrigin², I.A. Serova¹,
M.P. Moshkin^{1,4}, L.A. Gerlinskaya¹, N.R. Battulin^{1,2}✉

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia

⁴ Tomsk State University, Tomsk, Russia

Transgenesis has become a routine for modern biological studies. The most popular method for producing transgenic animals—pronuclear microinjection—frequently leads to host gene disruption due to a random transgene integration. In this paper, we report our analysis of morphophysiological parameters of the transgenic mouse line GM9, in which a transgene designed for milk-specific expression of the human granulocyte-macrophage colony-stimulating factor (GM-CSF) gene was integrated into the intron of the Contactin 5 gene (*Cntn5*). We studied *Cntn5* expression with RT-PCR and discovered that its expression in the brain, the primary organ of *Cntn5* activity, was unperturbed. However, transgenic animals had less *Cntn5* transcripts in other tissues such as the kidney and heart. In addition, we observed a decreased amount of splice variants of *Cntn5* exons that flank the transgene integration site. These data suggest that the transgene integration event might affect proper *Cntn5* splicing in some tissues. Publications exist that imply that some polymorphisms in the *Cntn5* gene are associated with obesity and arterial hypertension in humans. We evaluated core parameters of lipid metabolism and heart activity in mice homozygous and heterozygous for *Cntn5* mutation using wild-type animals as control. Our results uncovered that homozygous mutant mice have lower body weight than controls and that it is caused by slower accumulation of fat tissue. *Cntn5* mutants also exhibit abnormalities in blood circulation: homozygous *Cntn5* mutants are

characterized by a higher blood pressure and heart beat rate, as well as faster blood flow in the tail vessels. Heterozygous animals showed intermediate results for all of these parameters.

Key words: Contactin 5; transgenesis; lipid metabolism; insertional mutagenesis.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Смирнов А.В., Фефанова Н.А., Концевая Г.В., Анисимова М.В., Ковригин И.И., Серова И.А., Мошкин М.П., Герлинская Л.А., Баттулин Н.Р. Морфофизиологические эффекты инсерционного мутагенеза гена контактин 5 (*Cntn5*) у трансгенных мышей. Вавиловский журнал генетики и селекции. 2016;20(6):918-924. DOI 10.18699/VJ16.212

HOW TO CITE THIS ARTICLE:

Smirnov A.V., Feofanova N.A., Kontsevaya G.V., Anisimova M.V., Kovrigin I.I., Serova I.A., Moshkin M.P., Gerlinskaya L.A., Battulin N.R. Morphophysiological alterations caused by insertional mutagenesis of contactin 5 (*Cntn5*) gene in transgenic mice. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):918-924. DOI 10.18699/VJ16.212

В настоящее время трансгенез активно используется для решения широкого круга теоретических и прикладных задач биологии, медицины и сельского хозяйства. Несмотря на огромные достижения в этой области, получение трансгенных животных, стабильно экспрессирующих трансген, сопряжено с рядом проблем. Эффект положения встроенного трансгена и влияние генетического окружения, структура трансгена и его целостность, количество копий трансгена, его эпигенетические модификации, тип экспрессирующей ткани-мишени, эффекты присутствия гомологичных элементов в геноме – вот неполный перечень параметров, влияющих на экспрессию трансгена. Эти аспекты трансгенеза внимательно изучаются, поскольку напрямую связаны с уровнем экспрессии трансгена. Один из побочных эффектов трансгенеза – появление уникальных мутантных линий животных в случае, если интеграция трансгена нарушила функции какого-либо гена. Такие спонтанные нокауты или модификации генов, возникающие при встройке трансгенной ДНК в геном, являются важным источником информации о функциях генов, особенно если они слабо исследованы. В ранее выполненной нами работе были изучены функциональные характеристики генно-инженерной конструкции pGoatcasGMCSF, способной обеспечивать продукцию гранулоцит-макрофаг колоние-стимулирующего фактора человека в молоке трансгенных самок мышей. Среди четырех трансгенных линий была обнаружена одна, в которой сайт интеграции трансгена находился в интроне гена контактин-5 (*Cntn5*) на хромосоме 9 (Burkov et al., 2013). Известно, что интеграция трансгена в интрон может приводить к изменению активности гена (Gogliotti et al., 2011; Fu et al., 2016). Поэтому мы исследовали как повлияла интеграция трансгена на сплайсинг гена *Cntn5* у трансгенных мышей, а также провели оценку морфофункциональных характеристик, на которых могли сказаться изменения активности гена *Cntn5*.

Ген *Cntn5* входит в суперсемейство иммуноглобулинов, отвечающих за клеточное взаимодействие (Ogawa et al., 1996; Zuko et al., 2013). Характерной чертой контактинов является большой размер – порядка 1.2 М бп в случае *Cntn5*. Всего описано шесть схожих по структуре генов контактинов, которые экспрессируются в основном в нейронах головного (Kleijer et al., 2015) и спинного мозга (Ashrafi et al., 2014) и регулируют образование синапсов между нейронами, обеспечивая различные аспекты нервной деятельности. Контактин-5 экспрессируется главным образом в центральной нервной системе (Ogawa et al.,

1996), а его максимальная экспрессия обнаруживается у эмбрионов мышей начиная с 16-го дня и на 2-й день после рождения (Li et al., 2003). Зоны наибольшей экспрессии в мозге – таламус, кортекс, хвостатое ядро и скорлупа. Экспрессия *Cntn5* начинается на поздних стадиях эмбриогенеза и сохраняется после рождения. Единственный описанный нокаут по гену *Cntn5* у мышей был получен в лаборатории К. Watanabe, в которой была создана трансгенная линия со вставкой кассеты tau-LacZ-neo после первого кодирующего экзона *Cntn5*, что привело к полной инактивации гена (Li et al., 2003). Данные, опубликованные этой группой ученых, характеризуют влияние нокаута на деятельность нервной системы, в частности на формирование слуховой системы (Li et al., 2003). Психоневрологические эффекты генов контактинов находят подтверждение в популяционных исследованиях, которые показывают, что у людей полиморфизмы данной группы генов, включая и *Cntn5*, ассоциируются с расстройствами аутистического спектра (Fernandez et al., 2008; Morrow et al., 2008; Burbach, van der Zwaag, 2009; Roohi et al., 2009; Van Daalen et al., 2011; Zuko et al., 2011, 2013; Nava et al., 2013), с дефицитом внимания и гиперреактивностью (Lionel et al., 2011), нервной анорексией (Nakabayashi et al., 2009).

Вместе с тем установлено, что полиморфизмы *Cntn5* ассоциированы также со склонностью к ожирению и предрасположенностью к артериальной гипертензии (Nikpay et al., 2012). Эти эффекты могут иметь нейрогенную природу, но нельзя исключить и прямого влияния полиморфизмов *Cntn5* на функционирование внутренних органов, вовлеченных в процессы жирового обмена и гомеостатическое регулирование гемодинамики. Возможность такого эффекта следует из анализа баз данных (FANTOM5, GeneAtlas), согласно которым экспрессия *Cntn5* обнаруживается не только в мозге, но и во многих других тканях тела. Таким образом, полученная нами линия мышей GM9 с интеграцией трансгена в интрон гена контактин-5 на хромосоме 9 (Burkov et al., 2013) становится перспективным объектом для изучения особенностей экспрессии гена *Cntn5*, а также для анализа метаболических и гемодинамических последствий направленных изменений данного гена.

Методы

Животные. Работа выполнена на базе Центра генетических ресурсов лабораторных животных ФИЦ ИЦиГ СО РАН (RFMEFI61914X0005 и RFMEFI62114X0010).

Подопытных животных содержали в индивидуально вентилируемых клетках (OptiMice) при температуре 24 °C, фотопериоде 14C:10T, влажности 40–50 % и свободном доступе к воде и корму. Для кормления использовали брикетированные корма SNIFF (Германия) и ЧАРА, («Ассортимент – Агро», Россия) в соотношении 50:50, деионизированную воду обогащали солями в соответствии с требованиями к воде, используемой для поения SPF-мышей. Подстилочным материалом служили обеспыленная березовая крошка (ООО Альбион, Новосибирск). Все подопытные животные были свободны от видоспецифических патогенов (specific pathogen free – SPF).

В работе были использованы трансгенные мыши B6;CB-Cntn5^{Tg(gCsn1s1-hCsf2)lcr} (номенклатура по (Montoliu, Whitelaw, 2011)), несущие интеграцию генетической конструкции pGoatcasGMCSF в гене *Cntn5* (Burkov et al., 2013), далее будет использовано краткое обозначение линии – GM9.

Животных для фенотипирования получали скрещиванием мышей линий C57BL/6 (B6) и GM9. От 32 пар было получено 278 потомков, которые после генотипирования были разделены на три группы: мыши дикого типа (B6), гетерозиготы (B6/GM9) и гомозиготы (GM9). В возрасте трех недель животных взвешивали и рассаживали в клетки однополыми группами по пять-шесть особей. Фенотипирование проводили на самцах в возрасте 11–13 нед. Исследования выполняли в светлое время суток с 10:00 до 16 ч.

ПЦР для определения трансгенных животных. Генотипирование трансгенных мышей проводили с помощью полимеразной цепной реакции. Кончики хвостов мышей 1–3 мм лизировали в 100 мкл буфера PBNB (50 mM KCl; 20 mM Tris-HCl, pH 8.4; 2.5 mM MgCl₂; 0.45 % v/v Nonidet P40 (NP40); 0.45 % v/v Tween 20) с добавлением протеиназы К (конечная концентрация – 1 мкг/мл). Пробирки инкубировали при температуре 56 °C в течение 12–16 ч с последующей инактивацией протеиназы (95 °C, 10 мин). Реакционная смесь для ПЦР объемом 25 мкл содержала 1× ПЦР буфер (65 mM Tris-HCl, pH 8.8, 16 mM (NH₄)₂SO₄ и 0.01 % Tween 20), 1.5 mM MgCl₂, 0.2 mM каждого дезоксинуклеотида (dATP, dCTP, dGTP, dTTP) и по 0.4 мкМ каждого из праймеров. В реакционную смесь добавляли 0.5 ед. Taq ДНК полимеразы и 1 мкл выделенной геномной ДНК. Условия ПЦР включали денатурацию геномной ДНК при 95 °C в течение 3 мин, с последующей амплификацией (35 циклов): денатурация при 95 °C – 30 с, отжиг праймеров при температуре 60 °C – 30 с, элонгация при 72 °C – 30 с. После стадии амплификации следовала финальная элонгация при 72 °C в течение 3 мин. ПЦР проводили на автоматическом амплификаторе T100 Thermal Cycler (Bio-Rad, США). Последовательности праймеров были следующие: праймер mTg (трансген) – 5'-CTGA CACTACTTCCACCTTCTCCCC-3'; праймер 1F (аллель дикого типа) – 5'-ATCTTGCTTTTCGGGAGCGT-3'; праймер 4R (участок генома рядом с 5'-концом трансгена) – 5'-TTGGTACACAATGCCTGCAA-3' (Биоссет, Новосибирск).

Выделение РНК и ОТ-ПЦР на кДНК *Cntn5*. РНК была выделена из целых органов гомозиготной трансгенной и нетрансгенной мышей (головной мозг, почка, сердце) линии C57BL/6. В дополнение к этому была

выделена РНК из культуры ЭС клеток мыши (линия 129/Sv). Для выделения РНК использовали TRI reagent (Sigma-Aldrich, США). Небольшие кусочки органов гомогенизировали в стеклянных гомогенизаторах Даунса и обрабатывали 500 мкл TRI reagent. Выделение РНК проводили по протоколу производителя (<http://www.sigmaaldrich.com/technical-documents/protocols/biology/tri-reagent.html>). После необходимых этапов выделения и очистки осадок РНК растворяли в 100 мкл mQ воды. Раствор РНК хранили при –80 °C.

Для синтеза кДНК использовали набор реактивов RevertAid First Strand cDNA Synthesis Kit (ThermoFisher, США). 1 мкг РНК добавляли в реакционную смесь (20 мкл), содержащую 1× буфер (50 mM Tris-HCl, pH 8.3; 50 mM KCl, 4 mM MgCl₂, 10 mM DTT), 1 mM каждого дезоксинуклеотида (дАТФ, дЦТФ, дГТФ, дТТФ), 1 ед. ингибитора РНКазы (RiboLock RNase Inhibitor), 5 мкМ вырожденных праймеров (random hexamer primers), 200 ед. обратной транскриптазы (RevertAid M-MuLV RT). Условия реакции обратной транскрипции: 5 мин при 25 °C, 1 ч при 42 °C, 5 мин при 70 °C.

ОТ-ПЦР-анализ кДНК *Cntn5* проводили с использованием праймеров из статьи (Namba et al., 2013). Последовательности праймеров: 5'-TGGATTCTTCACTTGATTC-3' (F) и 5'-CACCCCTCTTGCCGCAG-3' (R). Эти праймеры комплементарны экзонам 8–10 *Cntn5* (если считать от первого кодирующего экзона). Для анализа транскрипции также были использованы праймеры на экзоны до интеграции трансгена (экзоны 1 и 2), а также на экзоны после сайта интеграции (экзоны 3 и 4). Последовательности этих праймеров: 5'-CCAGAGCGGCCTGTGACTA-3' (2F); 5'-AGTGGTACTGCTGAGCTGCC-3' (2R); 5'-CTGGCCTTCCCGCACTGAGA-3' (3R). Состав ОТ-ПЦР реакции был такой же, как и при ПЦР для генотипирования трансгенных мышей (см. выше). В реакцию брали 1 мкл кДНК. Условия ОТ-ПЦР: исходная денатурация при 95 °C в течение 3 мин, с последующей амплификацией (35 циклов), денатурация при 95 °C – 30 с, отжиг праймеров при температуре 58 °C, элонгация при 72 °C – 1 мин. Финальная элонгация после окончания амплификации – 3 мин.

Качество выделенной кДНК проверяли с помощью ПЦР с использованием праймеров на мРНК гена *β-актина* мыши. Последовательности использованных праймеров: Actbf 5'-ACGCACGATTTCCCTCTCAGC-3', Actbr 5'-GGCCAGAGCAAGAGAGGTATCC-3'.

Измерение массы тела и содержания жира. Самцы – 16 (+/+), 15 (–/+) и 16 (–/–) генотипов были взвешены в возрасте 3 нед в момент отъема от матерей и в возрасте 11–13 нед в день измерения содержания жира. Содержание жира и тощую массу определяли методом магниторезонансной спектроскопии с помощью низкочастотного ЯМР спектрометра (EchoMRI, США).

Измерение показателей кровообращения. У самцов генотипов B6 (15), B6/GM9 (16) и GM9 (16) в возрасте 11–13 нед были измерены систолическое и диастолическое давление, частота сердечных сокращений и объемная скорость кровотока в сосудах хвоста. Измерения выполнены на ненаркотизированных животных с помощью системы неинвазивного определения перечисленных выше

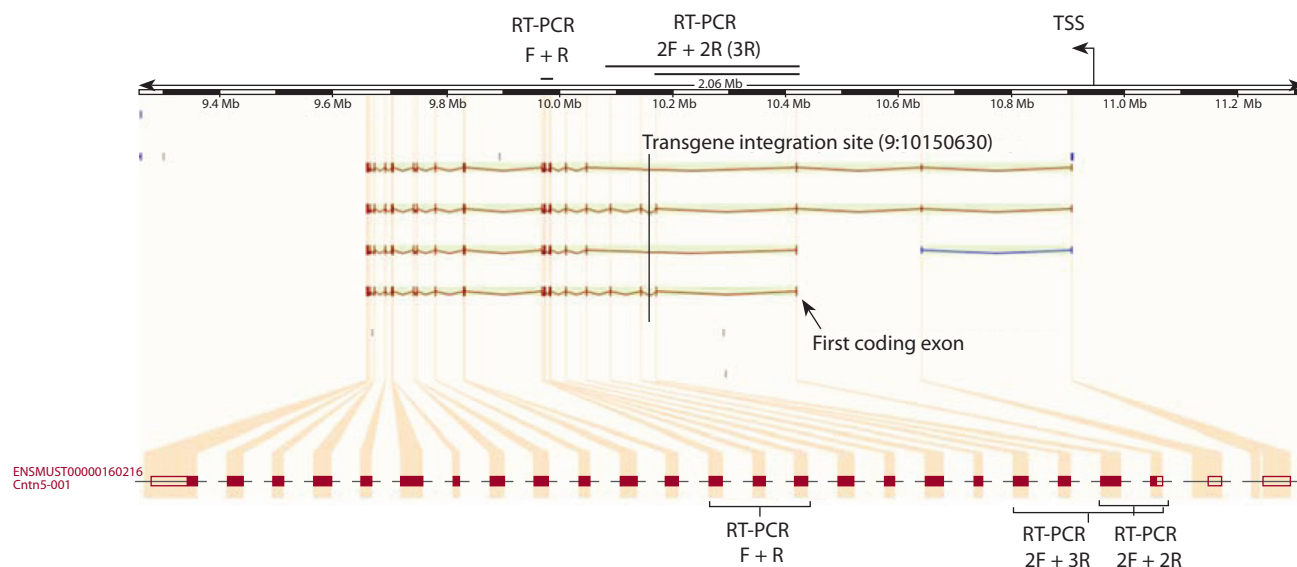


Fig. 1. Schematic representation of the murine *Cntn5* gene (ENSUSG0000039488, ensembl.org). The transgene integration site and primers used for RT-PCR analysis are indicated. TSS, transcription start site.

гемодинамических параметров (CODA, Kent Scientific Corporation, США). За 5 мин до измерения животное помещали в ограничивающий движения пенал (рестриктор), который находился на подогреваемом столике (37 °C). За каждый сеанс проводили 20 измерений. Качество измерений контролировали по графикам. Значения, искаженные помехами, обусловленными подвижностью животных в рестрикторе или движениями хвоста, исключали из дальнейшего анализа. Для привыкания к процедуре измерения каждое животное тестировали трижды – по одному сеансу в день. При анализе результатов использовали данные последнего измерения.

Статистика. Для определения статистической значимости различий между исследуемыми генотипами использовали однофакторный дисперсионный анализ и метод множественного сравнения средних (LSD-тест). Значения выражали как среднее $\pm$ SE.

Результаты

Генотипирование трансгенных мышей линии GM9. У мышей линии GM9 (Burkov et al., 2013) трансген *pGoatcasGMCSF* был встроен в интрон гена *Cntn5* на хромосоме 9 (9:10150630) (рис. 1). Сайт интеграции расположен внутри интрона размером примерно 29 т. п. н., при этом размер встройки предположительно составляет около 14 т. п. н. Для генотипирования были подобраны праймеры на участок между 5'-границей трансгена и геном (трансгенный аллель), а также на участок интрона гена *Cntn5* (нормальный, неповрежденный аллель). Некоторые затруднения создал тот факт, что трансген встроился в ретротранспозон (LINE/L1), который имеет гомологию порядка 90 % с другими родственными последовательностями в геноме, поэтому праймеры для генотипирования подбирались с учетом мутаций в конкретном мобильном элементе. Сначала образцы ДНК из хвостов мышей анализировали на наличие трансгенного аллеля (праймеры mTg + 4R), а затем всех животных, позитивных по встрой-

ке трансгена, также проверяли на наличие у них аллеля *Cntn5* дикого типа (праймеры 1F + 4R) (рис. 2).

ОТ-ПЦР анализ экспрессии гена *Cntn5* в трансгенных мышах. Ожидалось, что интеграция большого фрагмента экзогенной ДНК (14 т. п. н.) в сопоставимый по размеру интрон гена *Cntn5* (29 т. п. н.) повлияет на экспрессию гена. Однако предварительный анализ транскрипции гена *Cntn5* показал, что встройка трансгена не привела к исчезновению транскриптов гена *Cntn5* в мозге (рис. 3, экзоны 8–10 и 1–2). Ранее экспрессия *Cntn5* была описана лишь для головного и спинного мозга. Интересно, что мы детектировали транскрипты *Cntn5* в почках и сердце (но не в эмбриональных стволовых клетках), при этом экспрессия гена в сердце была снижена. Для более детального анализа были выбраны праймеры на экзоны *Cntn5*, которые расположены возле сайта интеграции и подвергаются альтернативному сплайсингу (см. рис. 1) (NM_001170787.1, NM_001033359.2). Мы подтвердили, что мРНК *Cntn5* присутствует в мозге трансгенных животных, но количество изоформ у мутантных мышей было ниже (рис. 3). В то же время данный участок мРНК *Cntn5* отсутствовал в почках и сердце трансгенных животных. Для более точного определения нарушений сплайсинга и транскрипции гена *Cntn5* в трансгенных мышах линии GM9 потребуются дополнительные эксперименты.

Масса тела и жировой обмен. Однофакторный дисперсионный анализ показал значимое влияние генотипа на массу тела самцов в возрасте 3 нед ($F_{2,46} = 7.53$, $p = 0.0015$). Самцы мутантного (GM9) генотипа имели меньшую 10.3 ± 0.35 массу тела по сравнению с самцами дикого типа (B6) 12.3 ± 0.37 , $p = 0.0005$ и гетерозиготами (B6/GM9) 11.8 ± 0.40 , $p = 0.008$ (LSD-тест). Различия по массе тела между генотипами B6 и GM9 сохранялись и в возрасте 11–13 нед (рис. 4).

Самцы мутантного генотипа (GM9) в этом возрасте имели меньшую по сравнению с диким генотипом (B6) массу тела, а гетерозиготы занимали промежуточное по-

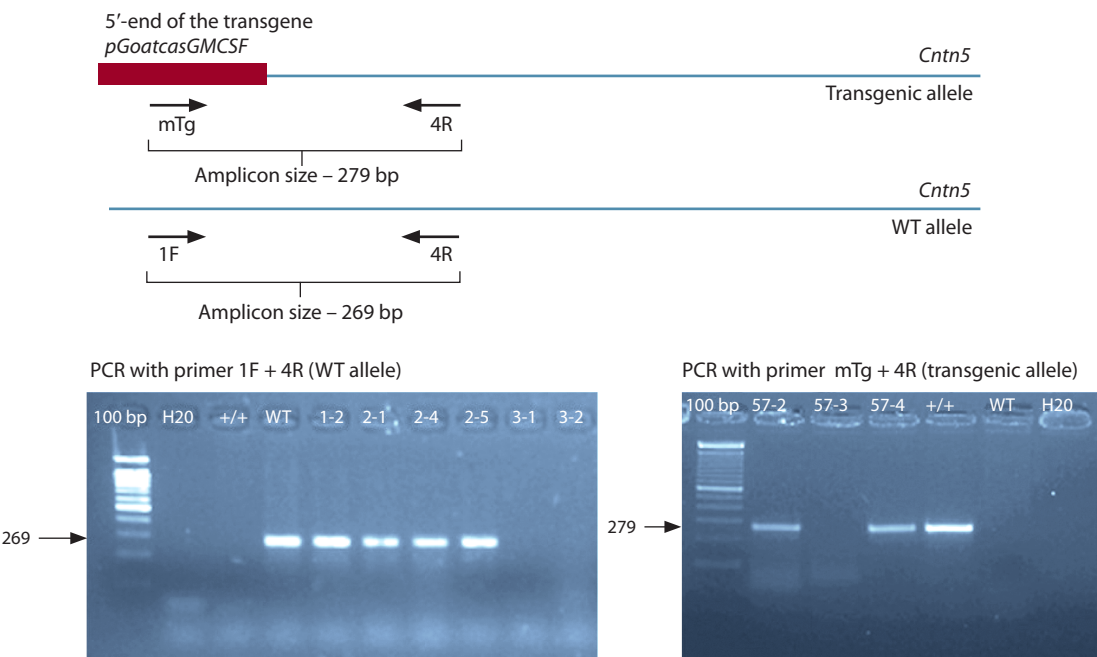


Fig. 2. Genotyping of transgenic mice. The primer arrangement and PCR results for some DNA samples from the tail tips of analyzed animals are shown. PCR with primers mTg and 1F was used to detect *Cntn5* with transgene integration, while primer pair 1F и 4R amplifies the wild-type allele. 100bp, DNA ladder; +/+, homozygote for the transgene; WT, wild-type mice; H20, negative control.

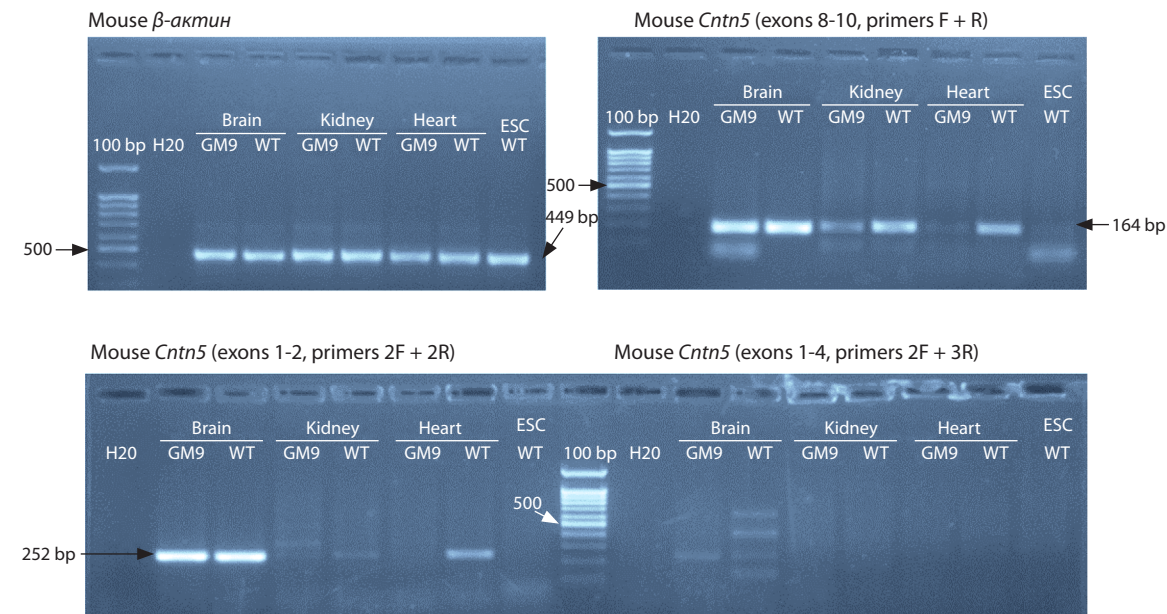


Fig. 3. RT-PCR detection of *Cntn5* mRNA in various tissues of transgenic (GM9, homozygote) and nontransgenic (WT) mice. 100bp, DNA ladder; ESC, mouse embryonic stem cells.

ложение. Исследованные генотипы различались между собой и по содержанию жира. Самцы B6 достоверно отличались от GM9 по средним значениям абсолютной и относительной массы жира (см. рис. 4). Абсолютное и относительное содержание жира у гетерозигот значимо не отличалось от значений, характерных для B6 и GM9. Абсолютные величины тощей массы достоверно не различались у самцов исследованных генотипов, тогда как

относительная доля тощей массы была максимальной у самцов GM9.

Сердечно-сосудистая система. Результаты межгрупповых сравнений (LSD-тест) показали, что систолическое и диастолическое давление, частота сердечных сокращений и объемная скорость кровотока у самцов мутантного генотипа (GM9) были выше, чем у дикого генотипа (B6). Гетерозиготы по всем исследованным показателям сер-

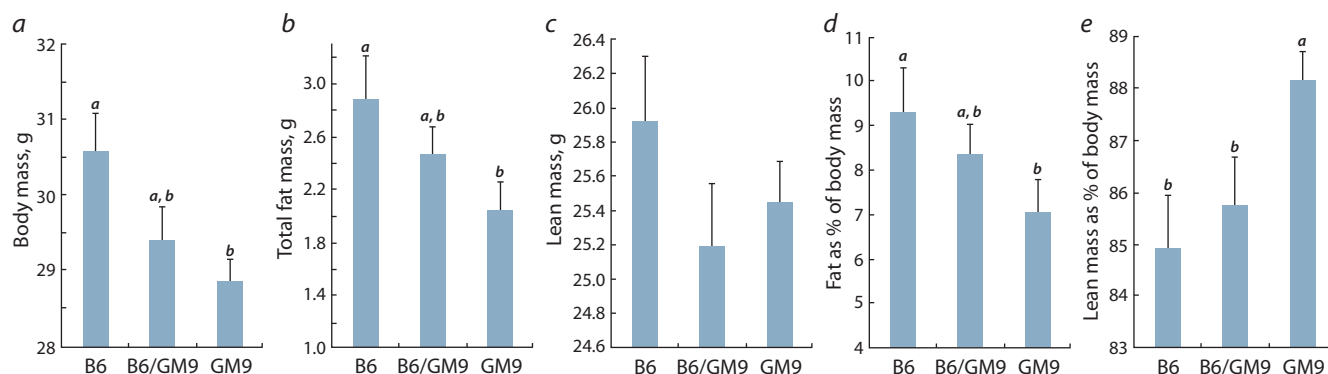


Fig. 4. Effect of the knockout of the *Cntn5* gene on body weight and composition in mice of different genotypes at the age of 11–13 weeks: (a) body mass; (b) total fat mass; (c) lean mass; (d) fat as % of body mass; (e) lean mass as % of body mass.

Here and in Fig. 5: B6, wild type; B6/GM9, heterozygotes; and GM9, homozygotes. Different letters above bars (a; a, b; b) denote significantly different means, a differing from b at $p < 0.05$, LSD test.

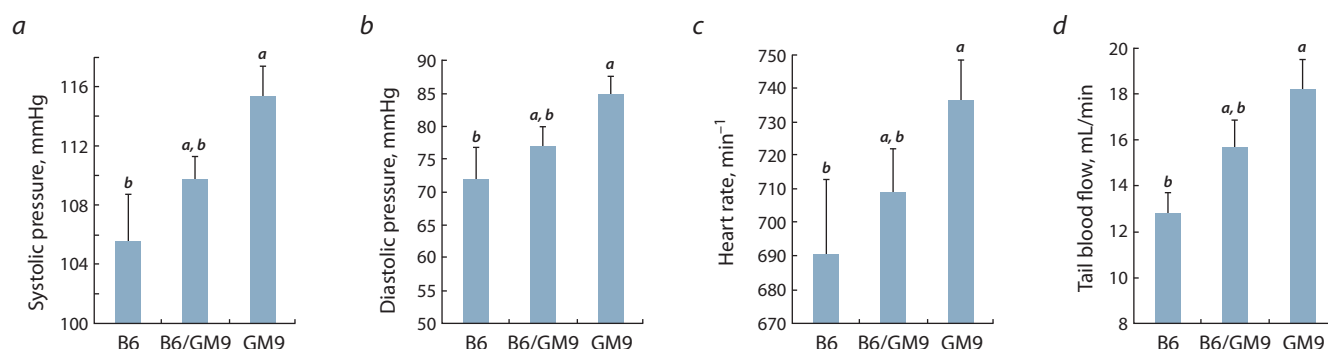


Fig. 5. Effect of the knockout of the *Cntn5* gene on circulation indices in mice of different genotypes at the age of 11–13 weeks: (a) systolic pressure; (b) diastolic pressure; (c) heart rate; (d) tail blood flow.

дечно-сосудистой деятельности не отличались от особей дикого и мутантного генотипов (рис. 5).

Обсуждение

Как было описано ранее, у трансгенных мышей линии GM9 трансген *pGoatcasGMCSF*, встроенный внутри гена контактин 5, функционировал так же, как и при интегрировании в другие области генома. Было показано, что у линии GM3 трансген локализован в гене *Prkcb1* (хромосома 7), у линии GM11 – в некодирующем участке хромосомы 9, а у линии GM20 – в мобильном SINE-элементе хромосомы 6. Это свидетельствует о том, что место интеграции изучаемых трансгенов в реципиентный геном не оказывает существенного влияния на экспрессию исследованных генно-инженерных конструкций (Burkov et al., 2013). Важно отметить отсутствие эктопической экспрессии трансгена в линии GM9 и, соответственно, влияния рекомбинантного белка GM-CSF на гемопоэтические и другие физиологические показатели у трансгенных животных (Burkov et al., 2013). Линия трансгенных мышей GM9 может быть интересна для изучения роли сплайсинга в функции гена *Cntn5* в мозге и других тканях тела. Мы показали, что интеграция трансгена в интрон *Cntn5* приводит к нарушению сплайсинга центральных

экзонов, следствием чего, возможно, является и снижение общего количества транскриптов в тканях тела, что было обнаружено для сердца. Для выяснения этих вопросов необходимо провести анализ экспрессии различных изоформ *Cntn5* в специфических отделах мозга и других органов животных.

Функциональная значимость изоформ контактина-5 становится предметом комплексного изучения лишь в последнее время. Тем не менее проведенное нами фенотипирование мышей линии GM9 показывает, что генно-инженерное изменение экспрессии гена *Cntn5* отражается на количественных характеристиках таких физиологически значимых процессов, как накопление жира и функционирование сердечно-сосудистой системы. Мутантные мыши (GM9) характеризуются меньшей по сравнению с особями дикого типа (B6) массой тела. Эти отличия, выраженные в момент отъема от матерей, сохраняются и у взрослых особей. Причем весовая разница между генотипами определяется не столько отставанием мутантов в развитии скелета и мышц, образующих в сумме тощую массу, сколько существенно меньшим накоплением жира. Исследуемые линии показывают также статистически значимые различия по параметрам, характеризующим интенсивность кровообращения. Значения

артериального давления у мышей дикого типа (B6) соответствуют величинам, характерным для данной линии (Mattson, 2001). Мутантные мыши (GM9) превосходят особей дикого типа (B6) по значениям артериального давления, частоте сердечных сокращений и скорости кровотока в сосудах хвоста. Следует отметить, что гетерозиготные особи показывают промежуточные значения между мутантами и диким типом по всем измеренным морфофункциональным параметрам.

Проведенное исследование является реализацией принципа «черного ящика» в приложении к расшифровке функциональной значимости как отдельных генов, так и генных сетей, в формировании которых эти гены участвуют. Наши результаты показывают, что инсерционный мутагенез гена контактин 5 дает целый спектр фенотипических эффектов. Эти эффекты, а именно изменения липидного обмена и параметров гемодинамики, согласуются с эпидемиологическими наблюдениями, в которых отмечаются ассоциации полиморфизмов гена *Cntn5* со склонностью людей к ожирению и с предрасположенностью к артериальной гипертензии (Nikraу et al., 2012). Высокая социальная значимость данных полиэтиологических патологий (ожирение и гипертензия) и недостаточное понимание их генетической детерминации указывают на необходимость детализации процессов, предопределяющих ассоциации индивидуальных вариаций экспрессии гена *Cntn5* с предрасположенностью к заболеваниям. Принципиально важно, что для продвижения в данном направлении, которое имеет не только теоретическую, но и прикладную значимость, создана доступная для исследований новая линия животных с целевой мутацией гена контактин-5.

Acknowledgments

The experiments and result processing were supported by the Russian Science Foundation, project 16-14-00095. The purchase of laboratory animals and equipment rental were supported by State Budgeted Project 0324-2015-0004.

Conflict of interest

The authors declare no conflict of interest.

References

- Ashrafi S., Betley J.N., Comer J.D., Brenner-Morton S., Bar V., Shimoda Y., Watanabe K., Peles E., Jessell T.M., Kaltschmidt J.A. Neuronal Ig/Caspr recognition promotes the formation of axoaxonic synapses in mouse spinal cord. *Neuron*. 2014;81(1):120-129.
- Burbach J.P.H., van der Zwaag B. Contact in the genetics of autism and schizophrenia. *Trends Neurosci*. 2009;32:69-72.
- Burkov I., Serova I., Battulin N., Smirnov A., Babkin I., Andreeva L., Dvoryanchikov G., Serov O. Expression of the human granulocyte-macrophage colony stimulating factor (hGM-CSF) gene under control of the 5'-regulatory sequence of the goat alpha-S1-casein gene with and without a MAR element in transgenic mice. *Transgenic Res*. 2013;22(5):949-964.
- Gogliotti R.G., Lutz C., Jorgensen M., Huebsch K., Koh S., Didonato C.J. Characterization of a commonly used mouse model of SMA reveals increased seizure susceptibility and heightened fear response in FVB/N mice. *Neurobiol. Dis*. 2011;43(1):142-151. DOI 10.1016/j.nbd.2011.03.002.
- Fernandez T., Morgan T., Davis N., Klin A., Morris A., Farhi A., Lifton R.P., State M.W. Disruption of Contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *Am. J. Hum. Genet*. 2008;82:1385.
- Fu C., Begum K., Overbeek P.A. Primary ovarian insufficiency induced by Fanconi anemia E mutation in a mouse model. *PLoS ONE*. 2016;11(3):e0144285. DOI 10.1371/journal.pone.0144285.
- Kleijer K.T., Zuko A., Shimoda Y., Watanabe K., Burbach J.P. Contactin-5 expression during development and wiring of the thalamocortical system. *Neuroscience*. 2015;310:106-113.
- Li H., Takeda Y., Niki H., Ogawa J., Kobayashi S., Kai N., Akasaka K., Asano M., Sudo K., Iwakura Y., Watanabe K. Aberrant responses to acoustic stimuli in mice deficient for neural recognition molecule NB-2. *Eur. J. Neurosci*. 2003;17(5):929-936.
- Lionel A.C., Crosbie J., Barbosa N., Goodale T., Thiruvahindrapuram B., Rickaby J., Gazzellone M., Carson A.R., Howe J.L., Wang Z., Wei J., Stewart A.F., Roberts R., McPherson R., Fiebig A., Franke A., Schreiber S., Zwaigenbaum L., Fernandez B.A., Roberts W., Arnold P.D., Szatmari P., Marshall C.R., Schachar R., Scherer S.W. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci. Transl. Med*. 2011;3(95):95ra75. DOI 10.1126/scitranslmed.3002464.
- Mattson D.L. Comparison of arterial blood pressure in different strains of mice. *Am. J. Hypertens*. 2001;14:405-408.
- Morrow E.M., Yoo S.Y., Flavell S.W., Kim T.K., Lin Y., Hill R.S., Mukaddes N.M., Balkhy S., Gascon G., Hashmi A., Al-Saad S., Ware J., Joseph R.M., Greenblatt R., Gleason D., Ertelt J.A., Apse K.A., Bodell A., Partlow J.N., Barry B., Yao H., Markianos K., Ferland R.J., Greenberg M.E., Walsh C.A. Identifying autism loci and genes by tracing recent shared ancestry. *Science*. 2008;321:218-223.
- Nakabayashi K., Komaki G., Tajima A., Ando T., Ishikawa M., Nomoto J., Hata K., Oka A., Inoko H., Sasazuki T., Shirasawa S. Identification of novel candidate loci for anorexia nervosa at 1q41 and 11q22 in Japanese by a genome-wide association analysis with microsatellite markers. *J. Hum. Genet*. 2009;54:531-537.
- Nava C., Keren B., Mignot C., Rastetter A., Chantot-Bastaraud S., Faudet A., Fonteneau E., Amiet C., Laurent C., Jacqueline A., Whalen S., Afenjar A., Périsse D., Doummar D., Dorison N., Leboyer M., Sifroi J.P., Cohen D., Brice A., Héron D., Depienne C. Prospective diagnostic analysis of copy number variants using SNP microarrays in individuals with autism spectrum disorders. *Eur. J. Hum. Genet*. 2013;975:1-8.
- Nikpay M., Seda O., Tremblay J., Petrovich M., Gaudet D., Kotchen T.A., Cowley A.W., Hamet P. Genetic mapping of habitual substance use, obesity-related traits, responses to mental and physical stress, and heart rate and blood pressure measurements reveals shared genes that are overrepresented in the neural synapse. *Hypertens. Res*. 2012;35:585-591.
- Ogawa J., Kaneko H., Masuda T., Nagata S., Hosoya H., Watanabe K. Novel neural adhesion molecules in the Contactin/F3 subgroup of the immunoglobulin superfamily: isolation and characterization of cDNAs from rat brain. *Neurosci. Lett*. 1996;218(3):173-176.
- Roohi J., Montagna C., Tegay D.H., Palmer L.E., DeVincent C., Pomeroy J.C., Christian S.L., Nowak N., Hatchwell E. Disruption of contactin 4 in three subjects with autism spectrum disorder. *J. Med. Genet*. 2009;46:176-182.
- Van Daalen E., Kemner C., Verbeek N.E., van der Zwaag B., Dijkhuizen T., Rump P., Houben R., van't Slot R., de Jonge M.V., Staal W.G., Beemer F.A., Vorstman J.A.S., Burbach J.P.H., van Amstel H.K.P., Hochstenbach R., Brilstra E.H., Poot M. Social responsiveness scale-aided analysis of the clinical impact of copy number variations in autism. *Neurogenetics*. 2011;12:315-323.
- Zuko A., Bouyain S., van der Zwaag B., Burbach J.P.H. Contactins: structural aspects in relation to developmental functions in brain disease. *Adv. Protein Chem. Struct. Biol*. 2011;84:143-180.
- Zuko A., Kleijer K.T., Oguro-Ando A., Kas M.J., van Daalen E., van der Zwaag B., Burbach J.P. Contactins in the neurobiology of autism. *Eur. J. Pharmacol*. 2013;719(1-3):63-74.



Эффективное получение химерных мышей с использованием новой линии эмбриональных стволовых клеток

Г.В. Концевая¹, Н.А. Феофанова^{1, 3}, А.Г. Мензоров^{1, 2}, И.Е. Пристяжнюк¹, А.В. Смирнов¹, Н.Р. Баттулин^{1, 2}✉, Л.А. Герлинская¹

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

³ Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт фундаментальной и клинической иммунологии», Новосибирск, Россия

Для получения трансгенных мышей широко используются эмбриональные стволовые клетки. При инъекции генетически модифицированных эмбриональных стволовых клеток в бластоцисту можно получить химерных животных, часть половых клеток которых будет содержать модифицированный участок генома или трансген. Такие мыши-основатели могут дать начало линиям с модификациями генома. В настоящее время запущено несколько проектов (KOMP Repository, EUCOMM, Lexicon Genetics) по созданию коллекций нокаутных по различным генам линий мышей. Тем не менее многие генно-инженерные задачи требуют сложных модификаций генома, таких как большие геномные делеции, встройка генов-репортеров в 3'-регуляторную область генов или сайт-специфичные мутации участков генома. Для этих целей требуется линия эмбриональных стволовых клеток мыши, клетки которой способны участвовать в формировании химерного животного даже после длительного культивирования. В лаборатории генетики развития Института цитологии и генетики СО РАН для проведения экспериментов по созданию трансгенных линий мышей было получено несколько линий эмбриональных стволовых клеток. Мы выбрали одну из них, DGE51 (нормальный кариотип $2n = 40, XY$, генотип 129S2/SvPasCrl), для получения химерных животных на базе Центра коллективного пользования «SPF-виварий» Института цитологии и генетики СО РАН. Эмбриональные стволовые клетки были введены в 136 бластоцист генотипа B6D2F1. Бластоцисты подсаживали мышам CD-1, всего родилось 66 мышат. Среди этих потомков 15 оказались химерами, 4 из них с химеризмом выше 80 %. Все родившиеся химеры были самцами и не имели отклонений в развитии. 10 из 15 химерных животных были фертильны. Анализ аллелей микросателлитов потомков химерных самцов показал вклад эмбриональных стволовых клеток DGE51 в формирование гамет. Таким образом, линию эмбриональных стволовых клеток DGE51 можно эффективно использовать для создания трансгенных линий мышей с бластоцистами генотипа B6D2F1 и мышей-реципиентов линии CD-1.

Ключевые слова: эмбриональные стволовые клетки; химерные мыши; плюрипотентность; трансгенез.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Концевая Г.В., Феофанова Н.А., Мензоров А.Г., Пристяжнюк И.Е., Смирнов А.В., Баттулин Н.Р., Герлинская Л.А. Эффективное получение химерных мышей с использованием новой линии эмбриональных стволовых клеток. Вавиловский журнал генетики и селекции. 2016;20(6):925-929. DOI 10.18699/VJ16.213

HOW TO CITE THIS ARTICLE:

Kontsevaya G.V., Feofanova N.A., Menzorov A.G., Pristazhnyuk I.E., Smirnov A.V., Battulin N.R., Gerlinskaya L.A. Efficient chimeric mouse production using a novel embryonic stem cell line. Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding. 2016;20(6):925-929. DOI 10.18699/VJ16.213

ORIGINAL ARTICLE

Received 03.11.2016

Accepted for publication 16.12.2016

© AUTHORS, 2016

✉ e-mail: battulin@bionet.nsc.ru

Efficient chimeric mouse production using a novel embryonic stem cell line

G.V. Kontsevaya¹, N.A. Feofanova^{1, 3}, A.G. Menzorov^{1, 2}, I.E. Pristazhnyuk¹, A.V. Smirnov¹, N.R. Battulin^{1, 2}✉, L.A. Gerlinskaya¹

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia

Embryonic stem cells are commonly used for generation of transgenic mice. Embryonic stem cells could participate in the development of chimeric animals after injection into a blastocyst. Injection of genetically modified embryonic stem cells could lead to germ line transmission of a transgene or genomic modification in chimeric mice. Such founders are used to produce transgenic lines of mice. There are several projects dedicated to production of knock-out mouse lines (KOMP Repository, EUCOMM, Lexicon Genetics). Nevertheless, there is a need for complex genome modifications, such as large deletions, reporter genes insertion into the 3' gene regulatory sequence, or site-specific modifications of the genome. To do that, researchers need an embryonic stem cell line that is able to participate in chimeric animal formation even after prolonged culture *in vitro*. Several lines of mouse embryonic stem cells were produced in the Laboratory of Developmental Genetics of the Institute of Cytology and Genetics SB RAS. We tested DGE51 cell line ($2n = 40, XY$) (129S2/SvPasCrl genetic background) for chimeric mice production at the Center for Genetic Resources of Laboratory Animals at ICG SB RAS. Embryonic stem cells were injected into 136 blastocysts (B6D2F1 genetic background), which were transplanted into CD-1 mice. Among 66 progeny, 15 were chimeric, 4 of which were more than 80 % chimeric judged by coat color. All chimeras were males without developmental abnormalities. 10 of 15 males were fertile. Microsatellite analysis of the progeny of chimeric mice revealed embryonic stem cell line DGE51 contribution to the gamete formation. Thus, a novel DGE51 embryonic stem cell line could be efficiently used for transgenic mouse production using B6D2F1 blastocysts and CD-1 recipients.

Key words: embryonic stem cells; chimeric mice; pluripotency; transgenesis.

Впервые эмбриональные стволовые (ЭС) клетки мыши были получены M.J. Evans и M.H. Kaufman (1981) путем изоляции клеток внутренней клеточной массы бластоцисты. ЭС клетки способны к самообновлению и дифференцировке *in vitro* в производные трех зародышевых листков (включая гаметы). При введении ЭС клеток в бластоцисту можно получить химерное животное, часть клеток которого будет иметь генотип ЭС клеток и, в том числе, формировать гаметы. Более того, при инъекции ЭС клеток в тетраплоидную бластоцисту возможно образование эмбриона, полностью сформированного из введенных ЭС клеток, так как тетраплоидные клетки не способны формировать эмбрион, но дифференцируются во внезародышевые ткани. Проведение генетических манипуляций с ЭС клетками с последующим субклонированием позволяет получать трансгенных мышей с желаемыми свойствами. С развитием новых методов геномной инженерии, таких как аденоассоциированные вирусы, транспозоны и сайт-направленные нуклеазы (ZFN, TALEN, CRISPR/Cas9), трансгенез обрел «новое дыхание». Стало возможным эффективное получение линий ЭС клеток с направленной встройкой интересующего исследователя гена, большими хромосомными перестройками или генами-репортерами развития (Wijshake et al., 2014).

Нужно отметить, что свойства линий ЭС клеток, полученных в разных лабораториях, индивидуальны. Основные параметры, по которым оценивают потенциал линий клеток, следующие: способность давать вклад в половые клетки при получении химерных животных; стабильность хромосомного состава при длительном культивировании *in vitro*; условия культивирования (питающие клетки (фидер)), компоненты среды для культивирования). Успех эксперимента по получению трансгенной линии мышей во многом зависит именно от свойств модифицируемых ЭС клеток (Kraus et al., 2010). Помимо линий ЭС клеток для получения химерных мышей необходима технологическая база, обеспечивающая возможность проведения инъекции ЭС клеток в бластоцисты, пересадку бластоцист реципиентным самкам и другие вспомогательные операции. Процедура создания химерных мышей требует тщательного планирования и координации, так как инъекцию ЭС клеток в бластоцисту необходимо проводить на определенной стадии эмбрионального развития (Longenecker, Kulkarni, 2009). В данной работе мы исследовали потенциал новой линии ЭС клеток, DGES1, полученной в лаборатории генетики развития ИЦиГ СО РАН. Использование линии DGES1 в сочетании со специфическими генотипами реципиентных бластоцист и приемных матерей привело к эффективному получению фертильных химерных мышей.

Материалы и методы

Культивирование клеток. Линию ЭС клеток DGES1 получили из 3.5D бластоцисты мышей *Mus musculus* линии 129S2/SvPasCrl (ЦКП «SPF-виварий» ИЦиГ СО РАН) по следующему протоколу (Bryja et al., 2006). Число хромосом линии DGES1 составляло $2n = 40$, XY. Дальнейшее культивирование проводили в среде для ЭС клеток на инактивированных митомицином С эмбриональных фибробластах мыши (фидере) на пластике, покрытом 0.1 % желатином. Для пересева ЭС клеток использовали

0.05 % трипсин-ЭДТА. Среда для ЭС клеток включала DMEM с 4.6 г/мл глюкозы (Thermo Fisher Scientific, США), 7.5 % FBS для ЭС клеток (Thermo Fisher Scientific, США), 7.5 % KSR (Thermo Fisher Scientific, США), 1X GlutaMAX (Thermo Fisher Scientific, США), 1X NEAA (Thermo Fisher Scientific, США), 1X пенициллин-стрептомицин (Thermo Fisher Scientific, США) и 1000 ед./мл LIF (PolyGene, США). Для инъекций в бластоцисту клетки культивировали без фидера на желатине, при этом в среду дополнительно добавляли 2i ингибиторы: 1 мкМ PD0325901 (Sigma-Aldrich, США) и 3 мкМ CHIR99021 (Sigma-Aldrich, США) (Gertsenstein et al., 2010). Культуры клеток содержались при 37 °C в 5 % CO₂ инкубаторе. Клетки пассировали каждые два дня. В данном исследовании для микроинъекции в бластоцисты использовали клетки DGES1 на 34-м пассаже.

Животные и условия содержания. Исследование выполнено на базе Центра генетических ресурсов лабораторных животных ИЦиГ СО РАН (RFMEFI61914X0005 и RFMEFI62114X0010). Использовали самцов инбредной линии мышей C57BL/6J, самок гибридов B6D2F1, полученных при скрещивании самок линии C57BL/6J и самцов DBA/2JRccHsd, а также самок и самцов аутбредной линии CD-1. Животных содержали в индивидуально вентилируемых клетках (Optimice, США), самцов по одному, а самок по пять в клетке при температуре 24 °C, фотопериоде 14C:10T, влажности 40–50 % и свободном доступе к воде и корму. Для кормления использовали брикетированные корма SNIFF (Германия), деионизированную воду обогащали солями в соответствии с требованиями к воде, используемой для поения SPF-мышей. Подстилочным материалом служили обеспыленная березовая крошка (Альбион, Новосибирск). Все подопытные животные были свободны от видоспецифических патогенов (specific pathogen free – SPF).

Получение вазэктомированных самцов. Процедуру вазэктомии проводили не менее чем за две недели до начала эксперимента. Самцов наркотизировали внутрибрюшинным последовательным введением препаратов Домитор (15 мкг/100 г массы тела) и Золетил (3 мг/100 г массы тела). После наркотизации самцов помещали на подогреваемый столик, обрабатывали операционное поле 70 % этиловым спиртом. Семьявыносящие протоки отделяли от сопряженных тканей и пережигали пинцетом, раскаленным на спиртовой горелке, и закрывали надрезы наложением двух-трех швов.

Получение бластоцист. В качестве доноров бластоцист использовали самок мышей B6D2F1 в возрасте 8–12 нед. Для синхронизации эстрального цикла и получения большого числа бластоцист индуцировали суперовуляцией. За два часа до отключения света в комнате содержания животных самкам внутрибрюшинно вводили 7.5 у.е. гонадотропина из сыворотки жеребых кобыл (Фоллигон, MSD Animal Health, Нидерланды). Через 46–48 ч внутрибрюшинно вводили 7.5 у.е. хорионического гонадотропина человека (Хоруллон, MSD Animal Health, Нидерланды). Сразу после инъекции хорионического гонадотропина человека самок подсаживали по одной к фертильным самцам линии C57BL/6J. На следующее утро после подсадки к самцу у самок проверяли наличие

вагинальных пробок и при обнаружении таковых отсаживали мышей в отдельные клетки. День обнаружения вагинальной пробки считали 0.5 днем беременности. Блостоцисты вымывали из матки самок-доноров через три дня после обнаружения вагинальной пробки (срок беременности – 3.5 дня). У самок, подвергнутых кранио-цервикальной дислокации, извлекали матку и освобождали от жировой ткани. Каждый рог матки через апикальный надрез промывали подогретой до 37 °C эмбриональной средой M2 (Sigma-Aldrich, США). Собранные бластоцисты переносили в 50 мкл среды KSOM (CosmoBio, Япония), покрывали минеральным маслом (Sigma-Aldrich, США) и помещали в CO₂-инкубатор.

Получение химер. Микроинъекции проводили под инвертированным микроскопом Axio Observer (Carl Zeiss, Германия), оснащенным системой микроманипуляторов и микроинъекторов (Narishige, Япония) и охлаждающим столиком (+6 °C), соединенным с водяной помпой. Суспензию ЭС клеток (50–100 мкл) и 15–20 бластоцист вносили в инъекционную камеру, содержащую среду, используемую для культивирования ЭС клеток DGE51. ЭС клетки собирали в инъекционный капилляр (TransferTip, Eppendorf, Германия) и, удерживая бластоцисту холдинговой пипеткой, вводили в бластоцель 12–15 клеток (рис. 1).

Бластоцисты с инъекрованными ЭС клетками и не имеющие признаков лизиса клеток переносили в чашку со средой KSOM и инкубировали 1–2 ч при 37 °C в атмосфере 5 % CO₂. Псевдобеременным самкам (2.5-й день беременности) CD-1, покрытым вазэктомированными самцами линии CD-1, хирургическим путем трансплантировали 12–13 здоровых бластоцист в левый рог матки. Операции проводили под газовым наркозом (Aerrane, Baxter Healthcare Corp., США).

ПЦР-анализ микросателлитов. ДНК из хвостов мышей и из культур ЭС клеток выделяли лизированием в буфере (100 mM NaCl; 10 mM Tris-HCl, pH 8.0; 10 mM EDTA; 0.5 % SDS) с добавлением протеиназы К (конечная концентрация – 1 мкг/мл). Пробирки инкубировали при темпе-



Fig. 1. Microinjection of DGE51 ES cells into a mouse blastocyst.

ратуре 56 °C в течение 12–16 ч с последующей инактивацией протеиназы К (95 °C, 10 мин). Затем проводили очистку ДНК фенол-хлороформом. Реакционная смесь для ПЦР объемом 25 мкл содержала 1× AS буфер (67 mM Tris-HCl (pH 8.8 при 25 °C); 16.6 mM (NH₄)₂SO₄; 1.5 mM MgCl₂; 0.01 % Tween 20), 0.2 mM каждого дезоксинуклеотида (дАТФ, дЦТФ, дГТФ и дТТФ), 0.4 мкМ каждого из праймеров, 0.5 ед. Taq полимеразы, 100 нг геномной ДНК. ПЦР проводили на автоматическом амплификаторе T100 Thermal Cycler (Bio-Rad, США). Для ПЦР использовали следующие условия амплификации: 95 °C в течение 3 мин, затем 35 циклов амплификации (95 °C – 30 с, 58 °C – 30 с, 72 °C – 20 с), заключительная элонгация 72 °C – 3 мин. Продукты амплификации визуализировали с помощью электрофореза в 3 % агарозном геле. Использовали праймеры для амплификации микросателлита D6Mit102: 5'-CCATGTGGATATCTTCCCTTG-3' (F) и 5'-GTATACCCAGTTGTAAATCTTGTGTG-3' (R) (Lamacchia et al., 2007).

Результаты

Получение химер и их потомков. В ходе эксперимента в 136 бластоцист микроинъектировали ЭС клетки DGE51 (таблица). Бластоцисты, не имеющие признаков повреждений после микроинъекций, были подсажены 10 реципиентным самкам, все они успешно выносили беременность и родили мышат. Количество детенышей в пометах варьировало от двух до девяти. Всего родилось 66 мышат, 15 из них оказались химерными (рис. 2). Оценка химеризма по окраске шерсти выявила долю агути около 80 % у четырех мышей, от 30 до 80 % у восьми и менее 30 % у трех. Для проверки фертильности и способности химер передавать потомству геном ЭС клеток химер, достигших возраста половой зрелости, ссаживали с самками C57BL/6J. Из 15 химерных самцов 10 были фертильны. От каждого из них получили не менее четырех пометов. Всего было получено 270 потомков, среди которых 20.7 % имели окраску агути (вклад генома ЭС клеток DGE51). Мышата с окраской агути родились в пометах от четырех из десяти химерных самцов. В пометах, полученных от самок, покрытых двумя самцами с выраженностью химеризма ~90 %, встречаемость агути среди потомков составила 100 и 54.7 % соответственно. У потомков двух самцов с выраженностью химеризма 50 и 30 % доля носителей агути составила 8 и 2.6 % соответственно.

ПЦР-анализ вклада ЭС клеток DGE51 в зародышевый путь. Для того чтобы подтвердить вклад ЭС клеток в половые клетки химерных животных, мы использовали полиморфный микросателлит D6Mit102 (Lamacchia et al., 2007), позволяющий различать родительское происхождение хромосомы 6 в линиях мышей C57BL/6J и 129S2/SvPasCrl. Анализ показал, что потомки от скрещивания химерных самцов, имеющих вклад из ЭС клеток DGE51 (генотип 129S2/SvPasCrl), и самок мышей линии C57BL/6J несут два аллеля,

Efficiency of chimera production from DGE1 ES cells injected into blastocysts

Number of successfully injected blastocysts	Number of recipient females	Total progeny size	Total number of chimeras	Number of male chimeras	Number of fertile male chimeras
124 of 136 (91.2 %)	10	66 (53.3 %)	15 (22.7 %)	15 (100 %)	10 (66.7 %)



Fig. 2. Male chimeras born after DGE1 ES cell microinjection into a mouse blastocyst.
Agouti (brown) patches of coat color indicate the contribution of DGE1 cells with the dominant agouti allele to animal tissues.

характерных для линий C57BL/6J и 129S2/SvPasCrl, так же, как и их химерные родители (рис. 3). В некоторых пометах от скрещивания химер на C57BL/6J рождались животные с черным цветом шерсти. Это объясняется тем, что у химерных животных сперматозоиды могут образовываться как из потомков клеток линии DGE1, так и из клеток blastocисты-реципиента. Мы проводили ПЦР-анализ микросателлитов только для потомков с цветом

шерсти агути. Результаты ПЦР-анализа демонстрируют, что химерные самцы, родившиеся в результате инъекции клеток линии DGE1, действительно передают свой генетический материал потомкам (germline transmission). Таким образом, ЭС клетки линии DGE1 подходят для получения трансгенных линий мышей.

Обсуждение

Получение трансгенных мышей с помощью инъекции ЭС клеток в эмбрионы мыши крайне востребовано, когда необходимо внести сложные модификации генома. Была создана новая линия ЭС клеток мыши, DGE1, пригодная для эффективного получения трансгенных животных. Кроме того, мы показали, что линия ЭС клеток DGE1 сохраняет плюрипотентность и способность формировать ткани и органы химерных мышей, в том числе гонады, на 34-м пассаже культивирования. Это означает, что накопление пассажей при проведении генетических модификаций клеток этой линии не будет ограничивать ее потенциал. Таким образом, новые ЭС клетки линии DGE1 можно эффективно использовать для генетической модификации и получения трансгенных мышей.

Не менее важную роль для рождения большого числа фертильных химерных животных играет сочетание генотипов ЭС клеток, blastocист и приемных матерей, поэтому постоянно ведутся поиски оптимальных комбинаций этих факторов (Seong et al., 2004; Fielder et al., 2012; Alcantar et al., 2016).

Для повышения эффективности при создании химер мы использовали 2i ингибиторы (Gertsenstein et al., 2010) и сочетание генотипов: ЭС клетки 129S2/SvPasCrl, blastocисты B6D2F1 и реципиенты CD-1. Сравнение пока-

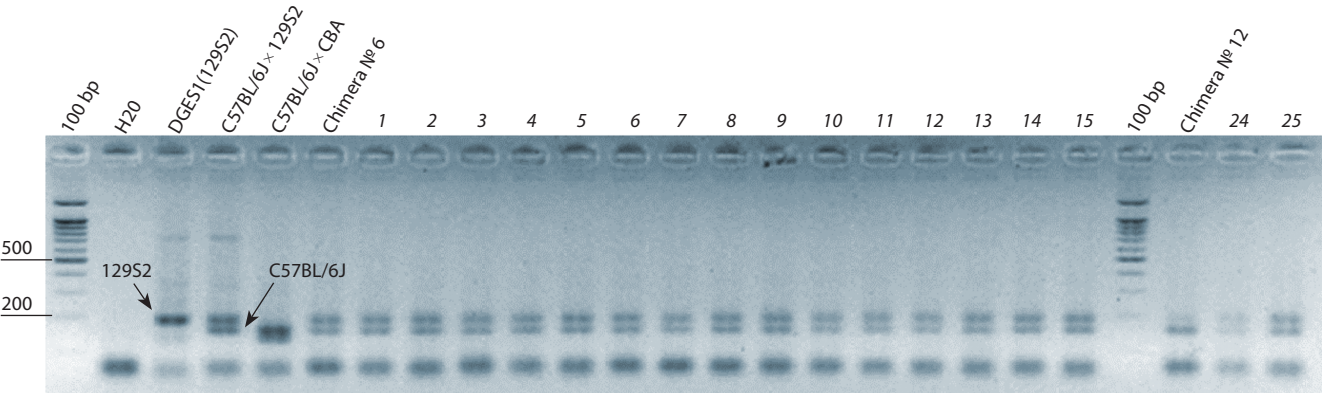


Fig. 3. Microsatellite PCR analysis of the DGE1 ES cells and the offspring from DGE1 chimeras. Only agouti-colored pups were selected for microsatellite PCR.
Lanes: 100 bp, 100-bp DNA ladder; H2O, negative control; DGE1 (129S2), DNA from the DGE1 cell line (129S2/SvPasCrl genotype); C57BL/6J × 129S2 and C57BL/6J × CBA, DNA from hybrid strains containing the C57BL/6J allele; chimeras №6 and 12, male DGE1 chimeras; 1–15, the offspring of chimera No.6 born after crossing to C57BL/6J females; 24–25, the corresponding offspring of chimera No.12.

зателей эффективности получения химер после инъекций ЭС клеток в бластоцисты, наблюдаемых в нашей работе, с результатами, представленными М. Ну с коллегами (2013), показало, что используемая нами комбинация H-2 гаплотипов ЭС клеток, бластоцист и приемных матерей повышает выход химер. В нашей работе доля химер, полученных из общего числа инъецированных бластоцист, составила 12.1 %, из них химерных самцов было 100 %, эти показатели превышают результативность получения химер 7 и 66.7 % соответственно, достигнутую в одной из последних работ (Ну et al., 2013). Наблюдаемое увеличение выхода химер может быть обусловлено иммуногенетическими различиями по главному локусу гистосовместимости (H-2) ЭС клеток, бластоцист и приемной матери. Показано, что особенности иммуногенетических взаимоотношений мать–плод, складывающихся на стадии формирования бластоцисты, оказывают существенное влияние на эндокринное обеспечение беременности и, как следствие, на жизнеспособность эмбрионов в периоды доимплантационного и постимплантационного развития (Gerlinskaya, Evsikov, 2001). Это обстоятельство может объяснять меньший выход химер (Ну et al., 2013) тем, что в данной работе для микроинъекций ЭС клеток использовали бластоцисты аутбредных самок линии CD-1 и в качестве реципиентов – самок этой же CD-1 линии. Наш подход отличался тем, что для микроинъекций ЭС клеток применялись гибридные бластоцисты B6D2F1, несущие H-2^d и H-2^b локусы, которые подсаживали аутбредным самкам CD-1. Проблеме влияния комбинации H-2 гаплотипов химерных бластоцист и организма приемной матери на выживаемость химер в организме приемной матери и жизнеспособность не уделяется должного внимания, хотя Р. Dvorak с коллегами (1995) показали, что определенные комбинации H-2 химерных бластоцист и приемных матерей оказывают существенное влияние на успешность имплантации. Наши результаты, в свою очередь, указывают на влияние комбинации H-2 гаплотипа химерных бластоцист и приемной матери на выживаемость в период беременности и фертильную способность взрослых химерных самцов. Таким образом, сочетание новой линии ЭС клеток DGES1 и комбинации H-2 гаплотипов бластоцист, используемых для инъекций ЕС клеток и приемных матерей, позволяет создавать новые трансгенные линии мышей с высокой эффективностью.

Acknowledgments

The experiments and result processing were supported by the Russian Science Foundation, project 16-14-00095. The

purchase of laboratory animals and equipment rental were supported by State Budgeted Project 0324-2015-0004.

Conflict of interest

The authors declare no conflict of interest.

References

- Alcantar T.M., Wiler R., Rairdan X.Y. Comparison of BALB/c and B6-albino mouse strain blastocysts as hosts for the injection of C57BL/6N-derived C2 embryonic stem cells. *Transgenic Res.* 2016;25(4):527-531. DOI 10.1007/s11248-016-9937-5.
- Bryja V., Bonilla S., Arenas E. Derivation of mouse embryonic stem cells. *Nat. Protoc.* 2006;1(4):2082-2087. DOI 10.1038/nprot.2006.355.
- Dvorak P., Yoshiki A., Dvorakova D., Flechon J.E., Kusakabe M. Cell mixing during the early development of mouse aggregation chimera. *Int. J. Dev. Biol.* 1995;39(4):645-652.
- Evans M.J., Kaufman M.H. Establishment in culture of pluripotent cells from mouse embryos. *Nature.* 1981;292(5819):154-156.
- Fielder T.J., Yi C.S., Masumi J., Waymire K.G., Chen H.W., Wang S., Shi K.X., Wallace D.C., MacGregor G.R. Comparison of male chimeric mice generated from microinjection of JM8.N4 embryonic stem cells into C57BL/6J and C57BL/6NTac blastocysts. *Transgenic Res.* 2012;21(6):1149-1158. DOI 10.1007/s11248-012-9605-3.
- Gerlinskaya L.A., Evsikov V.I. Influence of genetic dissimilarity of mother and fetus on progesterone concentrations in pregnant mice and adaptive features of offspring. *Reproduction.* 2001;121(3):409-417.
- Gertsenstein M., Nutter L.M., Reid T., Pereira M., Stanford W.L., Rosant J., Nagy A. Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS ONE.* 2010;5(6):e11260. DOI 10.1371/journal.pone.0011260.
- Hu M., Wei H., Zhang J., Bai Y., Gao F., Li L., Zhang S. Efficient production of chimeric mice from embryonic stem cells injected into 4- to 8-cell and blastocyst embryos. *J. Anim. Sci. Biotechnol.* 2013;4(1):12. DOI 10.1186/2049-1891-4-12.
- Kraus P., Leong G., Tan V., Xing X., Goh J.W., Yap S.P., Lufkin T. A more cost effective and rapid high percentage germ-line transmitting chimeric mouse generation procedure via microinjection of 2-cell, 4-cell, and 8-cell embryos with ES and iPS cells. *Genesis.* 2010;48(6):394-399. DOI 10.1002/dvg.20627.
- Lamacchia C., Palmer G., Gabay C. Discrimination of C57BL/6J Rj and 129S2/SvPasCrl inbred mouse strains by use of simple sequence length polymorphisms. *J. Am. Assoc. Lab. Anim. Sci.* 2007;46(2):21-24.
- Longenecker G., Kulkarni A.B. Generation of gene knockout mice by ES cell microinjection. *Curr. Protoc. Cell Biol.* 2009;19(14):1-36. DOI 10.1002/0471143030.cb1914s44.
- Seong E., Saunders T.L., Stewart C.L., Burmeister M. To knockout in 129 or in C57BL/6: that is the question. *Trends Genet.* 2004;20(2):59-62. DOI 10.1016/j.tig.2003.12.006.
- Wijshake T., Baker D.J., van de Sluis B. Endonucleases: new tools to edit the mouse genome. *Biochim. Biophys. Acta.* 2014;1842(10):1942-1950. DOI 10.1016/j.bbdis.2014.04.020.

Практическое руководство по редактированию геномов системой CRISPR/Cas9

А.Г. Мензоров^{1,2}, В.А. Лукьянчикова¹, А.Н. Кораблев¹, И.А. Серова¹, В.С. Фишман^{1,2} 

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Технология CRISPR/Cas за последние несколько лет совершила прорыв в области редактирования геномов. В силу высокой эффективности и простоты сборки отдельных компонент в условиях современной лаборатории, система CRISPR/Cas применяется огромным количеством исследователей в самых разных областях биологии. После появления первых сведений о редактировании генома млекопитающих системой CRISPR/Cas9 было разработано множество методов, предлагающих те или иные модификации белков семейства Cas или направляющих РНК. Целый ряд работ посвящен использованию технологий, основанных на системе CRISPR/Cas, для самых неожиданных целей – не только для редактирования геномов, но и для контроля экспрессии определенных генов, локализации и визуализации отдельных локусов ДНК в пространстве ядра, изменения статуса метилирования заданных сайтов в геноме млекопитающих и многое другое. Подробности работы системы CRISPR/Cas и способы ее применения детально рассмотрены нами ранее. Наиболее часто система CRISPR/Cas используется именно для редактирования геномов, однако, несмотря на кажущуюся простоту, существует ряд технических сложностей, с которыми можно столкнуться, применяя эту технологию впервые. В настоящей статье подробно описаны протоколы, связанные со сборкой, тестированием и использованием системы CRISPR/Cas9 для внесения мутаций в геном млекопитающих. Дано краткое пояснение теоретических аспектов процессов, связанных с внесением направленных модификаций в геном млекопитающих. Подробно описана методика клонирования и тестирования векторов для доставки компонент системы CRISPR/Cas9 в клетки. Приведены протоколы модификации генома клеток в культуре и создания трансгенных животных. В каждом из разделов упомянуты потенциальные сложности, связанные с использованием системы CRISPR/Cas9, и предложены способы их преодоления.

Ключевые слова: редактирование геномов; CRISPR/Cas9; эмбриональные стволовые клетки.

Genome editing using CRISPR/Cas9 system: a practical guide

A.G. Menzorov^{1,2}, V.A. Lukyanchikova¹,
A.N. Korablev¹, I.A. Serova¹, V.S. Fishman^{1,2} 

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

Over the past few years, the CRISPR/Cas techniques have become a revolution in genome editing. Since the original paper on CRISPR/Cas9 genome editing, researches have proposed numerous modifications of the key components of the CRISPR/Cas9 system to make it extremely efficient. Nowadays, CRISPR/Cas systems can be used not only to modify genomes, but also to control expression levels of defined genes, visualize loci of interest in the space of living cell nuclei, change methylation status of mammalian CpG sites, and to serve many other purposes. Due to an extremely high efficacy and ease of usage, the CRISPR/Cas system has been employed in a large number of studies in various areas of biology and biotechnology. We have recently published a review describing various CRISPR/Cas systems, mechanisms of their functioning, and applications of the techniques in details. Despite the broad range of potential applications of CRISPR/Cas systems, they are mostly used for genome editing. And, however simple the system may be, there is a number of potential pitfalls on the way towards its use in CRISPR/Cas-naïve laboratory settings. In this article, we describe protocols of CRISPR/Cas9 system generation. We start with a short description of theoretical aspects underlying Cas9-mediated genome editing. Next, we describe a step-by-step protocol of guide RNA vector design and assembly, and several ways of qualitative and quantitative evaluations of the system. Finally, we report protocols of genome editing for modification of embryonic stem cells and zygotes.

Key words: genome editing; CRISPR/Cas9; embryonic stem cells.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Мензоров А.Г., Лукьянчикова В.А., Кораблев А.Н., Серова И.А., Фишман В.С. Практическое руководство по редактированию геномов системой CRISPR/Cas9. Вавиловский журнал генетики и селекции. 2016;20(6):930-944. DOI 10.18699/VJ16.214

HOW TO CITE THIS ARTICLE:

Menzorov A.G., Lukyanchikova V.A., Korablev A.N., Serova I.A., Fishman V.S. Genome editing using CRISPR/Cas9 system: a practical guide. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):930-944. DOI 10.18699/VJ16.214



Модификация генома млекопитающих системой CRISPR/Cas9. Система CRISPR/Cas9 состоит из двух основных частей. Первой является белок, нуклеаза Cas9, которая способна вносить двуцепочечный разрыв в молекулу ДНК. Вторая представляет собой небольшую молекулу РНК размером около 120 нуклеотидов – химерную направляющую РНК (хнРНК). Далее будем использовать сокращение хнРНК; в английском языке распространена аббревиатура gRNA (от англ. guide RNA – направляющая РНК). Последовательность направляющей РНК можно условно разделить на две части. Около 100 нуклеотидов, расположенных на 3'-конце, являются одинаковыми для всех направляющих РНК и обеспечивают формирование специфической пространственной структуры, которая узнается белком Cas9. Около 20 нуклеотидов, расположенных на 5'-конце направляющей РНК, определяют последовательность ДНК, с которой свяжется белок Cas9. Белок Cas9 вносит двуцепочечный разрыв в молекулу ДНК в месте, которое комплементарно 5'-последовательности хнРНК при условии, что непосредственно за комплементарным хнРНК участком находится тринуклеотид NGG (здесь и далее N – любой из четырех нуклеотидов ДНК, А, Т, G или С), который называется РАМ (protospacer adjacent motif – мотив, прилегающий к протоспейсеру) (рис. 1, а). Разрыв в ДНК происходит на расстоянии 3–4 нуклеотидов от РАМ. На этом работа системы CRISPR/Cas9 заканчивается, и дальнейшая модификация генома происходит при участии системы репарации ДНК клетки.

Двуцепочечные разрывы в клетках млекопитающих могут репарироваться разными путями. В случае репарации по механизму нехомологичного соединения концов (non-homologous end joining – NHEJ) в месте разрыва могут возникать небольшие inserции или делеции. В случае репарации по механизму гомологичной рекомбинации для восстановления информации о последовательности ДНК используются сестринские хроматиды или искусственно введенная экспериментатором в клетку генетическая конструкция, имитирующая сестринскую хроматиду. Более подробно этот вопрос рассмотрен в обзоре (Смирнов и др., 2016).

Дизайн эксперимента, предполагающего использование системы CRISPR/Cas9, зависит от того, какую именно модификацию генома исследователь планирует получить. Можно выделить три типа таких модификаций.

Первый тип модификации – внесение небольших inserций и делеций (INDEL – INsertion or DEletion). Размер таких модификаций может варьировать от одной до нескольких десятков пар оснований, но, как правило, затрагивает от 1 до 10 нуклеотидов. Чаще всего внесение INDEL применяется для нокаутирования какого-либо гена, при этом INDEL может приводить к сдвигу рамки считывания или разрушению сайта сплайсинга. В первом случае (сдвиг рамки считывания) экспрессия гена и транскрибирующаяся мРНК меняются лишь незначительно, однако при трансляции образуется нефункциональный белок. Во втором случае (разрушение сайта сплайсинга) последовательность мРНК и уровень ее экспрессии могут существенно изменяться, и, кроме того, при трансляции тоже образуется нефункциональный белок. Для реализа-

ции такого эксперимента исследователю достаточно обеспечить экспрессию в клетке одной хнРНК (специфичной модифицируемому участку) и белка Cas9.

Второй тип модификации – введение целевой последовательности ДНК в определенный участок генома. Для этого системой CRISPR/Cas9 вносится разрыв в участок генома, который необходимо модифицировать. Одновременно в клетку доставляют фрагмент ДНК, содержащий встраиваемый участок, фланкированный гомологичными месту разрыва последовательностями (конструкцию для гомологичной рекомбинации). Конструкция для гомологичной рекомбинации имитирует сестринскую хроматиду, и информация из нее «копируется» в место разрыва. Таким образом, для реализации эксперимента исследователю требуется экспрессия в клетке хнРНК, Cas9 и наличие конструкции для гомологичной рекомбинации.

Третий тип модификации – создание крупных (от сотен до сотен тысяч пар оснований) делеций. Наиболее простой способ такой модификации – экспрессия в клетке белка Cas9 и двух хнРНК, каждая из которых будет нести на 5'-конце последовательность нуклеотидов, комплементарную одной из границ предполагаемой делеции. В результате в хромосоме возникнут два разрыва, концы которых могут быть соединены по механизму NHEJ, что и приведет к возникновению необходимой делеции.

Протокол внесения модификаций в геном млекопитающих при помощи системы CRISPR/Cas9

Этап 1. Выбор последовательностей хнРНК и дизайн эксперимента

Теоретически любая последовательность геномной ДНК, содержащая два подряд идущих гуанина, может быть использована как сайт для внесения разрывов системой CRISPR/Cas9. Например, в случае последовательности ДНК 5'-GTAGGCGACGAATTGACACATGG-3' (здесь и далее подчеркнут и выделен жирным шрифтом РАМ) хнРНК должна на 5'-конце иметь 20 нуклеотидов, прилегающих к РАМ-последовательности: GUAGGCGACGAAUUGACACA.

Важно! Обратите внимание, что сама последовательность РАМ не входит в состав хнРНК, но должна обязательно присутствовать в ДНК.

Важным аспектом выбора хнРНК является анализ ее специфичности. Активация системы CRISPR/Cas9 и внесение разрывов могут происходить даже в случае неполной комплементарности хнРНК и ДНК. Наибольшую роль в специфическом связывании комплекса хнРНК-Cas9 с ДНК играют нуклеотиды хнРНК, располагающиеся ближе к 3'-концу, наименьшую – располагающиеся ближе к 5'-концу. Таким образом, внесение разрывов возможно даже при совпадении 12–16 из 20 нуклеотидов хнРНК, задающих специфичность, с последовательностью геномной ДНК (Jinek et al., 2012). Неспецифические сайты узнавания называют также off-target сайтами, и их наличие можно легко определить в геномах с известной нуклеотидной последовательностью для каждой конкретной хнРНК.

В ряде работ было показано, что нуклеотидный состав хнРНК может влиять на эффективность работы систе-

мы CRISPR/Cas9 (Chari et al., 2015). Поэтому следует оценивать нуклеотидный состав выбранной для таргетирования последовательности при помощи специального программного обеспечения.

Наконец, для эффективной транскрипции выбранной хнРНК *in vitro* или *in vivo* желательно, чтобы первым нуклеотидом, расположенным на 5'-конце хнРНК (с которого будет начинаться синтез молекулы РНК), был гуанин. (Это связано с использованием U6 промотора для транскрипции направляющей РНК. При использовании других промоторов такой необходимости нет.) Поэтому предпочтительно выбирать для таргетирования последовательности ДНК вида G(N)₁₉**NGG**. В случае, если это невозможно, первая буква выбранной последовательности должна быть заменена на гуанин, что в большинстве случаев не будет фатальным для узнавания таргетируемой последовательности, поскольку специфичность системы CRISPR/Cas9 в большей степени задается 12–16 буквами, расположенными ближе к 3'-концу специфической последовательности хнРНК.

Таким образом, для выбора последовательности хнРНК в определенном участке генома исследователю необходимо:

1. Найти все тринуклеотиды NGG (PAM-последовательности) в этом участке*
2. Для каждого тринуклеотида проанализировать нуклеотидную композицию двадцати нуклеотидов, расположенных на 5'-конце PAM
3. Выбрать последовательности с наименьшим числом неспецифических мест связывания в геноме (желательно вообще не имеющие таких мест связывания)
4. При необходимости заменить первый нуклеотид в выбранной последовательности на гуанин

* Кроме последовательностей вида «NGG» подходят и последовательности «CCN», в этом случае необходимо выполнять пункты 1–4 с комплементарной цепью ДНК.

** Данные о неспецифических сайтах и эффективности получены при помощи программы https://crispr.med.harvard.edu/sgRNA_Scorer для генома человека. Функционал программы обсуждается ниже.

Выполнение пунктов 1–3 достаточно трудоемкий процесс, для оптимизации которого были созданы специальные онлайн-сервисы. Мы рекомендуем использовать такие сервисы, а не проводить «ручной» подбор хнРНК.

Достаточно простой по функциональности сервис расположен на

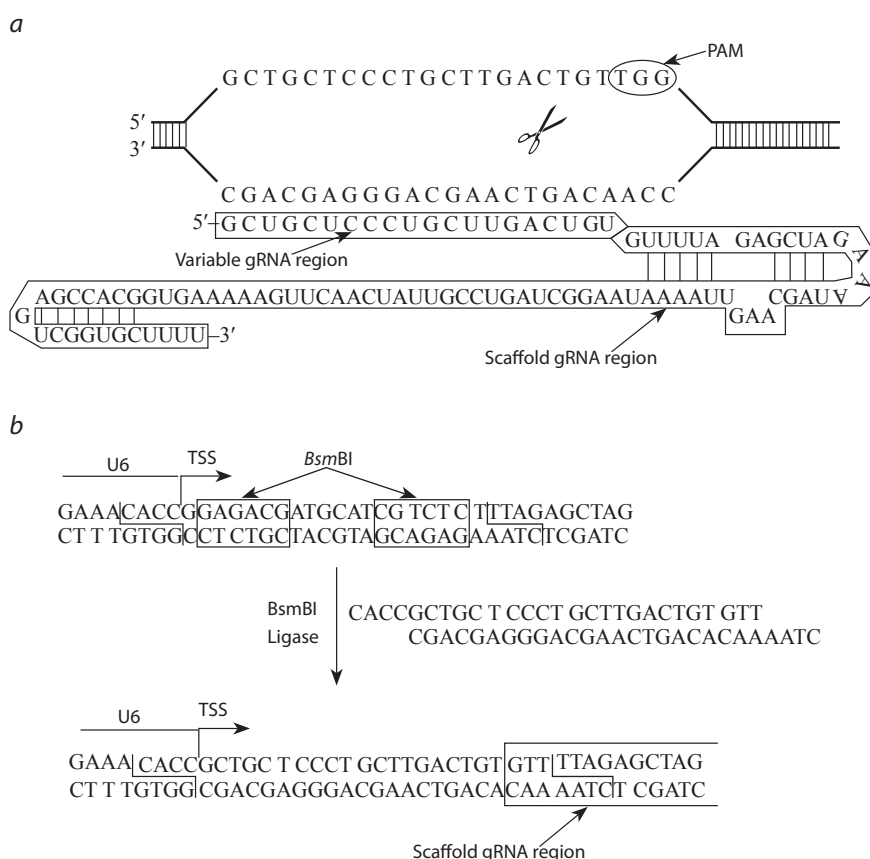


Fig. 1. Generation of a gRNA-carrying vector. (a) Schematic representation of gRNA and a targeting locus. (b) A strategy of gRNA vector cloning.

Ligase, T4 DNA ligase; U6, promoter of the U6 gene; TSS, position of the transcription start site.

AATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGG
ATCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTATAAAAAAAAGGCTTGACCGA
CAATTGCATGAAGAATCTGCTTAGTTACGTTTTGCGCT
TCTGCTCCCTGCTTGTGTGT – низкая эффективность**
AAATTTAAATTATAAAAAAA – высокая эффективность

TCTGCTCCCTGCTTGTGTGT – 7 неспецифических сайтов
AAATTTAAGCTATAAAAAAA – более 40 неспецифических сайтов

GCTGCTCCCTGCTTGACTGT
GAATTTAAGCTATAAAAAAA

сайте <http://crispr.mit.edu/>. Он позволяет определять в последовательности ДНК все потенциальные сайты модификации системой CRISPR/Cas9 и находить для каждого из них неспецифические сайты связывания в выбранном геноме (off-target сайты). Для каждой хнРНК программа вычисляет рейтинг (score), показывающий количество потенциальных неспецифических сайтов. Чем ближе рейтинг к 100, тем меньше неспецифических сайтов найдено.

Другая, более функциональная программа sgRNA Scorer 1.0 доступна по адресу <https://crispr.med.harvard.edu/>. В случае применения этой программы пользователю высылается на адрес электронной почты отчет, основной

файл которого (под названием CasFinderOutput.input.FinalOutput) содержит все потенциальные сайты модификации системой CRISPR/Cas9 в целевой последовательности ДНК, предполагаемую эффективность модификации в зависимости от нуклеотидного состава хнРНК, количество и позиции off-target сайтов в выбранном геноме. Предсказанная эффективность модификации (столбец score) варьирует от 0 до 100, большее значение score соответствует большей эффективности. Мы рекомендуем использовать хнРНК, имеющие score более 90.

Внимание! Программа *sgRNA Scorer 1.0* требует ввода данных в формате FASTA. Это означает, что перед каждой из анализируемых последовательностей ДНК, вводимой в окно «Paste Sequence», должна располагаться строка с названием этой последовательности, начинающаяся с символа «>». Пример:

```
>Test_DNA_1
AATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCT
CCCTGCTTGTGTGTTGGATCGCTGAGTAGTGCGCGAGC
AAAAATTAAGCTATAAAAAAAGGCTTGACCGACAATTGC
ATGAAGAATCTGCTTAGTTACGTTTTGCGCT
```

Наконец, следует обратить внимание на то, что выбранная последовательность хнРНК не должна содержать сайтов узнавания для фермента *BsmBI* (см. этап 2 «Создание вектора для экспрессии хнРНК»).

Особенности выбора последовательностей хнРНК и дизайна эксперимента для определенных типов модификаций. При выборе последовательности хнРНК важно учесть некоторые особенности, связанные с дизайном планируемого эксперимента. Например, если предполагается тестирование ДНК при помощи рестрикционного анализа (см. этап 3), необходимо при выборе последовательности убедиться, что на расстоянии 3–4 нуклеотидов от PAM (в месте внесения двуцепочечного разрыва системой CRISPR/Cas9) содержится сайт узнавания какой-либо эндонуклеазы рестрикции.

В случае внесения INDEL и их последующей идентификации при помощи ПЦР (см. этап 4) необходимо одновременно с выбором хнРНК подобрать соответствующие праймеры для ПЦР-анализа. После того как праймеры, фланкирующие место посадки комплекса хнРНК-Cas9, подобраны, мы рекомендуем провести подбор условий ПЦР и только после этого начинать эксперимент по модификации генома. Чаще всего при выборе последовательности хнРНК используются данные из официальных версий сборки генома. Мы рекомендуем провести секвенирование полученных ПЦР-продуктов, чтобы удостовериться, что геном организмов (или клеток), которые предполагается использовать в эксперименте, совпадает с референсным геномом в таргетируемой области.

Если предполагается, что INDEL должен привести к нокауту какого-либо гена, необходимо тщательно проанализировать варианты альтернативного сплайсинга этого гена, наличие альтернативных промоторов и старто-кодонов, доменную организацию белка. В противном случае существует вероятность того, что внесенная мутация не приведет к полной утрате биологической функции.

Если планируется использование конструкций для гомологичной рекомбинации для внесения последовательностей ДНК в определенный участок генома, обратите

внимание на то, что сайт узнавания хнРНК должен отсутствовать в этой конструкции. В рамках данной статьи мы не останавливаемся подробно на протоколах создания такой конструкции, отсылая читателя к классическим руководствам по молекулярной биологии и генной инженерии (Щелкунов, 2004). Отметим только, что, исходя из нашего опыта, размер плечей гомологии должен быть не менее 500 нуклеотидов. Литературные данные свидетельствуют о том, что увеличение размера плечей гомологии до нескольких тысяч пар оснований приводит к увеличению эффективности модификации генома, в том числе и при использовании системы CRISPR/Cas9 (Li et al., 2014).

Наконец, эффективность разных хнРНК может существенно различаться. Мы рекомендуем выбирать и клонировать несколько разных хнРНК и определять эффективность для каждой из них в эксперименте (см. этап 3 «Тестирование хнРНК»).

Этап 2. Создание вектора для экспрессии хнРНК

Вектор для экспрессии хнРНК должен содержать следующие элементы: бактериальный ориджин репликации (для наработки вектора в бактериях); бактериальный селективный маркер (для селекции трансформированных бактерий, несущих вектор; например, ген устойчивости к канамицину); эукариотический промотор, обеспечивающий эффективную транскрипцию в клетках млекопитающих (традиционно используются промоторы, процессируемые РНК-полимеразой III, поскольку хнРНК относится к некодирующим РНК; например, промотор гена U6); последовательность хнРНК (начиная с 5'-гуанина); сигнал терминации транскрипции.

Как говорилось ранее, хнРНК состоит из специфической части, представленной 20 нуклеотидами на 5'-конце, и конститутивной части, одинаковой для всех экспериментов (около 100 нуклеотидов). Это означает, что векторы для экспрессии различных хнРНК будут различаться только 20 буквами на 5'-конце последовательности хнРНК. Поэтому были предложены эффективные системы клонирования, позволяющие заменять только небольшой фрагмент вектора на специфическую последовательность, содержащую 20 нуклеотидов необходимой хнРНК.

Мы предлагаем протокол клонирования хнРНК, основанный на использовании вектора *gRNA_Cloning_Vector_BsmBI*. Вектор *gRNA_Cloning_Vector_BsmBI* содержит бактериальный ориджин репликации, ген устойчивости к канамицину, промотор U6, два сайта рестрикции *BsmBI* и терминатор транскрипции. При обработке рестриктазой *BsmBI* из вектора удаляется небольшой фрагмент, расположенный между промотором U6 и конститутивной частью хнРНК (см. рис. 1, б). Поскольку фермент *BsmBI* относится к типу ферментов IIS, он узнает асимметричный сайт (CGTCTC) и вносит двуцепочечный разрыв на некотором расстоянии от него. В результате гидролиза вектора *gRNA_Cloning_Vector_BsmBI* из него удаляются сайты узнавания фермента *BsmBI* и образуются липкие концы для лигирования фрагмента ДНК, кодирующего специфическую последовательность хнРНК. Таким образом, обработку ферментом *BsmBI* и лигирование можно проводить в одной реакции.

Протокол клонирования вектора для экспрессии хнРНК

Материалы

- Вода. Здесь и далее в протоколах используется деионизованная вода (ddH_2O) с удельным сопротивлением более 18 МОм·см, например Sigma 7732-18-5.
- Плазмидный вектор gRNA_Cloning_Vector_BsmBI, концентрация 300 нг/мкл. Вектор был клонирован в лаборатории генетики развития на основе плазмиды Addgene #41824. Вектор и его полная нуклеотидная последовательность доступны по запросу в лаборатории генетики развития ИЦиГ СО РАН.
- Олигонуклеотиды. Для выбранной последовательности хнРНК вида $\text{G(N)}_{19}\text{NGG}$ и комплементарной ей последовательности $\text{CCX(X)}_{19}\text{C}$ (где азотистые основания X комплементарны соответствующим основаниям N) необходимо заказать химический синтез двух олигонуклеотидов: gRNA_F 5'-CACCG(N)₁₉GTT-3' и gRNA_R 5'-CTAAAC(X)₁₉C-3'.

Пример: для таргетируемой последовательности GCTGCTCCCTGCTTGACTGTGGG закажите олигонуклеотиды 5'-CACCGCTGCTCCCTGCTTGACTGTGTT-3' и 5'-CTAAACACAGTCAAGCAGGGAGCAGC-3'.

Для секвенирования полученных векторов необходимо заказать олигонуклеотид M13F: 5'-TGTAACACGACGGCCAGT-3'. Полученные олигонуклеотиды должны быть разведены до концентрации 100 мкМ в ddH_2O .

- Полинуклеотидкиназа фага T4 (T4 PNK, NEB M0201S) и соответствующий буфер (10X T4 PNK Buffer, поставляется вместе с ферментом).
- Лигаза фага T4 (T4 DNA ligase, NEB M0202S) и соответствующий буфер (T4 DNA Ligase Buffer, поставляется вместе с ферментом).
- Фермент BsmBI (NEB R0580S).
- Бычий сывороточный альбумин (BSA), концентрация 20 мг/мл, качество, пригодное для молекулярно-биологических работ (например, NEB B9000S).
- Компетентные клетки *Escherichia coli* (например, NEB C3019).
- Канамицин (например, Sigma 25389-94-0), концентрация стока 100 мг/мл.
- Жидкая среда LB (на 100 мл содержит: бакто-триптон 1 г; бактериально-дрожжевой экстракт 0.5 г; NaCl 0.5 г).
- Агаризованная среда LB (на 100 мл содержит: бакто-триптон 1 г; бактериально-дрожжевой экстракт 0.5 г; NaCl 0.5 г; агар 1.5 г).
- Наборы для выделения плазмидной ДНК (Biosilica PLD-100 или аналог).

Оборудование

- Программируемый термошейкер и/или амплификатор.
- Прибор для трансформации в случае использования электрокомпетентных клеток (BioRad Micropulser или аналог) и специализированные кюветы (Gene Pulser/Micropulser Cuvette, Bio-Rad).
- Пластиковые чашки Петри ($d = 10$ см).

Протокол

I. Отжиг и фосфорилирование нуклеотидов. Смешайте в пробирке в указанном порядке:

1. 6.5 мкл ddH_2O .

2. По 1 мкл каждого из двух олигонуклеотидов gRNA_F и gRNA_R.

3. 1 мкл T4 DNA Ligase Buffer*.

4. 0.5 мкл T4 PNK.

Объем реакции составляет 10 мкл.

Перемешайте содержимое пробирки пипетированием, затем инкубируйте в следующем температурном режиме:

1. 37 °C → 30 мин.

2. 95 °C → 5 мин.

3. Равномерное снижение температуры до 25 °C по 5 °C в минуту.

II. Реакция рестрикции–лигирования. Смешайте в пробирке в указанном порядке:

1. 6.5 мкл ddH_2O .

2. 1 мкл T4 DNA Ligase Buffer.

3. 0.1 мкл BSA.

4. 0.4 мкл реакционной смеси, полученной на шаге I (содержит фосфорилированные двуцепочечные фрагменты ДНК).

5. 1 мкл плазмиды gRNA_Cloning_Vector_BsmBI.

6. 0.5 мкл T4 DNA Ligase.

7. 0.5 мкл BsmBI.

Объем реакции составляет 10 мкл.

*Лигазный буфер (T4 DNA Ligase Buffer) содержит рибозадеинозинтрифосфорную кислоту (pATФ), которая разрушается при повторном размораживании и замораживании буфера. Поэтому не рекомендуется использовать размороженные аликвоты буфера повторно. Кроме того, в случае неэффективного лигирования можно дополнительно добавлять pATФ (NEB P0756) в реакцию.

Перемешайте содержимое пробирки пипетированием, затем инкубируйте в следующем температурном режиме:

1. 37 °C → 5 мин.

На этом шаге происходит гидролиз вектора ферментом BsmBI.

2. 20 °C → 5 мин.

На этом шаге происходит лигирование вставки и гидролизованного вектора.

3. Повторить шаги 1 и 2 15–35 раз.

На шаге 2 в гидролизированный вектор может быть как лигирован целевой фрагмент, так и обратно лигирован фрагмент, вырезанный из вектора на шаге 1. Однако в случае лигирования целевого фрагмента в образовавшемся векторе нет сайта узнавания фермента BsmBI, тогда как при обратном лигировании этот сайт восстанавливается. Поэтому увеличение количества циклов рестрикции–лигирования приводит к увеличению выхода целевого продукта.

На этом этапе протокол может быть остановлен, продукт реакции II может в течение нескольких часов храниться при температуре 10 °C или в течение нескольких дней при температуре –20 °C.

III. Трансформация бактериальных клеток. Используйте 0.5–1 мкл продукта реакции II для трансформации компетентных клеток *E. coli*. Необходимо высадить трансформированные бактериальные клетки на чашку Петри на твердую агаризованную среду LB (20–25 мл), содержащую селективный антибиотик канамицин с конечной концентрацией 50 мкг/мл. Убрать чашки с бактериями в термостат, инкубировать при 37 °C в течение 12–18 ч.

Существует несколько различных протоколов приготовления компетентных клеток и их трансформации. Мы не приводим их подробного описания, рекомендуя читателю ознакомиться с классическими молекулярно-биологическими методиками (Sambrook, Russell, 2001; Щелкунов, 2004). Для лабораторий, не использующих генно-инженерные методы рутинно, рекомендуем покупку компетентных клеток (например, NEB C2989K) и проведение трансформации в соответствии с протоколом производителя (например, <https://www.neb.com/protocols/1/01/01/electroporation-protocol-c2989>). Экспериментатору надо учитывать, что компетентные клетки НЕ ДОЛЖНЫ нести устойчивость к канамицину.

IV. Скрининг трансформантов. Выберете 2–4 клон из полученных трансформантов (в обычном эксперименте мы получаем десятки или сотни трансформантов). Нарастите соответствующие бактериальные культуры и выделите плазмидные ДНК, как описано в стандартных молекулярно-биологических протоколах (Sambrook, Russell, 2001; Щелкунов, 2004) или в руководствах к соответствующим наборам. Проведите секвенирование плазмидной ДНК с использованием праймера M13F, чтобы удостовериться в корректности последовательности хнРНК.

Этап 3. Тестирование хнРНК

Во многих публикациях показано, что эффективность работы CRISPR/Cas9 может сильно зависеть от выбора конкретной хнРНК. Поэтому перед началом эксперимента мы рекомендуем провести тестирование хнРНК. Для этого предложены два протокола тестирования, один из которых основан на системе pDR, другой – на методе разрушения сайта эндонуклеазы рестрикции.

Тестирование хнРНК с помощью системы pDR

Впервые система на основе pDR-плазмиды была создана и использована для оценки эффективности процесса гомологичной рекомбинации у млекопитающих в работе (Pierce et al., 1999). На клетках китайского хомячка, дефицитных по продукту гена *XRCC3*, предполагалось проверить, насколько критично отсутствие данного белка и будут ли в этом случае осуществляться процессы репарации по гомологичному механизму. В других работах с использованием pDR-вектора исследовались каталитические активности различных эндонуклеаз (Windbichler et al., 2007; Smith et al., 2009), механизмы репарационного процесса при некоторых вирусных заболеваниях (Kulkarni, Fortunato, 2011; Datta et al., 2014) и при патологических изменениях (Golding et al., 2004), а также создавались модельные системы для изучения процессов клеточной репарации (Pierce et al., 2001; Bindra et al., 2013).

В первоначальной модели (Pierce et al., 1999) плазида pDR содержит две «неполноценные» копии гена *GFP*, разделенные геном селективного маркера пурамицина, которые не дают функционального белкового продукта, но при запуске процесса гомологичной рекомбинации экспрессия *GFP* восстанавливается. В первую копию, которая находится под промотором куриного β -актина, слитого для усиления эффекта с CMV-энхансером, встроены уникальный сайт узнавания редкой рестриктазы *I-SceI*,

содержащий несколько стоп-кодона и нарушающий правильную рамку считывания гена *GFP*. Поскольку ни один из известных геномов млекопитающих исходно не содержит сайтов узнавания данной рестриктазы, ее действие может быть направлено исключительно на векторную конструкцию. Вторая копия представлена геном *GFP*, усеченным с 5' и 3'-концов, но содержащим центральную кодирующую часть, которая как раз нарушена в первой копии. В обоих вариантах гена *GFP* присутствуют гомологичные друг другу последовательности, разделенные 3.7 т.п.н., что в случае повреждения плазмидной ДНК направляет репарацию по механизму гомологичной рекомбинации.

При образовании двуцепочечного разрыва на месте сайта рестрикции *I-SceI* существует несколько альтернативных вариантов развития репарационных событий. В первом случае может происходить некроссоверный вариант репарации двуцепочечных разрывов (DSBR) – процесс, при котором разрыв репарируется по гомологичному аналогу, представленному второй копией *GFP*, с образованием классических структур Холлидея, но разрешающихся без обмена участками. Во втором случае запускается путь синтез-зависимого отжига цепей, при котором также не происходит обмена участками. Продуктом в обоих случаях будет появление функционального белка *GFP*.

Альтернативным продуктом гомологичной рекомбинации является транскрибированный по 3'-концу белок *GFP*, который может образоваться также несколькими путями: по кроссоверному варианту DSBR-репарационных событий с образованием промежуточных структур Холлидея и последующим обменом фланкирующими разрыв участками или по механизму гомологичной репарации путем отжига одиночной цепи, при котором целиком деградирует фрагмент между гомологичными участками. Для исследовательских целей второй вариант продукта не представляет большого интереса, поскольку не может быть идентифицирован так же легко, как появление флуоресцентного сигнала, но при этом и не вносит препятствий в оценку эффективности работы нуклеазы.

Для того чтобы использовать pDR-систему при оценке эффективности работы комплекса «Cas9/направляющая химерная РНК», вместо сайта *I-SceI* необходимо встроить выбранную целевую последовательность, которая бы распознавалась Cas9-нуклеазой. Белок Cas9, попадая в клетку и связываясь с направляющей молекулой РНК, будет узнавать целевой фрагмент и направленно вносить двуцепочечный разрыв в предварительно доставленную в клетки pDR-векторную конструкцию. В случае корректной и стабильной работы системы гомологичной рекомбинации в клетке будет возникать зеленая флуоресценция уже на вторые сутки после этапа трансфекции. Таким образом, эффективность работы системы CRISPR/Cas9 для конкретного геномного локуса может быть оценена на проточном цитометре прямо пропорционально количеству *GFP*-позитивных клеток.

Тестовый эксперимент с pDR-системой можно разделить на два основных этапа: предварительная подготовка векторных конструкций на основе плазмиды pDR и проведение эксперимента на клетках с последующей детекцией *GFP*-позитивных клеток.

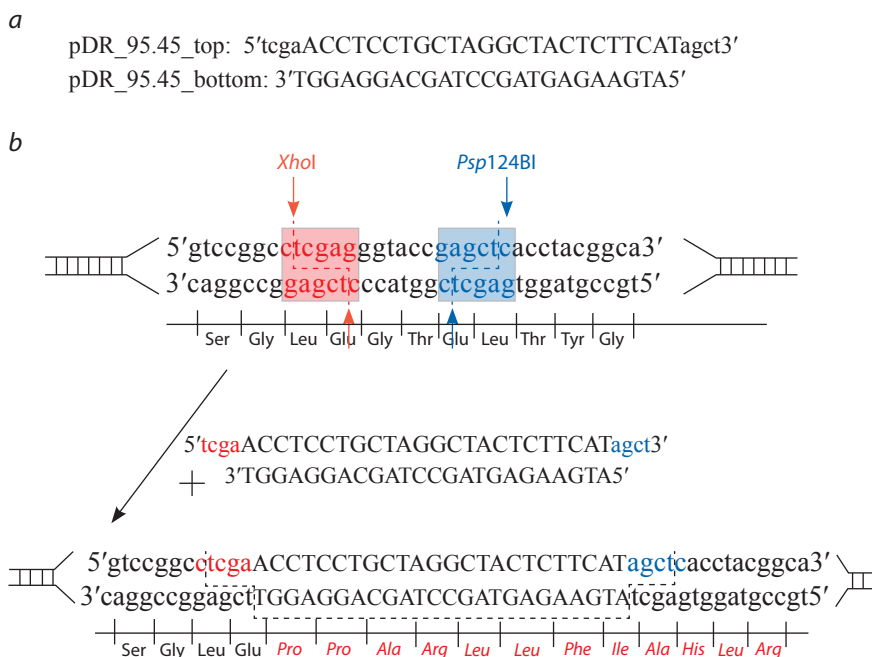


Fig. 2. pDR plasmid cloning. (a) Oligonucleotide primers used for synthesis of the target fragment, whose annealing results in the formation of a template flanked by *Psp124BI* and *XhoI* restriction sites and recognized by Cas9 nuclease. (b) The insertion of the target fragment into the linear pDR-vector through the *XhoI* and *Psp124BI* restriction sites and further shift of the *gfp* gene reading frame as a result of the insertion.

Материалы

- Плазмидные векторы pDR GFP univ (Addgene, #46085), hCas9 (вектор для экспрессии гена *Cas9*, Addgene #41815), pgRNA (вектор для экспрессии хнРНК, полученный на этапе 2).
- Олигонуклеотиды. Необходимо заказать синтез двух одноцепочечных олигонуклеотидов (top и bottom) такого состава, чтобы продукт их отжига был фланкирован сайтами рестриктаз *XhoI* и *Psp124BI* и содержал целевую последовательность для узнавания Cas9-нуклеазой (рис. 2, а). Встройка синтезированного фрагмента в первую из копий гена *GFP* pDR-плазмиды должна приводить к смещению рамки считывания данного гена и, как следствие, к нарушению экспрессии зеленого белка (см. рис. 2, б).
- Ферменты *XhoI* (Neb, R0146S), T4-лигаза (Neb, M0202), *Taq*-полимераза (Sibenzyme, E337), *Psp124BI* (Sibenzyme, E107), буфер Cut Smart (Neb, B7204S).
- NaCl (растворы с концентрациями 5 мМ и 1 М).
- Буфер TE (10 мМ Tris HCl, 1 мМ EDTA).
- Компетентные клетки, например XL1-Blue Supercompetent Cells (Neb, C2992).
- Клеточная линия Phoenix (доступна в лаборатории генетики развития ИЦиГ СО РАН либо под заказ ATCC® CRL-3215™, ATCC).
- Ампициллин (развести сток 100 мг/мл) (Sigma, A9393).
- Pen/Strep (смесь пенициллин–стрептомицин, рабочая концентрация 100 ед/мл пенициллин, 100 мкг/мл стрептомицин, 11074440001 Roche).
- Набор для выделения плазмидной ДНК (Biosilica PLD-100 или аналог).
- Липофектамин Lipofectamine® 3000 Transfection Reagent (Thermo Fisher Scientific, L3000001).
- Бактериальная среда LB (описана выше).
- Среда Opti-MEM (Opti-MEM® I Reduced Serum Medium, Thermo Fisher Scientific, 11058021), DMEM (DMEM, high glucose, GlutaMAX™ Supplement, HEPES, Thermo Fisher Scientific, 32430027).
- Fetal Bovine Serum (Thermo Fisher Scientific, 10270).

- GlutaMAX™ Supplement (Thermo Fisher Scientific, 35050061).
- Среда для клеток Phoenix (DMEM + 10 % FBS + 1 % GlutaMax + 1 % PenStrep).

Оборудование

- Электропоратор, в случае использования электрокомпетентных клеток (MicroPulser Electroporator, Bio rad или аналог), и специализированные кюветы (Gene Pulser/MicroPulser Cuvette, Bio-Rad).
- Амплификатор (T100™ Thermal Cycler, Bio-Rad или аналог).
- Термостат с температурным диапазоном 25–95 °C.
- Пластиковые чашки Петри ($d = 10$ см).
- Флуоресцентный микроскоп Axioskop 2 Plus (Carl Zeiss Microscopy, LLC, США) либо аналог.
- Проточный цитометр BD FACS Aria II (BD Biosciences, США) или аналог с возможностью детекции сигнала GFP.

Протокол дизайна и клонирования плазмидных конструкций pDR-тестовой системы

1. Развести одноцепочечные олигонуклеотиды до концентрации 100 мМ. Для отжига в 1.5 мл пробирку добавить ~1.5 мкл каждого из олигонуклеотидов (2–4 мкг), 5 мкл NaCl (стоковая концентрация 5 мМ) и довести объем смеси до 50 мкл буфером TE. Инкубировать реакционную смесь при 95 °C в течение 5 мин, затем снижать температуру на 0,5 °C каждые 20 с до достижения 25 °C. В дальнейших реакциях использовать рабочую концентрацию двуцепочечных олигонуклеотидов 20 нг/мкл. Хранить смесь при +16 °C.
2. Для линейаризации плазмиды и получения липких концов обработать вектор pDR GFP univ эндонуклеазами рестрикции *XhoI* и *Psp124BI*. Для этого в одной пробирке смешать 2 мкл буфера Cut Smart, по 1 мкл каждой из рестриктаз *XhoI* и *Psp124BI*, 2–3 мкг плазмидной ДНК и довести объем реакции до 20 мкл добавлением ddH₂O. Инкубировать 2–4 ч при 37 °C.
3. Инактивировать ферменты нагреванием при 80 °C в течение 20 мин.
Мы не рекомендуем дальнейшую очистку вектора от побочных продуктов рестрикции (с использовани-

ем специальных колонок или методом выделения специфического бэнда из агарозного геля), поскольку в этом случае происходят большие потери материала, а эффективность реакции лигирования принципиально не изменяется.

4. На основе линеаризованного плазмидного вектора pDR GFP univ собрать векторную конструкцию со встройкой целевой нуклеотидной последовательности, распознаваемой Cas9-нуклеазой, поставив лигазную реакцию с T4-лигазой по предложенному протоколу. Для этого в пробирку добавить 2 мкл лигазного буфера, 1 мкл T4-лигазы, 30–50 нг вектора, 10–20 нг целевого двуцепочечного олигонуклеотида и довести объем реакции до 20 мкл добавлением ddH₂O. Рекомендуемое соотношение клонирующего вектора к олигонуклеотидному встраиваемому фрагменту 1:3. Инкубировать при комнатной температуре в течение 10–20 мин.
5. Инактивировать действие лигазы нагреванием до 65 °C в течение 10 мин.
6. Для наработки целевой плазмидной конструкции необходимо трансформировать бактериальные компетентные клетки инактивированной лигазной смесью (см. пункт III этапа 2; используемые для трансформации клетки НЕ ДОЛЖНЫ быть устойчивыми к ампициллину). Для трансформации желателен трехкратный разбавить инактивированную лигазную смесь добавлением ddH₂O, чтобы уменьшить количество солей, исходно присутствующих в лигазном буфере. Повышенное количество ионов лигазного буфера негативно сказывается на эффективности процесса электропорации. На 40 мкл электрокомпетентных клеток рекомендуется использовать 1.5 мкл разбавленной инактивированной лигазной смеси. Высадить трансформированные бактерии необходимо на чашку Петри на твердую агаризованную среду LB (20–25 мл), содержащую селективный антибиотик ампициллин с конечной концентрацией 100 мкг/мл. Затем убрать чашки с бактериями в термостат и инкубировать при 37 °C в течение 12–18 ч.
7. На следующий день следует проанализировать бактериальные колонии на наличие клонирующего вектора с целевой встройкой методом ПЦР с использованием следующих праймеров: forward – использовать top-праймер, заказанный для синтеза целевого олигонуклеотида (пример на рис. 2, а), reverse – 5'ctgttacagctcgctccatgc3'. Рекомендуется использовать классический протокол приготовления ПЦР-реакции на 25 мкл, за тем исключением, что вместо выделенной ДНК использовать небольшой фрагмент бактериальной колонии. Удобно осуществлять перенос фрагмента анализируемой колонии носиком автоматической пипетки, затем тщательно пипетировать полученную смесь. Оптимальные условия ПЦР для данного набора праймеров и Taq-полимеразы приведены в табл. 1. Целевой ПЦР-продукт должен иметь размер 636 п. н. и может быть детектирован методом гель-электрофореза в 2 % агарозном геле.
8. Для амплификации целевой векторной конструкции со встройкой выбрать единичную позитивную колонию клеток используемого штамма и поставить с нее ночную культуру на питательной среде LB (5–10 мл) в присутствии селективного антибиотика ампициллина с конечной концентрацией 100 мкг/мл.

Table 1. Basic PCR protocol for verification of correct insertion into the pDR cloning vector

Step No.	Temperature, °C	Duration, sec	Number of cycles
I	95	300	1
II	95	30	25–30
	59	30	
	72	60	
III	72	300	1
IV	16	∞	

The optimum number of reaction cycles depends on the initial amount of plasmid DNA. Usually, 25 cycles suffice.

Table 2. Sanger sequencing reaction conditions

Step No.	Temperature, °C	Duration, sec	Number of cycles
I	96	1	1
II	96	10	40
	50	10	
	60	240	

Prior to the loading onto the sequencer, the reaction mix should be purified by gel filtration, e. g., with Sephadex G-50 (semisynthetic sorbent, highly hydrophilic dextran).

9. Выделить плазмидную ДНК при помощи специализированного набора для выделения по рекомендованному протоколу.
10. Проверить полученную векторную конструкцию на основе pDR GFP univ на наличие необходимой встройки, а также на целостность промоторного региона секвенированием по Сэнгеру целевых локусов. Использовать любой из указанных выше праймеров (forward или reverse) для контроля правильности встройки. Рекомендуемые параметры реакции секвенирования приведены в табл. 2.

Оценка эффективности работы системы CRISPR/Cas9 при помощи плазмидных конструкций pDR-тестовой системы

После того как все pDR-конструкции, а также ключевые компоненты системы CRISPR/Cas9 собраны и проверены на соответствие референсным последовательностям, можно приступать к тестовому эксперименту на клеточных культурах. Для дальнейшей работы обычно используются клетки HEK293 или их аналоги (Phoenix, 293T), которые быстро делятся и эффективно трансфицируются.

1. Предварительно, за несколько дней до эксперимента по трансфекции, необходимо разморозить клетки и нарастить их до рекомендуемой плотности (70–90 %). Мы используем рассадку на 12-луночных планшетах по (4–8) × 10⁵ клеток на лунку.
2. Провести трансфекцию липофектаминоом 3000 по стандартному протоколу с использованием P3000™ Enhancer-реагента, с максимально рекомендованной

концентрацией липофектамина, на среде Opti-MEM. Следует использовать соотношение плазмидных векторов 2.4:1:2.2, рассчитанное исходя из размеров конструкций, Cas9 (9553 п. н.):gRNA (3975 п. н.):pDR (8661 п. н.).

4. Инкубировать клетки в течение 6–8 ч в термостате при стандартных условиях (37 °C и 5 % CO₂), затем сменить среду на специализированную для фибробластов и оставить инкубироваться еще на 20–36 ч.
5. Появление зеленого свечения можно детектировать с использованием флуоресцентного микроскопа через 10–12 ч после этапа трансфекции.
6. По нашим оценкам, свечение GFP достигает максимума на вторые сутки после трансфекции. Именно в это время оптимально оценивать количество светящихся клеток на проточном цитометре. По результатам FACS-анализа (Fluorescence Activated Cell Sorting – активированная флуоресценцией сортировка клеток; несмотря на общее название, сортировка клеток в этом эксперименте не требуется, проводится только подсчет доли флуоресцирующих клеток) можно судить об эффективности действия комплекса Cas9/направляющая РНК – чем больше клеток экспрессируют зеленый флуоресцентный белок, тем лучше работает Cas9-нуклеаза, внося двуцепочечный разрыв в тестовую pDR-конструкцию, – и таким образом отбирать наиболее перспективные направляющие РНК.

Тестирование хнРНК на основе детекции разрушения сайта эндонуклеазы рестрикции в таргетированном районе

Эффективность внесения разрывов системой CRISPR/Cas9 может зависеть от многих параметров: нуклеотидной композиции хнРНК, доступности хроматина в данном типе клеток, эффективности доставки компонентов системы в клетки и времени их жизни и т. д. Наиболее адекватно отражают эффективность системы CRISPR/Cas9 тесты, проводимые непосредственно на культуре целевых клеток.

Предлагаемый тест основан на том, что при внесении двуцепочечного разрыва и его репарации по механизму NHEJ в клетках возникают мутации, разрушающие сайт узнавания какой-либо эндонуклеазы рестрикции. Следует отметить, что при этом возможно тестирование только тех хнРНК, для которых на расстоянии 3–4 нуклеотидов от РАМ (в месте внесения двуцепочечного разрыва) расположен сайт узнавания эндонуклеазы рестрикции. Для тестирования других хнРНК следует использовать систему pDR, описанную выше.

Тест проводят следующим образом: в клетки вводят компоненты системы CRISPR/Cas9, спустя несколько дней выделяют геномную ДНК. Участок, фланкирующий место внесения разрыва, амплифицируют, после чего проводят рестрикционный анализ соответствующей эндонуклеазой рестрикции. Наличие негидролизованного продукта свидетельствует об эффективной работе системы CRISPR/Cas9.

Материалы и оборудование

- Олигонуклеотиды. На основе последовательности модифицируемого генома необходимо выбрать праймеры

(gRNA_test_F и хнРНК_test_R), фланкирующие таргетируемый район. Мы рекомендуем программу NCBI Primer Blast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) для дизайна праймеров. При дизайне праймеров нужно учитывать следующие рекомендации:

- размер ПЦР-продукта должен находиться в диапазоне 200–800 п. н.;
- сайт(ы) эндонуклеазы рестрикции, используемой в тесте, должны находиться достаточно далеко от начала и конца ПЦР-продукта (т. е. в результате гидролиза размер ПЦР-продукта изменяется более чем на 10 %);
- если в составе ПЦР-продукта сайт гидролиза эндонуклеазы рестрикции встречается более одного раза, все продукты гидролиза должны иметь различную длину и могут быть визуализированы в агарозном геле.
- Материалы и оборудование, необходимые для культивирования клеток млекопитающих и проведения трансфекций (описаны выше).
- Стандартные материалы (например, Sigma G1N10) и оборудование для выделения ДНК и проведения ПЦР.
- Фермент рестрикции и соответствующий ему буфер. Сайт узнавания выбранного фермента рестрикции должен присутствовать в редактируемом геноме на расстоянии 3–4 нуклеотидов от РАМ. Кроме того, при выборе фермента рестрикции следует учитывать рекомендации, относящиеся к дизайну ПЦР-продукта (приведены выше).
- Плазмидные векторы, экспрессирующий хнРНК (приготовление описано в рамках этапа 2) и Cas9 (pCas9, описан выше).

Протокол

1. Провести трансфекцию клеток плазмидными векторами, экспрессирующими хнРНК и Cas9 (с использованием реактива Lipofectamin 3000, как указано в протоколе производителя). Параллельно проводить культивирование контрольных клеток, не трансфицированных компонентами системы CRISPR/Cas9.
2. Инкубировать клетки 2–4 сут, при необходимости выполнять смену среды и пассирование клеток.
3. Выделить из контрольных и трансфицированных клеток геномную ДНК. Геномная ДНК может быть выделена при помощи коммерчески доступных наборов или фенол-хлороформной экстракцией.
4. Для каждого образца геномной ДНК провести ПЦР с праймерами хнРНК_test_F и хнРНК_test_R в объеме 50 мкл. Использовать 10 мкл ПЦР-продукта для анализа в агарозном геле. В результате реакции должен с высокой эффективностью образовываться единственный продукт.
5. Разделить оставшиеся 40 мкл продукта ПЦР на две части. Подготовить тестовую реакцию и отрицательный контроль:

Тестовая реакция:	Отрицательный контроль:
Продукт ПЦР 20 мкл*	Продукт ПЦР 20 мкл
Буфер эндонуклеазы рестрикции 2.3 мкл	Буфер эндонуклеазы рестрикции 2.3 мкл
Эндонуклеаза рестрикции 1 мкл	ddH ₂ O 1 мкл

*Для большинства ферментов очистка продуктов ПЦР перед гидролизом не требуется. Однако при возникновении сложностей с рестрикционным гидролизом мы рекомендуем

провести очистку ПЦР-продукта стандартным методом (например, переосаждением этиловым спиртом).

6. Инкубировать реакции в оптимальных для используемого фермента условиях 12–24 ч. Оценить количество и размер продуктов гидролиза в агарозном геле.

В случае корректной работы системы CRISPR/Cas9 ожидается неполный гидролиз ПЦР-продуктов в образцах, полученных из трансфицированных клеток. В случае контрольных клеток гидролиз должен быть полным. Гидролиз должен наблюдаться только в тестовых реакциях, в случае отрицательного контроля ПЦР-продукт должен оставаться интактным.

Следует отметить, что даже для высокоэффективных хнРНК мы наблюдали лишь небольшую долю негидролизованного продукта в тестовых реакциях. Надо учитывать, что эффективность трансфекции многих типов клеток далека от 100 %. Кроме того, при использовании диплоидных культур одна и та же клетка может давать как модифицированный, так и немодифицированный продукт ПЦР, если модификация затрагивает лишь один из аллелей.

Этап 4. Модификация генома клеток *in vitro*

Внесение генетических изменений удобно проводить *in vitro* на культурах клеток. В этом случае можно осуществлять выбор клонов, в которых произошло желаемое изменение генома, и даже последовательно вносить несколько генетических модификаций. В зависимости от целей экспериментатора есть две основные стратегии селекции клонов, которые можно сочетать: отбор события попадания в клетку генетической конструкции, в которой работает белок Cas9, например для получения INDEL или делеций протяженных участков ДНК (типы модификаций 1 и 3 описаны выше); и отбор события модификации генома с выбором клонов на последующих этапах культивирования (тип модификации 2).

В зависимости от цели и удобства для отбора генетически модифицированных клонов можно использовать флуоресцентные белки (GFP и т. д.) и/или устойчивость к антибиотикам. При высокой эффективности доставки компонентов системы CRISPR/Cas9 в клетки и ее активности можно вообще отказаться от проведения селекции, а все полученные клоны анализировать на присутствие генетической модификации.

Селекция попадания в клетку компонент системы CRISPR/Cas9

Для того чтобы сразу отобрать клоны с возможной генетической модификацией, используем кольцевые плазмиды, несущие ген *Cas9*. Известно, что случайная негомологичная встройка кольцевой ДНК менее эффективна, чем встройка линейного фрагмента (Brinster et al., 1985). При использовании плазмиды Cas92AGFP (Addgene #48138), кодирующей GFP и Cas9, соединенные через 2A пептид, по свечению GFP на первый-второй день после трансфекции можно отобрать клетки с экспрессией белка Cas9. Для отбора клеток мы используем BD FACSAria II (BD Biosciences, США). После сортировки GFP-позитивных клеток их можно редко рассадить для получения индивидуальных колоний или же сразу про-

вести сортировку единичных клеток в 96-ячеечное плато для дальнейшего культивирования и анализа генетической модификации.

Еще один вариант – использование плазмиды pHCas9 (Addgene #41815), коэкспрессирующей Cas9 и ген устойчивости к антибиотику генетицину (G418). С каждым делением клеток плазида разбавляется и разрушается ферментами. Пока в клетках остается плазида, можно провести экспресс-селекцию на устойчивость к определенному антибиотику.

Селекция направленной модификации генома

Следующий вариант селекции – использование для отбора клеток экспрессии введенной конструкции. При необходимости можно ввести в геном GFP или другой флуоресцентный белок и отбирать клоны со встройкой. Для упрощения селекции (или в случае, когда введение флуоресцентного белка в геном нежелательно) используется конструкция, обеспечивающая устойчивость к антибиотику. Добавление в среду для культивирования антибиотика позволяет оставить только устойчивые клоны, несущие генетическую конструкцию.

Однако необходимо учитывать, что вводимая генетическая конструкция может встроиться не только в целевой локус, но и в случайное место в геноме. По нашим данным, в ряде случаев частота такой «случайной» встройки сопоставима с частотой встройки в таргетируемый участок. При этом клетки, в которых встройка происходит в случайный район генома, также будут устойчивы к действию селективного антибиотика. Поэтому необходимо во всех случаях проводить анализ полученных клонов на присутствие необходимой модификации генома.

Рассмотрим протокол генетической модификации на примере эмбриональных стволовых (ЭС) клеток мыши.

Материалы

- Среда для ЭС клеток мыши: среда DMEM с 4.5 г/мл глюкозы (Thermo Fisher Scientific, 32430-100), 7.5 % FBS для ЭС клеток (Thermo Fisher Scientific, 16141079), 7.5 % KSR (Thermo Fisher Scientific, 10828-028), 1X GlutaMAX (Thermo Fisher Scientific, 35050038), 1X NEAA (Thermo Fisher Scientific, 11140050), 1X пенициллин-стрептомицин (Thermo Fisher Scientific, 15140122), 1000 ед/мл LIF (PolyGene, PG-A1140-0100).
- 2i ингибиторы: PD0325901 (Sigma-Aldrich, PZ0162), стоковая концентрация 1 мМ, рабочая концентрация 1 мкМ; CHIR99021 (Sigma-Aldrich, SML1046), стоковая концентрация 3 мМ, рабочая концентрация 3 мкМ.
- Трипсин-ЭДТА 10X 0.5 % (Thermo Fisher Scientific, 15400054), рабочая концентрация 0.05 %.
- Желатин (Sigma, G1890).
- Митомицин С (Sigma, M4287). Стоковая концентрация 200 мкг/мл, рекомендуемая рабочая концентрация 10 мкг/мл, мы используем 5 мкг/мл. Продолжительность обработки эмбриональных фибробластов мыши 2–3 ч. Так как митомицин С очень токсичен, после обработки необходимо не менее трех раз промыть клетки фосфатным буфером.
- Opti-MEM I (Thermo Fisher Scientific, 11058021).
- Пуромицин (Thermo Fisher Scientific, A1113803). Стоковая концентрация 1 мг/мл, рекомендуемая рабочая

- концентрация 1–10 мкг/мл, мы используем 2 мкг/мл. Продолжительность селекции 2–7 дней, в день пересева клеток добавлять антибиотик нельзя.
- Генитинин (сульфат G418) (Sigma, A1720). Стоковая концентрация 100 мг/мл, рекомендуемая рабочая концентрация 100–1000 мкг/мл, мы используем 400 мкг/мл. Продолжительность селекции не менее 7 дней.
 - Липофектамин 3000 (Thermo Fisher Scientific, L3000001).
 - Среда для заморозки клеток млекопитающих: 90 % KSR (см. выше) и 10 % ДМСО (Amresco, Am-0231). Заморозка клеток проводится в криоконтейнере (Thermo Fisher Scientific, 5100-0001), хранение в жидком азоте.
 - Линия ЭС клеток. Линию ЭС клеток DGES1 получали по следующему протоколу (Bryja et al., 2006). Она доступна в лаборатории генетики развития ИЦиГ СО РАН и в «Коллекции клеточных культур плюрипотентных клеток человека и животных для проведения фундаментальных исследований в области биомедицины и генетики развития» ИЦиГ СО РАН.

Протокол

1. ЭС клетки культивировать в среде для ЭС клеток. Для пересадки использовать 0.05 % трипсин-ЭДТА. В зависимости от задачи ЭС клетки рассаживать на пластик, покрытый только 0.1 % желатином, или дополнительно использовать инактивированные митомицином С эмбриональные фибробласты мыши (фидер) в концентрации 15 тыс. кл./см². ЭС клетки лучше растут в присутствии фидера, однако при добавлении антибиотиков фидер нельзя использовать. Свойства ЭС клеток при переводе с фидера на желатин меняются. Если ЭС клетки полностью изменяют морфологию, можно использовать 2i ингибиторы или кондиционированную среду: через день после рассадки фидера промыть клетки фосфатным буфером и добавить ЭС среду. На следующий день собрать кондиционированную среду, профильтровать через фильтр 0.22 мкм и добавить LIF.
2. ЭС клетки рассаживать так, чтобы на следующий день (день проведения трансфекции) плотность культуры была 50–80 %. Трансфекцию проводить с помощью липофектамина 3000 в соответствии с рекомендациями производителя, использовать рекомендуемое производителем количество P3000 реагента и среднее от рекомендованного липофектамина 3000. Для трансфекции смешать плазмиды pHCas9 (или Cas92AGFP) и pgRNA (приготовлена на этапе 2) в соотношении 1 : 1 (для внесения INDEL) или pHCas9 (или Cas92AGFP), pgRNA и вектор для гомологичной рекомбинации в соотношении 1 : 1 : 1 (для получения встройки конструкции по механизму гомологичной рекомбинации). Смесь ДНК, P3000 и липофектамина 3000 готовить в среде Opti-MEM I, добавлять к клеткам в ЭС среде, культивировать до следующего дня и на следующий день сменить среду.
3. Селекция клонов.
 - А. Селекция на попадание в клетку системы CRISPR/Cas9. На следующий день после трансфекции клетки рассадить и провести отбор с помощью FACS или добавлением антибиотика. Полученные клоны нарастить и провести анализ генетической модификации.

- Б. Селекция на направленную модификацию генома. На следующий день после трансфекции клетки рассадить в соотношении 1 : 5 – 1 : 20 и селекцию провести с помощью антибиотика или, в случае встройки генетической конструкции с флуоресцентным белком, по свечению колоний. Полученные клоны нарастить и провести анализ генетической модификации. Для анализа генетической модификации нарастить отобранные клоны до необходимого количества – обычно одна лунка планшета на 24 ячейки (2 см²) – и разделить культуру на две части: половину заморозить, половину использовать для выделения ДНК. Геномная ДНК может быть выделена при помощи коммерчески доступных наборов или фенол-хлороформной экстракцией.

4. Анализ генетической модификации.

- А. При анализе на присутствие INDEL разработать праймеры, фланкирующие область генетической модификации. Экспресс-анализ клонов можно проводить с помощью электрофореза в 3 % агарозном геле. Желательно, чтобы размер продукта ПЦР был около 100 п. о., в этом случае различная длина аллелей может быть выявлена электрофорезом. В случае, если генетическая модификация разрушает известный сайт рестрикции, можно использовать рестрикционный анализ. Для подтверждения генетической модификации необходимо провести секвенирование.
- Б. При анализе направленной модификации генома в случае инсерции генетического материала необходимо разработать праймеры, фланкирующие обе границы трансгена. То есть один из пары праймеров должен быть комплементарен последовательности генома за пределами плечей гомологии, использовавшихся в конструкции для гомологичной рекомбинации, а второй праймер должен быть комплементарен участку трансгена и не иметь районов комплементарности в интактном геноме. При таком дизайне праймеров присутствие продукта ПЦР нужного размера возможно только при встройке в ожидаемое место. Правильность встройки необходимо подтвердить секвенированием. При анализе делеции разработать праймеры, фланкирующие ожидаемое место разрыва, провести ПЦР и в случае присутствия продукта ожидаемой длины подтвердить секвенированием.
- В. Известно, что система CRISPR/Cas9 может разрезать участки генома с частичной гомологией к хнРНК (Fu et al., 2013). Для проверки таких событий можно провести полногеномное секвенирование или секвенировать потенциальные сайты, которые могут быть разрезаны системой (имеющие 15 и более букв, гомологичных хнРНК). В случае вставки в геном генетической конструкции дополнительные копии можно обнаружить с помощью ОТ-ПЦР в реальном времени, Digital PCR (http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf) или с помощью инвертированного или TAIL ПЦР (Liu, Chen, 2007).

Этап 5. Создание трансгенных животных с использованием системы CRISPR/Cas9

После появления системы редактирования CRISPR/Cas9 существенно упростилось получение нокаутных и транс-

генных животных, в частности мышей. Существует три разных способа для редактирования генома при помощи системы CRISPR/Cas9 путем микроинъекции в зиготы. Принципиальные отличия в этих процедурах заключаются в белке Cas9. В первом случае это введение мРНК Cas9 совместно с хнРНК (Yang et al., 2014). Поскольку вводится матричная РНК, то нет необходимости в инъекции в пронуклеус, что технически упрощает процедуру. Поэтому смесь из РНК (мРНК Cas9 и хнРНК) вводится в цитоплазму зигот. Во втором случае система CRISPR/Cas9 представлена готовым комплексом из белка Cas9 и хнРНК (Kouranova et al., 2016). В этом случае нет необходимости в трансляции и наработке белка Cas9 в цитоплазме, и комплекс вводится непосредственно в пронуклеус. В третьем случае используется трансгенная мышь, которая несет эндогенный Cas9, и белок уже присутствует в зиготе в пронуклеусе (Sakurai et al., 2016). Поэтому вводится не вся система CRISPR/Cas9, а только хнРНК. Мы остановимся на первом варианте, потому что он технически наиболее простой.

Получение хнРНК

Материалы и оборудование

- HiScribe T7 High Yield RNA synthesis Kit (NewEngland BioLabs, E2040S).
- QIAprep Spin miniprep Columns (QUAGEN, 27115).
- RNA Clean and Concentrator 25 (ZymoResearch, R1017).
- Матричная РНК Cas9 (ThermoFisher A29378).
- Спектрофотометр NanoDrop 2000 Spectrophotometer (Thermo Scientific).

Для получения хнРНК в первую очередь надо определить целевой участок генома и последовательность для хнРНК (см. этап 1). Затем необходимо синтезировать матрицу для *in vitro* транскрипции, как описано в (Kistler et al., 2015).

После получения ДНК-матрицы производят ее очистку на колонках (QIAprep Spin miniprep Columns). Далее, следуя протоколу от производителя, получают хнРНК, используя набор для проведения *in vitro* транскрипции (HiScribe T7 High Yield RNA synthesis Kit). Перед внесением ДНК-матрицы в реакцию *in vitro* транскрипции важно проверить ее при помощи гель-электрофореза, а также измерить конечную концентрацию продукта на спектрофотометре.

После получения хнРНК необходимо отобрать аликвоту для последующего гель-электрофореза и далее очистить ее при помощи набора (RNA Clean & Concentrator) в соответствии с рекомендациями производителя. Очищенную хнРНК и отобранную до очистки аликвоту хнРНК оценивают при помощи гель-электрофореза. Количество хнРНК оценивают на спектрофотометре NanoDrop 2000. Полученную РНК следует хранить при температуре -80°C .

Предварительно можно приготовить готовые рабочие растворы для инъекции из расчета 50 нг/мкл мРНК Cas9 и 10 нг/мкл хнРНК. Для удобства и экономии растворов используются пробирки объемом 200 мкл, в которые помещается по 15 мкл раствора РНК. Предварительно пробирки необходимо промыть водой (RNase free) для удаления микрочастиц пластика. Полученные аликвоты

готовых растворов хранить при -80°C . После инъекции аликвоты рабочих растворов повторно не замораживать.

Работа с животными. Микроинъекция

Материалы и оборудование

- Гибридные мыши первого поколения C57Bl/6 $\times$ CBA/Lac или C57Bl/6 $\times$ DBA/2JrccHsd, самки, возраст два месяца.
- Мыши линии C57Bl/6, самцы.
- Мыши линии CD-1, самцы и самки.
- PMSG (Фоллигон (Folligon), Интервет Интернэшнл Б.В., Нидерланды), препарат доступен в ветеринарных аптеках.
- hCG (Хорулон (Chorulon), Интервет Интернэшнл Б.В., Нидерланды), препарат доступен в ветеринарных аптеках.
- Домитор (методомина гидрохлорид) (Orion Corporation, Finland), препарат доступен в ветеринарных аптеках.
- Золетил (VIRBAC, Франция). Препарат доступен в ветеринарных аптеках.
- Центрифуга Centrifuge 5810 R, Eppendorf или аналогичная.
- Стекланные капилляры (Harvard Apparatus, GC100-15).
- Пуллер (P-97, Sutter Instrument Co.).
- Micro Forge MF-90 (Narishige).
- Инвертированный микроскоп (Olympus IX71) с интерференционной оптикой Номарского или аналогичный.
- Держатель (холдер) Eppendorf™ VacuTip (кат. ном. 930001015).
- Наконечники Eppendorf™ Femtotips™ Microloader Tips for Femtojet Microinjector (кат. ном. 5242956003).
- Компрессор Transjector 5246, Eppendorf или аналогичный.
- Микроинъекционный шприц Narishige или аналогичный.
- Гидравлические манипуляторы Narishige или аналогичные.
- Среда для культивирования M2 (Sigma, M7167).
- Среда для культивирования M16 (Sigma, M7292).

Животные. Супероуляция. Получение зигот

Для проведения эксперимента по инъекции системы CRISPR/Cas9 в цитоплазму зигот мыши надо предварительно подготовить необходимое количество животных. В первую очередь нужны животные-доноры зигот. Для этих целей подходят любые линии мышей, но наиболее приемлемой схемой является получение зигот от самок-гибридов первого поколения (C57Bl/6 $\times$ CBA/Lac или C57Bl/6 $\times$ DBA/2JrccHsd), покрытых самцами C57Bl/6. Возможно применение других линий мышей, но эффективность супероуляции и выживаемости после цитоплазматической микроинъекции при этом будут ниже. Для более эффективной работы следует использовать протокол супероуляции по следующей схеме.

В первый день самкам-донорам производится внутрибрюшинная инъекция 7.5 ед. PMSG в 16:00. На третий день этим же самкам производится внутрибрюшинная инъекция 7.5 ед. hCG в 13:00. Наиболее результативной супероуляция будет в случае использования самок двух-

месячного возраста. Подсадка стимулированных самок к самцам производится в 17:00. Важно не производить смену подстилки у мышей перед тем, как производится подсадка, а самки по одной должны подсаживаться к самцам, а не наоборот. Данная схема подходит для искусственного 12-часового цикла со световым временем суток с 7:00 до 19:00. Для удобства дозирования препараты для супероуляции следует дополнительно развести физиологическим раствором до рабочей концентрации 0.05 ед./мкл, расфасовать полученные растворы в пробирки 1.5 мл и хранить при -80°C . Растворы не подлежат повторной заморозке, после размораживания их надо хранить на льду. Эффективность супероуляции при использовании данных гормонов в среднем составляет 30 зигот на одну гибридную самку. На четвертый день с 9:00 до 10:00 производится осмотр самок мышей на наличие пробок (Hogan et al., 1994). Самки с пробками используются для получения зигот с 10:00 до 11:00. При расчете необходимого количества зигот для одного дня эксперимента следует учитывать, что покрытие самок самцами происходит не в 100 % случаев и зависит от индивидуальных особенностей конкретного самца. Поэтому следует вести дневник покрытия самок самцами на этикетках клеток, отмечая дату покрытия. При постоянной работе с каждым из самцов можно отобрать группу наиболее активных, тем самым повысив долю покрытых самок. В среднем для одного дня работы достаточно 8–10 супероулированных самок. При проведении эксперимента следует опираться также на личный опыт, поскольку влияние оказывают и такие факторы, как корм, условия содержания, возраст и линия животных и пр.

Помимо мышей-доноров, для эксперимента по инъекции системы CRISPR/Cas9 в цитоплазму зигот также необходимы суррогатные псевдобеременные самки. Для этих целей предпочтительнее использовать аутбредную линию мышей CD-1. Минимум за одну-две недели до начала эксперимента выполняется вазэктомия самцов CD-1 по стандартной методике (Hogan et al., 1994). Для упрощения операции вместо перевязывания семенных канатиков можно пережигать их при помощи паяльника с небольшим жалом. После вазэктомии самцов в течение одной-двух недель следует содержать с самками любых линий для тренировки, после этого данных самок использовать в экспериментах нельзя.

Для получения псевдобеременных животных самок подсаживают в клетки к вазэктомизированным самцам по 2–3 самки на одного самца в тот же день, когда производится подсадка супероулированных самок-доноров. В данном случае следует учитывать, что самки находятся в естественном овуляторном цикле, длительность которого четыре дня. Этот фактор является решающим для получения нужного количества псевдобеременных самок, но важна также активность самца. Как описано выше, следует отмечать активных самцов и выбраковывать неактивных. В среднем для успешной работы необходимо иметь 20–25 вазэктомизированных самцов. Осмотр самок на наличие пробок производится в то же время, что и осмотр супероулированных самок. Далее самки с пробками будут использоваться для проведения операции трансплантации эмбрионов.

Микроинъекция

После получения зигот их следует культивировать в каплях среды M16 под минеральным маслом в инкубаторе при 37°C и 5 % CO_2 . С правилами культивирования эмбрионов можно подробно ознакомиться в (Hogan et al., 1994). Микроинъекцию следует начинать после 13:00. До этого времени надо подготовить растворы для микроинъекции, иглы и оборудование к работе.

После размораживания раствор РНК необходимо хранить на льду. Перед тем как наполнять раствором РНК инъекционные иглы, его следует центрифугировать при 14000 g, 0°C , 10 мин.

Для получения игл для микроинъекции используются стеклянные капилляры, которые оттягивают на пуллере при следующих параметрах: Heat 535, Pull 180, Vel. 60, Time 125, Pressure 200, Ramp 497. Полученные иглы загибаются на Micro Forge. Готовые иглы не следует хранить более одной недели.

Микроинъекция осуществляется на инвертированном микроскопе с интерференционной оптикой Номарского, оснащенном гидравлическими манипуляторами Narishige. Для удержания зигот используются холдеры, которые подключаются к микроинъекционному воздушному шприцу. Иголка заполняется 2–3 мкл рабочего раствора смеси РНК при помощи удлиненных носиков (Eppendorf™ Femtotips™ Microloader Tips), а затем аккуратно обламывается о холдер. Следует добиться слома с острым краем иглы, диаметр скола иглы в каждом случае подбирается индивидуально. После этого при помощи максимального давления нужно протолкнуть раствор РНК на кончик иголки, чтобы не осталось пузырька воздуха. Важным фактором при инъекции является отсутствие сильного смятия зиготы иглой. Игла должна прокалывать блестящую оболочку достаточно легко.

Для создания положительного давления в игле используется компрессор (трансжектор). Обычное инъекционное давление 30–40 Па. Микроинъекция осуществляется в капле среды M2 (объем около 100 мкл) под минеральным маслом. Важно производить инъекцию на стекле, а не на пластике, из-за плохих преломляющих свойств последнего. Камеру для инъекции можно сделать самостоятельно, для этого к обычному гистологическому стеклу приклеиваются импровизированные бортики из культурального пластика, для того чтобы минеральное масло не стекало вниз и могло полностью покрыть каплю.

При микроинъекции в цитоплазму важно не повредить пронуклеусы зиготы и контролировать введение раствора в цитоплазму. Иголка не всегда может сразу пройти насквозь плазматическую мембрану, и тогда приходится прокалывать зиготу до противоположной стенки блестящей оболочки.

Признаком успешной инъекции является локальное расширение и разрежение цитоплазмы. Зиготы оптимально прокалывать партиями по 20–30 шт. Для этого их нужно предварительно отмыть от среды M16 в среде M2, а затем перенести в один из полюсов капли. Холдер и инъекционную иглу удобнее расположить чуть выше плоскости зигот, опускаясь только холдером и забирая по одной зиготе; прокалывать зиготы нужно в противоположном полюсе, создавая группу проинъекцированных эмбрионов.

После инъекции зиготы помещаются в капли с M16 для культивирования в инкубаторе.

Трансплантация зигот

Перед операцией подсадки эмбрионы следует культивировать более часа, для того чтобы остались только жизнеспособные зиготы. После этого нужно отобрать жизнеспособные зиготы по характерному блеску в стереомикроскопе и четкой границе между блестящей оболочкой и плазматической мембраной.

Перед подсадкой зиготы снова отмывают в среде M2 и подсаживают псевдобеременным матерям по 15–20 шт., как описано в (Hogan et al., 1994). Надо учитывать, что могут встречаться «ложные пробки», и при выполнении операции по подсадке эмбрионов необходимо удостовериться в наличие ампулы яйцевода. Если ампула отсутствует, то операцию производить не следует.

Операция проводится под действием общей анестезии. Для наркоза можно применять следующую схему. В качестве премедикации используется домитор в количестве 0.125–0.25 мкг на 1 г массы тела мыши. Спустя 10–15 мин после введения домитора осуществляют инъекцию золетила в дозировке 15–30 мг на 1 г массы тела мыши. При использовании этой схемы мыши входят в длительный глубокий сон. Недостатки данного наркоза – охлаждение тела животного и сухость слизистых глаз. Для предотвращения этих эффектов необходимо дополнительное обогревание мышей на термостолке при 37 °C либо использование тепла от лампы накаливания. Глаза следует накрыть небольшим лоскутом марлевой салфетки, смоченным физиологическим раствором. Препараты после разведения по инструкции производителя можно хранить в морозильной камере при –20 °C. Целесообразно сделать небольшие аликвоты данных медикаментов и размораживать их по мере необходимости. После размораживания аликвоты доводят физиологическим раствором до концентрации домитора 0.25 мкг на 7.5–10 мкл, а золетила – 30 мг на 7.5–10 мкл. После размораживания и разведения препараты не подлежат повторному замораживанию, но возможно хранение готовых растворов при +4 °C в течение недели. Данную схему анестезии можно использовать и при операции по вазэктомии самцов.

Как правило, после успешной трансплантации развивается беременность, которая заметна через две недели. Забор материала для генотипирования от родившихся детенышей можно проводить со второй недели после рождения.

Заключение

Ежегодно публикуются тысячи статей, в которых применяется технология CRISPR/Cas9. Несмотря на общее сходство, детали протоколов использования CRISPR/Cas9 системы в этих работах часто различаются. Все приведенные нами протоколы были многократно проверены в нашей лаборатории и показали себя как эффективное средство внесения направленных геномных модификаций. Надеемся, что публикация этих протоколов будет способствовать быстрому и успешному освоению технологии CRISPR/Cas9 научными коллективами лабораторий, специализирующихся в самых разнообразных областях биологии и биотехнологии.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research, project 16-04-01453a, and State Budgeted Project 0324-2015-0004. The authors are grateful to the Center for Genetic Resources of Laboratory Animals, Institute of Cytology and Genetics (RFMEFI61914X0005).

Conflict of interest

The authors declare no conflict of interest.

References

- Bindra R.S., Goglia A.G., Jasin M., Powell S.N. Development of an assay to measure mutagenic non-homologous end-joining repair activity in mammalian cells. *Nucl. Acids Res.* 2013;41(11):e115-e115. DOI 10.1093/nar/gkt255.
- Brinster R.L., Chen H.Y., Trumbauer M.E., Yagle M.K., Palmiter R.D. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *Proc. Natl. Acad. Sci. USA.* 1985;82(13):4438-4442.
- Bryja V., Bonilla S., Arenas E. Derivation of mouse embryonic stem cells. *Nat. Protoc.* 2006;1(4):2082-2087. DOI 10.1038/nprot.2006.355.
- Chari R., Mali P., Moosburner M., Church G.M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods.* 2015;12(9):823-826. DOI 10.1038/nmeth.3473.
- Datta S., Roychoudhury S., Ghosh A., Dasgupta D., Ghosh A., Chakraborty B., Roy S., Gupta S., Santra A.K., Datta S., Das K. Distinct distribution pattern of hepatitis B virus genotype C and D in liver tissue and serum of dual genotype infected liver cirrhosis and hepatocellular carcinoma patients. *PloS One.* 2014;9(7):e102573. DOI 10.1371/journal.pone.0102573.
- Fu Y., Foden J.A., Khayter C., Maeder M.L., Reyon D., Joung J.K., Sander J.D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013;31(9):822-826. DOI 10.1038/nbt.2623.
- Golding S.E., Rosenberg E., Khalil A., McEwen A., Holmes M., Neill S., Povirk L.F., Valerie K. Double strand break repair by homologous recombination is regulated by cell cycle-independent signaling via ATM in human glioma cells. *J. Biol. Chemistry.* 2004;279(15):15402-15410. DOI 10.1074/jbc.M314191200.
- Hogan B., Beddington R., Costantini F., Lacy E. *Manipulating the Mouse Embryo. A Laboratory Manual.* 2nd ed. Cold Spring Harbor Laboratory Press, 1994.
- Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337(6096):816-821. DOI 10.1126/science.1225829.
- Kistler K., Vossell L., Matthews B. Genome engineering with CRISPR-Cas9 in the mosquito *Aedes aegypti*. *Cell Report.* 2015;11(1):51-60. DOI 10.1016/j.celrep.2015.03.009.
- Kouranova E., Forbes K., Zhao G., Warren J., Bartels A., Wu Y. CRISPRs for optimal targeting: Delivery of CRISPR components as DNA, RNA, and protein into cultured cells and single-cell embryos. *Hum. Gene Ther.* 2016;27(6):464-475. DOI 10.1089/hum.2016.009.
- Kulkarni A.S., Fortunato E.A. Stimulation of homology-directed repair at I-SceI-induced DNA breaks during the permissive life cycle of human cytomegalovirus. *J. Virology.* 2011;85(12):6049-6054. DOI 10.1128/JVI.02514-10.
- Li K., Wang G., Andersen T., Zhou P., Pu W.T. Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS One.* 2014;9(8):e105779. DOI 10.1371/journal.pone.0105779.
- Liu Y.G., Chen Y. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *Biotechniques.* 2007;43(5):649-650,652,654.
- Pierce A.J., Hu P., Han M., Ellis N., Jasin M. Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mam-

- malian cells. *Genes & Development*. 2001;15(24):3237-3242. DOI 10.1101/gad.946401.
- Pierce A.J., Johnson R.D., Thompson L.H., Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes & Development*. 1999;13(20):2633-2638.
- Sakurai T., Kamiyoshi A., Kawate H., Mori C., Watanabe S., Tanaka M. A non-inheritable maternal Cas9-based multiple-gene editing system in mice. *Sci. Reports*. 2016;6:20011. DOI 10.1038/srep20011.
- Sambrook J.F., Russell D.W. *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, 2001.
- Shchelkunov S.N. *Geneticheskaya inzheneriya* [Genetic Engineering]. Novosibirsk, 2004. (in Russian)
- Smirnov A.V., Yunusova A.M., Lukyanchikova V.A., Battulin N.R. CRISPR/Cas9: a universal tool for genomic engineering. *Vavilovskii Zhurnal Genetiki i Seleksii* = Vavilov Journal of Genetics and Breeding. 2016;20(4):493-510. DOI 10.18699/VJ16.175. (in Russian)
- Smith A.M., Takeuchi R., Pellenz S., Davis L., Maizels N., Monnat R.J., Stoddard B.L. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-Anil LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci.* 2009;106(13):5099-5104. DOI 10.1073/pnas.0810588106.
- Windbichler N., Papathanos P.A., Catteruccia F., Ranson H., Burt A., Crisanti A. Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucl. Acids Res.* 2007;35(17):5922-5933. DOI 10.1093/nar/gkm632.
- Yang H., Wang H., Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat. Protocols*. 2014;9:1956-1968.



Зачем нужны коллекции линий клеток

А.Г. Мензоров , О.Л. Серов

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Прогресс в области клеточной биологии в значительной мере связан с разработкой технологии репрограммирования геномов соматических клеток. В результате из соматических клеток человека, например клеток кожи, можно получить индуцированные плюрипотентные стволовые клетки. Индуцированные плюрипотентные стволовые клетки так же, как и эмбриональные стволовые клетки, обладают плюрипотентностью. Создание индуцированных плюрипотентных стволовых клеток открыло перспективу развития пациент-специфической клеточной терапии. Многие исследователи полагают, что плюрипотентные клетки являются реальной основой будущей регенеративной медицины. Получение плюрипотентных стволовых клеток пациента, их направленная дифференцировка в соматические клетки и трансплантация пациенту позволят избежать иммунологического отторжения. Кроме того, недавно появившаяся технология направленного изменения геномов CRISPR/Cas делает возможным исправление генетических дефектов в индуцированных плюрипотентных стволовых клетках, т. е. генетический дефект можно исправить в клетках пациента *in vitro* и, проведя дифференцировку в желаемый клеточный тип, вновь ввести эти клетки пациенту. Система CRISPR/Cas позволяет направленно вносить практически любые мутации в индуцированные плюрипотентные стволовые клетки для создания модельных линий, позволяющих изучать механизмы заболеваний и тестировать фармацевтические препараты. Можно как выключать один ген или группы генов, так и эффективно вносить генетические конструкции в выбранное место генома. Система направленного изменения генома также позволяет временно включать и выключать гены, удалять участки хромосом. Для эффективного использования индуцированных плюрипотентных стволовых клеток в биомедицинских исследованиях необходимо иметь общедоступные коллекции линий клеток. В настоящее время в Российской Федерации нет коллекций плюрипотентных стволовых клеток. Кроме того, необходима разработка стандартов культивирования и хранения клеток этого типа. В мини-обзоре представлено обоснование необходимости создания коллекций плюрипотентных линий клеток в Российской Федерации, подробное описание характеристик линий клеток, которые вносятся в коллекцию культур клеток, и рекомендуемый оптимальный протокол депонирования и использования линий клеток.

Ключевые слова: эмбриональные стволовые клетки; плюрипотентность; индуцированные плюрипотентные стволовые клетки; коллекция линий клеток; клеточная терапия.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Мензоров А.Г., Серов О.Л. Зачем нужны коллекции линий клеток. Вавиловский журнал генетики и селекции. 2016;20(6):945-948. DOI 10.18699/VJ16.215

HOW TO CITE THIS ARTICLE:

Menzorov A.G., Serov O.L. The necessity of cell banks. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6): 945-948. DOI 10.18699/VJ16.215

REVIEW

Received 27.10.2016

Accepted for publication 08.12.2016

© AUTHORS, 2016

The necessity of cell banks

A.G. Menzorov , O.L. Serov

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

Current progress in cell biology is connected with the development of somatic cell reprogramming technology. As a result of this technology, it is possible to produce induced pluripotent stem cells (iPSCs) from human somatic cells, for instance, from skin cells. As well as embryonic stem cells, these iPSCs possess pluripotency. Production of iPSCs opened new horizons for patient-specific cell therapy. Many researchers consider iPSCs a real basis for future regenerative medicine. Production of a patient's iPSCs, their differentiation into somatic cells, and subsequent transplantation to a patient would allow them to avoid immunological rejection. In addition, a recently developed technology of directed genome modification, CRISPR/Cas, allows correction of genetic mutations in iPSCs. Thus, genetic mutations could be corrected *in vitro*, and after differentiation into a desired cell type, these cells could be transplanted to a patient. In addition, CRISPR/Cas could be used to introduce practically any mutations into iPSCs for the creation of disease-specific model cell lines that would facilitate disease mechanism studies and pharmaceutical drug testing. It is possible to turn off any gene or genes as well as to insert a genetic construct into a selected genomic region to temporarily turn on and off genes and remove chromosomal regions. Cell banks that are open to general use are necessary for efficient usage of iPSCs in biomedical research. Currently, there are no pluripotent stem cell lines in Russian Federation cell banks. Moreover, it is essential to develop standardized practice of culture and storage of that cell type. This mini-review focuses on the necessity of the creation of a pluripotent stem cell bank in the Russian Federation, a detailed description, and a recommended protocol for cell line deposition and usage.

Key words: embryonic stem cells; pluripotency; induced pluripotent stem cells; cell bank; cell therapy.

Прошло 10 лет с появления статьи К. Takahashi и S. Yamanaka (2006), в которой авторы предложили эффективный способ получения индуцированных плюрипотентных стволовых клеток (ИПСК) из фибробластов с помощью экзогенной экспрессии четырех транскрипционных факторов, названных «коктейль Яманаки». ИПСК так же, как и эмбриональные стволовые клетки (ЭСК), плюрипотентны. ЭСК получают из клеток внутренней клеточной массы преимплантационных эмбрионов, а затем их можно дифференцировать во все типы соматических и зародышевых клеток взрослого организма. Важным параметром технологии, предложенной S. Yamanaka, является возможность создания персонализированных ИПСК человека, тем самым снимается проблема иммунологической совместимости. Многие исследователи считают, что будущее регенеративной медицины будет основано на широком использовании ИПСК в качестве источника для дифференцировки в различные типы соматических клеток с их дальнейшей трансплантацией пациентам. Таким образом, в результате создания технологий получения ЭСК и ИПСК в клеточной биологии появилась новая категория культур клеток, отличная от прежних, традиционных, преимущественно перевиваемых «бессмертных» культур клеток или кратковременно культивируемых культур первичных соматических клеток.

Есть еще три важных аспекта, связанных с совершенствованием технологий получения персонализированных ИПСК. Во-первых, пролиферативный потенциал ИПСК человека дает реальную возможность получения ИПСК от пациентов-носителей мутантных генов или хромосомных перестроек. Такая коллекция ИПСК может служить материальной базой для изучения молекулярных механизмов эффектов мутаций в развитии патологий. Во-вторых, разработка системы направленного трансгеноза CRISPR/Cas, позволяющей осуществлять адресную модификацию отдельных генов или даже фрагментов хромосом, дает экспериментаторам инструмент для создания мутантных линий ИПСК, которые можно использовать для моделирования тех или иных наследственных патологий (Смирнов и др., 2016). Более того, система CRISPR/Cas позволяет осуществлять коррекцию мутаций в ИПСК от пациентов-носителей мутаций, тем самым получать плюрипотентные клетки с «нормальными» аллелями, что может стать первым шагом для их дальнейшего использования в клеточной терапии наследственных заболеваний. Значение таких линий ИПСК трудно переоценить, и актуальность создания коллекций таких линий очевидна. В-третьих, создание коллекции мутантных клеток обеспечивает базу для тестирования фармакологических препаратов для коррекции эффектов мутаций.

Самостоятельным компонентом современных коллекций линий клеток являются культуры специализированных соматических клеток, например, таких как кератиноциты или клетки волосных фолликулов, и нейральные клетки-предшественники, например нейроэпителиальные или нейральные стволовые клетки. Эти клетки также являются новой категорией в клеточной биологии со специфическими условиями культивирования и криоконсервации. Учитывая то, что эти типы клеток наиболее

близки к применению в клинической практике, актуальность создания коллекций таких линий также очевидна.

В связи с этим следует отметить своевременность недавнего появления Федерального закона от 23.06.2016 г. № ФЗ-180 «О биомедицинских клеточных продуктах», регламентирующего обращение биомедицинских клеточных продуктов в Российской Федерации. Одновременно повышаются и требования к исследовательской деятельности в разных областях клеточной биологии и эмбриологии, которые должны соответствовать международным стандартам «Global Standards for Stem Cell Research and Clinical Translation: The 2016 ISSCR Guidelines» (<http://www.isscr.org/docs/default-source/guidelines/isscr-guidelines-for-stem-cell-research-and-clinical-translation.pdf>).

При создании коллекций плюрипотентных клеток человека и животных так же, как и специализированных зрелых и дифференцированных клеток, актуальна разработка новых регламентов и стандартов их культивирования и криоконсервации. Унификация требований к депонируемым клеточным культурам – необходимый и обязательный элемент клеточной коллекции. Ниже приведены общепринятые международным сообществом критерии для характеристики клеток, депонируемых в коллекциях.

Характеристики депонируемых линий клеток

Данные о коллекции клеток. Основная информация о коллекции клеток должна включать название коллекции и контактные сведения.

Название культуры. Уникальное название культуры клеток, позволяющее найти информацию об этой культуре в других источниках данных.

Вид. Видовая принадлежность культуры – одна из базовых характеристик. Наиболее распространенные линии клеток млекопитающих принадлежат к двум видам – человек и мышь – как наиболее широко используемым экспериментальным объектам.

Каталожный номер. Каждая культура помимо уникального названия должна иметь каталожный номер в данной коллекции. Может возникнуть ситуация, когда одна и та же линия клеток получена для коллекции из разных источников, и, соответственно, названия для идентификации недостаточно. Например, перевивная линия клеток НЕК293 очень широко используется в клеточной биологии. Эти клетки получены из клеток эмбриональной почки человека, трансформированных аденовирусом, модальное число хромосом – 64, в отличие от нормального для человека – 46, кроме того, в линии присутствуют хромосомные перестройки и в целом для этих клеток характерна нестабильность кариотипа (<https://www.lgstandards-atcc.org/Products/All/CRL-1573.aspx>). Популярность этой линии объясняется двумя причинами: простота культивирования и очень высокая эффективность трансфекции ДНК. Это позволяет эффективно использовать клетки НЕК293 для трансфицирования ДНК и продукции вирусов. Так, наиболее распространенный протокол получения ИПСК – трансдукция лентивирусами, произведенными клетками НЕК293. Нестабильность кариотипа этой линии при культивировании и дополнительные модификации (линия 293T содержит Т-антиген, Phoenix, производная этой

линии, — компоненты геномов вирусов и т. д.) приводят к необходимости учитывать не только название, но и происхождение культуры.

Происхождение клеток — важная характеристика. Например, ЭСК мыши получают из эмбрионов мышей на стадии 3–3.5 дней, чаще всего от мышей линии 129. ЭСК плюрипотентны, т. е. способны к неограниченному делению *in vitro* и к дифференцировке в клетки, производные трех зародышевых листков (Evans, Kaufman, 1981). ИПСК чаще всего получают из фибробластов (Takahashi, Yamanaka, 2006; Takahashi et al., 2007). Мезенхимальные стволовые клетки, как правило, выделяют из костного мозга, жировой ткани и плаценты.

Морфология — одна из ключевых характеристик линии клеток. Важными являются размер клеток, ядерно-цитоплазматическое соотношение, формирование при росте монослоя или колоний определенной формы. Эти сведения позволяют при получении культуры из коллекции и культивировании следить за состоянием клеток и при необходимости менять условия культивирования.

Способ культивирования. В целом способы культивирования клеток можно разделить на монослойный и суспензионный. При монослойном способе клетки растут прикрепленными к пластику, который можно покрыть различными матриксами для лучшей адгезии, например желатином, матригелем или витронектином. В некоторых случаях матрикс необходим не только для культивирования клеток, но и для направления дифференцировки, например поли-L-орнитин и ламинин часто используют для нейральной дифференцировки плюрипотентных клеток. В экспериментах с плюрипотентными клетками мыши и человека в качестве питающих клеток используют эмбриональные фибробласты мыши, обработанные митомицином С. Такие клетки не делятся, но сохраняют способность синтезировать ростовые факторы, необходимые для плюрипотентных клеток. Суспензионный способ культивирования, при котором клетки находятся в среде без прикрепления к поверхности, применяется для клеток, которые не растут в монослой. Например, на одной из стадий дифференцировки ИПСК в нейроны можно получить нейросферы — конгломераты незрелых предшественников нейронов. Кроме того, при необходимости получения большой биомассы некоторых типов клеток можно применять биореакторы для наращивания клеток в суспензии.

Условия культивирования. Это, возможно, самый важный пункт описания культуры. Условия культивирования обычно включают в себя среду для культивирования (например, DMEM), сыворотку крови (например, эмбрионов крупного рогатого скота) и дополнительные добавки, такие как антибиотики, аминокислоты и т. д. Для культивирования практически любого типа клеток существует множество протоколов и их модификаций. Важно, чтобы описание культуры в коллекции клеток включало в себя апробированные условия культивирования, гарантирующие эффективное культивирование клеток, такие как температура, концентрация CO₂ и процедуры пересева и криоконсервации.

Контроль контаминации. Культивирование клеток проводится в условиях, близких к стерильным. К сожа-

лению, иногда культуры клеток контаминируются бактериями или грибами. Для предотвращения роста бактерий в среду для культивирования добавляют антибиотики, обычно это смесь пенициллина и стрептомицина, хотя риск контаминации сохраняется даже при соблюдении всех правил безопасности. Кроме того, некоторые лаборатории даже рекомендуют постоянно вести культуры без антибиотиков, так как в некоторых случаях бактерии могут выживать в присутствии антибиотиков и постепенно приобретать устойчивость к ним. Среди особо «злостных» патогенов следует выделить микоплазму — эндосимбиотическую бактерию, не имеющую клеточной стенки. Контаминация клеточных культур микоплазмой осложняется тем, что изменяется физиологическое состояние клеток и нередко возникают перестройки хромосом. По мнению H.G. Drexler и C.C. Uphoff (2002), исследования, проведенные на клетках, зараженных микоплазмой, не могут быть полноценными, поскольку искажаются физиологические параметры клеток. Для обнаружения контаминации клеток оптимально использовать чувствительные молекулярно-биологические методы, например ПЦР. Культуры клеток человека необходимо дополнительно проверять на зараженность вирусами: вирус гепатита В, папилломавирусы человека, ВИЧ, цитомегаловирус, вирус Эпштейна — Барр.

Контроль видовой идентичности. При долговременном культивировании возрастает риск контаминации не только микоплазмой и другими микроорганизмами, но и клетками других линий. Один из самых известных примеров — массовая контаминация культур клетками HeLa, перевивными раковыми клетками (Lucey et al., 2009). Таким образом, необходимо проводить контроль видовой идентичности. В этом случае основной метод анализа — цитогенетический в сочетании с анализом специфических ДНК-маркеров.

Кариотип. Одна из важнейших характеристик клеточной линии — ее кариотип. В большинстве случаев ЭСК, ИПСК и нейральные стволовые клетки имеют нормальный диплоидный кариотип. Однако при культивировании клеток, особенно при длительном, в них могут возникать хромосомные перестройки и/или изменение численности хромосом. Это делает необходимым постоянный мониторинг кариотипа. Показано, например, что ЭСК мыши даже на ранних пассажах культивирования могут иметь нестабильный состав хромосом (Menzorov et al., 2016).

Пассаж криоконсервации. Продолжительное культивирование может вызывать нарушение хромосомного состава и увеличивает риск контаминации. В связи с этим для каждой культуры необходимо указывать пассаж, на котором клетки были криоконсервированы.

Дополнительные характеристики ЭСК и ИПСК. Известно, что разные клоны ЭСК или ИПСК, полученные из одного источника, могут иметь индивидуальные особенности. Для ЭСК и ИПСК наиболее важной характеристикой является их плюрипотентность. Оценка плюрипотентности включает несколько тестов: присутствие типичных маркеров, характерных для клеток внутренней клеточной массы бластоцисты, таких как Oct4, Sox2 и Nanog; способность образовывать эмбриоидные тельца, содержащие маркеры трех зародышевых листков; способность

формировать тератомы, содержащие производные трех зародышевых листков, при введении иммунодефицитным мышам (самый информативный тест для ЭСК и ИПСК человека); способность участвовать в формировании химерных мышей со вкладом в соматические и половые клетки; и, наконец, получение методом тетраплоидной комплементации мышей, полностью состоящих из клеток, потомков ЭСК или ИПСК. Данные транскриптомного анализа ЭСК и ИПСК также информативны для характеристики этих клеток.

Публикации. Дополнительная характеристика – публикации, описывающие депонируемые клетки. Оригинальные статьи позволяют получить дополнительные сведения о культуре клеток, которые по формату не могут быть указаны в базовом описании культуры клеток.

Коллекции линий клеток

Необходимость в централизованном доступе к линиям клеток привела к возникновению коллекций культур клеток. Мы рассмотрим несколько коллекций клеток с акцентом на наличие в них плюрипотентных стволовых клеток.

АТСС – один из самых больших в мире банков биологических образцов (<https://www.lgcstandards-atcc.org>). В нем содержится более 3400 линий клеток, в том числе 19 линий ИПСК человека и 14 линий ЭСК мыши.

Центр биоресурсов RIKEN Японии (<http://en.brc.riken.jp/>) содержит 11 линий ИПСК человека от здоровых доноров, а также более 400 линий от пациентов с различными наследственными заболеваниями. В этом банке представлено более 100 линий ЭСК мыши.

Специализированный европейский банк ИПСК на сегодняшний день имеет более 250 линий ИПСК, полученных из клеток здоровых людей и пациентов с наследственными заболеваниями (<https://cells.ebisc.org/>).

Российская коллекция клеточных культур содержит сотни линий позвоночных, представленных главным образом перевиваемыми или первичными соматическими клетками (www.cytspb.rssi.ru/rkkk/rkkk_ru.htm).

Рекомендуемый протокол депонирования линий клеток

Так как продолжительное культивирование клеток увеличивает риск изменения их хромосомного состава, желательно минимизировать число пассажей перед криоконсервацией. Далее описан рекомендуемый протокол депонирования линий клеток.

Этап 1 – основной сток. Следует размножить полученную линию клеток и заморозить три-пять ампул основного стока культуры клеток. На этом этапе необходимо проверить культуру на контаминацию бактериями, грибами и микоплазмой, а также подтвердить видовую принадлежность с помощью анализа кариотипа и специфических маркеров ДНК.

Этап 2 – рабочий сток. Размораживается одна ампула основного стока, клетки размножаются и криоконсервируются 10–100 ампул. В конце этого этапа также необходимо провести проверку культуры клеток.

Этап 3 – пользовательский сток. Размораживается одна ампула рабочего стока и снова после размножения

замораживается 10–100 ампул с оценкой культуры. Эти ампулы предоставляют заказчикам культур клеток.

Такой протокол депонирования дает возможность получить большое количество материала подтвержденного качества. При этом число пассажей культуры до получения пользовательского стока минимально, например для ЭСК мыши оно может составлять всего семь. Это позволяет минимизировать риск контаминации и сохранить свойства линий клеток.

Заключение

Прогресс в развитии новых клеточных технологий открыл широкие возможности как фундаментальных, так и прикладных исследований в биомедицине. Потребность в организации инфраструктуры для этих исследований очевидна и, в первую очередь, это относится к необходимости создания коллекций линий таких типов клеток, как плюрипотентные клетки, ЭСК и ИПСК и клетки-предшественники, детерминированные к дифференцировке в клетки определенной ткани. Такие коллекции существуют в Европе, США, Японии и других странах. В Российской Федерации также назрела необходимость в создании подобных коллекций для обеспечения как академических научных работ, так и активно развивающейся биомедицины, в которой предполагается внедрение достижений клеточной биологии в клиническую практику.

Acknowledgments

This work was supported by State Budgeted Project 0324-2016-0005 «Collection of pluripotent cultures of human and mammalian cells for general biological and biomedical purposes».

Conflict of interest

The authors declare no conflict of interest.

References

- Drexler H.G., Uphoff C.C. Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention. *Cytotechnology*. 2002;39(2):75-90. DOI 10.1023/A:1022913015916.
- Evans M.J., Kaufman M.H. Establishment in culture of pluripotent cells from mouse embryos. *Nature*. 1981;292(5819):154-156.
- Lucey B.P., Nelson-Rees W.A., Hutchins G.M. Henrietta Lacks, HeLa cells, and cell culture contamination. *Arch. Pathol. Lab. Med.* 2009; 133(9):1463-1467. DOI 10.1043/1543-2165.133.9.1463.
- Menzorov A., Pristiyazhnyuk I., Kizilova H., Yunusova A., Battulin N., Zhelezova A., Golubitsa A., Serov O. Cytogenetic analysis and Dlk1-Dio3 locus epigenetic status of mouse embryonic stem cells during early passages. *Cytotechnology*. 2016;68(1):61-71. DOI 10.1007/s10616-014-9751-y.
- Smirnov A.V., Yunusova A.M., Lukyanchikova V.A., Battulin N.R. CRISPR/Cas9, A universal tool for genomic engineering. *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov Journal of Genetics and Breeding*. 2016;20(4):493-510. DOI 10.18699/VJ16.175 (in Russian).
- Takahashi K., Tanabe K., Ohnuki M., Narita M., Ichisaka T., Tomoda K., Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 2007;131(5):861-872. DOI 10.1016/j.cell.2007.11.019.
- Takahashi K., Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126(4):663-676. DOI 10.1016/j.cell.2006.07.024.



Сравнительное изучение репликативных свойств противоопухолевых рекомбинантных вирусов осповакцины на культурах клеток глиобластомы человека U87 и почки мартышки CV-1

Р.А. Максютлов¹, И.В. Колосова¹, Т.В. Трегубчак¹, И.А. Разумов², С.Н. Щелкунов^{1, 2}

¹ Федеральное бюджетное учреждение науки Государственный научный центр вирусологии и биотехнологии «Вектор», Кольцово, Новосибирская область, Россия

² Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

В современном мире одно из активно развивающихся направлений в лечении онкологических заболеваний – виротерапия, особенностью которой является избирательное уничтожение раковых клеток при минимальном воздействии на здоровые клетки организма. Перспективной основой для создания онколитических препаратов могут стать ортопоксвирусы, имеющие ряд преимуществ перед другими вирусными векторами, одно из которых – большая емкость генома, что позволяет встраивать в их геном гены, кодирующие белки с различными противоопухолевыми свойствами. В данной работе проведено сравнительное изучение репликативных свойств десяти различных вариантов штамма ЛИВП вируса осповакцины (ВОВ), в которых встроены дополнительные гены гранулоцитарно-макрофагального колоние-стимулирующего фактора, апоптоз-индуцирующего белка TRAIL и гена зеленого флуоресцирующего белка, на перевиваемой культуре клеток глиобластомы человека U87 и почки африканской зеленой мартышки CV-1. Кроме того, изучен вирус с пятью удаленными генами вирулентности (гемагглютинаина, γ-интерферон-связывающего белка, тимидинкиназы, комплемент-связывающего белка и Bcl-2-подобного ингибитора апоптоза), характеризующийся значительно меньшей реактогенностью и нейровирулентностью по сравнению с исходным штаммом ЛИВП ВОВ. Полученные данные свидетельствуют о том, что в культуре клеток глиобластомы активнее всего реплицируются различные варианты ВОВ с нарушенным геном тимидинкиназы.

Ключевые слова: культура клеток глиобластомы; виротерапия; онколитический вирус.

A comparative study of replicative properties of antitumor recombinant vaccinia viruses on human glioblastoma cell culture U87 and monkey kidney cell culture CV-1

R.A. Maksyutov¹, I.V. Kolosova¹, T.V. Tregubchak¹, I.A. Razumov², S.N. Shchelkunov^{1, 2}

¹ State Research Center of Virology and Biotechnology "Vector", Koltsovo, Novosibirsk region, Russia

² Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

In the world of today, virotherapy is one of the rapidly developing areas in the treatment of cancer, and its advantage is selective destruction of cancer cells with minimizing the destructive effect on normal cells of the body. A promising basis for the creation of oncolytic drugs is orthopoxviruses, which have a number of advantages over other viral vectors, and one of these advantages is a large capacity of the genome, which allows genes encoding proteins with antitumor properties to be cloned into their genome. In this study, we compared the replicative properties of ten variants of vaccinia virus (the strain LIVP of VACV) using human glioblastoma cell culture; some of these viruses have additional genes, such as the gene encoding granulocyte-macrophage colony stimulating factor, gene encoding apoptosis-inducing protein TRAIL and gene encoding green fluorescent protein. Furthermore, the virus with five virulence genes deleted (genes encoding hemagglutinin, γ-interferon-binding protein, thymidine kinase, complement-binding protein and Bcl2-like inhibitor of apoptosis), which has significantly lower reactogenicity and neurovirulence compared to the original strain LIVP of VACV, was studied. These data suggest that variants of vaccinia virus with a defective gene encoding thymidine kinase most actively replicate in glioblastoma cell culture.

Key words: glioblastoma cell culture; virotherapy; oncolytic virus.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Максютлов Р.А., Колосова И.В., Трегубчак Т.В., Разумов И.А., Щелкунов С.Н. Сравнительное изучение репликативных свойств противоопухолевых рекомбинантных вирусов осповакцины на культурах клеток глиобластомы человека U87 и почки мартышки CV-1. Вавиловский журнал генетики и селекции. 2016;20(6):949-953. DOI 10.18699/VJ16.207

HOW TO CITE THIS ARTICLE:

Maksyutov R.A., Kolosova I.V., Tregubchak T.V., Razumov I.A., Shchelkunov S.N. A comparative study of replicative properties of antitumor recombinant vaccinia viruses on human glioblastoma cell culture U87 and monkey kidney cell culture CV-1. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):949-953. DOI 10.18699/VJ16.207

ORIGINAL ARTICLE

Received 26.04.2016

Accepted for publication 14.06.2016

© AUTHORS, 2016



e-mail: snshchel@vector.nsc.ru

Глиомы образуют гетерогенную группу опухолей центральной нервной системы (Louis et al., 2007; Perry, Wesseling, 2016). Одна из наиболее агрессивных форм глиом – глиобластома. За последние 30 лет появление новых методов лечения и терапевтических препаратов позволило увеличивать прогнозируемую продолжительность жизни людей с данным диагнозом всего лишь на несколько месяцев (Hulou et al., 2016), в связи с чем проблема разработки новых препаратов, модификации уже имеющихся веществ, усовершенствования схем лечения остается актуальной.

Основным методом лечения онкологических заболеваний является хирургический, который также применяется для резекции опухолей головного мозга. Однако в случае центральной нервной системы оперативное вмешательство сопряжено с высокой степенью риска возникновения послеоперационных осложнений. Такие осложнения могут нести как региональный характер (травмы функционально значимых зон мозга, послеоперационный отек ткани мозга и геморрагические осложнения), так и системный (острые язвы желудочно-кишечного тракта, тромбозы, тромбоэмболии) (Jackson et al., 2016). Только у 26.5 % пациентов, перенесших резекцию глиомы совместно с другими методами лечения, прогнозируемая продолжительность жизни составляет два года (Stupp et al., 2005).

К другим методам лечения, прежде всего, относится лучевая терапия, которая характеризуется своими преимуществами и недостатками. Зачастую в клинической практике глиомы оказываются наименее чувствительными к излучению даже при высоких дозировках, при этом наблюдается значительный токсический эффект на нетрансформированные клетки (Snider, Mehta, 2016). Наиболее результативно комбинирование радиотерапии с химиотерапией. Однако большинство химиотерапевтических препаратов невозможно применять при лечении опухолей головного мозга, что вызвано, в первую очередь, непроницаемостью гематоэнцефалического барьера для многих химических веществ. Существуют различные стратегии по разработке препаратов, способных преодолевать гематоэнцефалический барьер, – это и химическая модификация существующих препаратов, и создание наноносителей, которые способны доставлять вещества к клеткам головного мозга (Karim et al., 2016). Таким образом, только комплексное лечение, включающее хирургическое вмешательство, химио- и лучевую терапию, позволяет увеличить продолжительность и улучшить качество жизни больных.

В последние годы появилось еще одно активно развивающееся направление в терапии онкологических заболеваний – биотерапия, которая может быть успешно применена в комплексе с уже существующими методами лечения. Принцип такой терапии заключается, прежде всего, в активации защитных систем организма при помощи противоопухолевых вакцин, моноклональных антител, клеточных факторов и т. д. Такой подход в лечении позволяет вовлечь в противоопухолевую защиту иммунную систему, а также воздействовать на факторы и механизмы, контролирующие процессы ангиогенеза и апоптоза (Ding, Wei, 2012).

К одному из перспективных направлений биотерапии злокачественных опухолей относится виротерапия, которая позволяет избирательно уничтожать раковые клетки при минимальном воздействии на здоровые клетки организма. Также вирусы можно использовать в качестве способа доставки в клетки факторов, необходимых для поддержания естественных антиканцерогенных свойств организма (Bienkowska-Szewczyk, Szewczyk, 1999; Weibel et al., 2011).

Данная работа посвящена оценке потенциальной возможности использования созданного ранее набора онколитических вирусов с иммуностимулирующими свойствами на основе вируса осповакцины (BOV) (Максютов и др., 2013; Петров и др., 2013; Якубицкий и др., 2015) для терапии глиом, а именно изучению репликативных свойств различных вариантов BOV на перевиваемой культуре клеток глиобластомы человека U87, для которой показана успешная ортотопическая ксенотрансплантация мышам линии SCID (Завьялов и др., 2015).

Материалы и методы

Культуры клеток. В работе использовали перевиваемую культуру клеток почки африканской зеленой мармозетки CV-1 из коллекции клеточных культур ГНЦ ВБ «Вектор» и перевиваемую культуру клеток глиобластомы человека U87 из криобанка ЦКП «SPF-виварий» ИЦиГ СО РАН.

Культивирование и ростовая среда. Монослойное культивирование линий клеток проводили при температуре 37 °C в присутствии 5 % CO₂ с использованием сред ДМЕМ (для культуры клеток CV-1) и ДМЕМ/F12 (для культуры клеток U87) производства «Биолот» (Россия), содержащих 10 % эмбриональной сыворотки коров («HyClone», США).

Вирусы. Изучали полученный из двух независимых клоновых вариантов BOV ЛИВП (клоны 5 и 14) набор из семи вирусов с направленно введенными инактивирующими делециями по генам *J2R* (кодирует тимидинкиназу) и *C11R* (вирусный фактор роста) и встройками генов, кодирующих фенотипический маркер (GFP), иммуностимулирующую (GMCSF) и противораковую (TRAIL) терапевтические молекулы. Также был использован аттенуированный вариант BOV 1421ABJCN с пятью направленно введенными делециями по генам *A56R* (кодирует гемагглютинин), *B8R* (гамма-интерферон-связывающий белок), *C3L* (комплемент-связывающий белок), *J2R* (тимидинкиназу) и *NIL* (Bcl2-подобный ингибитор апоптоза) (табл. 1).

Определение динамики роста вариантов BOV *in vitro*. Для изучения динамики роста вариантов BOV ЛИВП в системах *in vitro* 90 % монослой клеток линий CV-1 или U87, полученный на 12-луночных планшетах, инфицировали со множественностью заражения 0.1 БОЕ/клетку каждым из исследуемых вирусов в трех повторах с инкубацией 0, 24, 48 или 72 ч при 37 °C в условиях 5 % CO₂. После трехкратного замораживания–оттаивания содержащие вирус препараты титровали методом бляшек в монослой клеток CV-1.

Результаты

При разработке потенциального онколитического препарата на основе вируса важными аспектами являются

Table 1. VACV variants tested

Nº	Recombinant VACV	Disturbed VACV gene	Inserted gene	References
1	VACV cl. 5	–	–	Maksyutov et al., 2013
2	VACVΔJ2R	<i>J2R</i>	–	
3	VACVΔJ2R/GMCSF(+)	<i>J2R</i>	<i>GMCSF</i>	
4	VACVΔJ2RΔC11R	<i>J2R, C11R</i>	–	
5	VACVΔJ2RΔC11R/GMCSF(+)	<i>J2R, C11R</i>	<i>GMCSF</i>	Yakubitskiy et al., 2015
6	VACVΔJ2RΔC11R/GMCSF(+)/GFP(+)	<i>J2R, C11R</i>	<i>GMCSF, GFP</i>	
7	VACVΔJ2RΔC11R/GMCSF(+)/TRAIL(+)	<i>J2R, C11R</i>	<i>GMCSF, TRAIL</i>	
8	VACV cl. 14	–	–	
9	1421ABJCN	<i>A56R, B8R, J2R, C3L, N1L</i>	–	Petrov et al., 2013
10	LIVP-GFP	<i>J2R</i>	<i>GFP</i>	

Viruses 2–7 are derived from VACV clonal variant 5 (No. 1); Viruses 9, 10 are derived from VACV clonal variant 14 (No. 8).

Table 2. Growth dynamics of the studied VACV variants on the cell culture CV-1

Virus	Virus titer*, log ₁₀ PFU/ml in a certain time after infection (h)			
	0	24	48	72
VACV cl. 5	3.81 ± 0.02	5.64 ± 0.01	7.13 ± 0.00	6.86 ± 0.02
VACVΔJ2R	3.95 ± 0.02	5.44 ± 0.07	7.05 ± 0.01	6.82 ± 0.01
VACVΔJ2R/GMCSF(+)	3.96 ± 0.02	4.63 ± 0.01	7.04 ± 0.00	6.58 ± 0.03
VACVΔJ2RΔC11R	3.89 ± 0.02	4.99 ± 0.01	6.77 ± 0.00	6.54 ± 0.04
VACVΔJ2RΔC11R/GMCSF(+)	4.00 ± 0.03	4.59 ± 0.06	6.65 ± 0.01	6.18 ± 0.08
VACVΔJ2RΔC11R/GMCSF(+)/GFP(+)	3.77 ± 0.01	4.58 ± 0.02	6.63 ± 0.03	6.40 ± 0.05
VACVΔJ2RΔC11R/GMCSF(+)/TRAIL(+)	3.80 ± 0.07	4.51 ± 0.03	6.75 ± 0.01	6.16 ± 0.04
VACV cl. 14	3.67 ± 0.03	5.56 ± 0.05	7.14 ± 0.02	6.78 ± 0.03
1421ABJCN	3.51 ± 0.03	5.19 ± 0.02	6.96 ± 0.02	6.66 ± 0.02
LIVP-GFP	3.71 ± 0.03	5.82 ± 0.02	6.98 ± 0.02	6.55 ± 0.02

* Virus titers are presented as the mean value (μ) ± standard deviation (σ).

селективность такого вируса в отношении онкогенно трансформированных клеток и эффективность размножения вируса в таких клетках. Поэтому нами было проведено сравнительное изучение репликативных свойств 10 вариантов BOB (см. табл. 1) на перевиваемых культурах клеток высокочувствительной для BOB линии CV-1 и широко используемой в исследованиях глиобластомы человека U87 с целью определения комбинации модификаций в геноме BOB, обеспечивающей наибольшую селективность размножения BOB в клетках глиом.

Проведенные эксперименты не выявили значимых различий в ростовых характеристиках исследуемых рекомбинантных вариантов BOB в сравнении с исходными клоновыми вариантами BOB в культуре клеток CV-1 (табл. 2).

В группе вирусов с происхождением из клона 5 BOB на культуре клеток U87 наблюдалась более высокая эффективность роста в сравнении с исходным вирусом для вариантов BOB с нарушением одного гена, *J2R*, или двух генов, *J2R* и *C11R*, без встройки чужеродных генов. При этом встройка гена *GMCSF*, независимо от комбинации с

другими нарушаемыми и встраиваемыми генами, значительно (до полутора порядков) снижала репродуктивную способность рекомбинантных вирусов на данной глиомной культуре клеток U87 (рис. 1).

В группе вирусов с происхождением из клона 14 BOB на культуре клеток U87 также наблюдалась большая эффективность роста в сравнении с исходным вирусом для обоих вариантов BOB с нарушением гена *J2R* (и одновременной встройкой гена *GFP*) и с нарушением, помимо гена *J2R*, дополнительно четырех вирусных генов: *A56R*, *B8R*, *C3L* и *N1L* (рис. 2).

Обсуждение

Вирусы являются естественными природными иммуномодуляторами, активно взаимодействующими как с иммунными, так и с другими, в частности опухолевыми, клетками и обладающими повышенным тропизмом к последним (Уразова, Кузнецова, 2013). Именно это природное свойство вирусов – тропность к опухолевым клеткам – позволяет рассматривать их в качестве терапевтических средств лечения онкологических заболеваний.

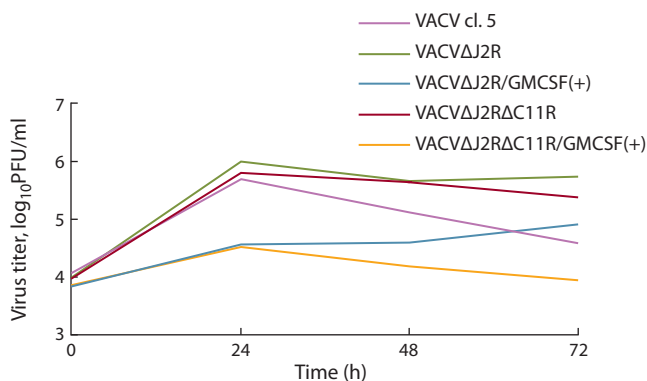


Fig. 1. Growth dynamics of VACV cl. 5 and its recombinants on the cell culture U87.

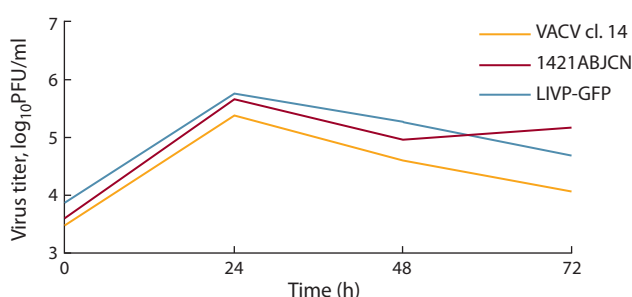


Fig. 2. Growth dynamics of VACV cl. 14 and its recombinants LIPV-GFP and 1421ABJCN on the cell culture U87.

В настоящее время известно немало препаратов на основе вирусов, которые успешно проходят клинические испытания. Для их создания используются вирусы, принадлежащие к широкому спектру вирусных семейств как ДНК-, так и РНК-содержащих. Одним из наиболее известных лекарственных препаратов вирусной природы является Rигvir, созданный на основе вируса семейства Picornaviridae. С 2004 г. Rигvir зарегистрирован в Латвии для лечения меланомы, локальной терапии кожных и подкожных метастазов меланомы, профилактики рецидивов и метастазов после операции (Donina et al., 2015). В 2015 г. для лечения меланомы в США был зарегистрирован герпесвирусный препарат Imlygic, его применяют в случае неоперабельных опухолей (Pol et al., 2015). Практически одновременно с Rигvir компания Shanghai Sunway Biotech зарегистрировала в Китае аденовирусный препарат Oncorine, который в сочетании с химиотерапией используется для лечения рака носоглотки (Liang, 2012).

В случае вирусной терапии опухолей головного мозга наиболее изучаемыми являются вирусы простого герпеса и аденовирусы. Так, вирус простого герпеса при попадании в клетку обуславливает синтез вирусной тимидинкиназы, что при совместной терапии с ганцикловиром приводит к нарушению клеточного цикла и, как следствие, — к клеточной гибели. Данный подход был описан еще в 1986 г. (Moolten, 1986) и остается актуальным. В недавних клинических испытаниях на пациентах с диагностированной глиобластомой изучали действие аденовирусного вектора,

несущего ген тимидинкиназы вируса простого герпеса. На третьей фазе клинических испытаний, проводимых в Европе (ASPECT), данный препарат не рекомендовал себя как увеличивающий общую выживаемость по сравнению с препаратами, используемыми при стандартных методах лечения (Westphal et al., 2013). Однако аналогичный вирус (аденовирус, содержащий ген тимидинкиназы вируса простого герпеса) при совместной химиолучевой терапии показал обнадеживающие результаты на первой и второй стадиях испытаний (Chiocci et al., 2011). Большая часть потенциальных онколитических терапевтических средств в настоящее время находится на первой стадии клинических испытаний.

Нами предлагается использовать БОВ как основу для создания онколитического терапевтического средства. Данный вирус имеет ряд преимуществ перед другими вирусами. Во-первых, репликация вируса происходит в цитоплазме клетки, что исключает мутагенез за счет интеграции вирусной ДНК в геном клетки; во-вторых, большая емкость двуцепочечного вирусного ДНК-генома позволяет осуществлять вставку в его состав многочисленных генов, кодирующих иммуностимулирующие и противораковые терапевтические молекулы, и адресно доставлять их в опухолевые клетки; в-третьих, долгая история клинического применения БОВ в борьбе с натуральной оспой гарантирует эффективность его применения для человека.

В используемых нами вариантах БОВ имеются различные комбинации модификаций вирусного генома. Варианты вируса с нарушенными генами тимидинкиназы и вирусного фактора роста позволяют снизить способность вируса к репликации в здоровых клетках организма (Максютов и др., 2013; Петров и др., 2013). Вставка в вирусный геном генов гранулоцитарно-макрофагального колониестимулирующего фактора и апоптоз-индуцирующего белка TRAIL позволяет представлять раковые антигены иммунной системе организма и запускать апоптоз. Полученный ранее вариант БОВ с пятью удаленными генами вирулентности (генами гемагглютинина, γ -интерферон-связывающего белка, тимидинкиназы, комплемент-связывающего белка и Bcl-2-подобного ингибитора апоптоза) характеризуется значительно меньшей реактогенностью и нейровирулентностью по сравнению с исходным штаммом ЛИВР БОВ (Якубский и др., 2015), что открывает перспективы для применения этого вируса как более безопасного вирусного вектора.

Ранее была отработана модель ортотопической ксенотрансплантации клеток глиобластомы человека U87 иммунодефицитным мышам линии SCID, которая может быть использована в качестве *in vivo* модели для проверки новых противоопухолевых препаратов и схем лечения онкологических заболеваний головного мозга человека (Завьялов и др., 2015).

Полученные нами данные по изучению динамики роста вирусов в перевиваемой культуре глиобластомы человека U87 свидетельствуют о том, что активнее всего реплицируются в культуре клеток глиом различные варианты БОВ с нарушенным геном тимидинкиназы (см. рис. 1, 2). Вставка гена GMCSF, независимо от комбинации с другими нарушаемыми и встраиваемыми генами, значитель-

но (до полутора порядков) снижает репродуктивную способность рекомбинантных вирусов на данной глиомной культуре клеток U87 (см. рис. 1).

Полученные результаты в эксперименте *in vitro* показывают, что, по крайней мере, некоторые исследованные кандидатные онколитические ВОВ могут быть после испытаний в системе *in vivo* использованы при создании новых препаратов для лечения глиом человека.

Acknowledgments

This study was supported by State Budgeted Project 0324-2015-0004; the Russian Foundation for Basic Research, project 15-04-01326a, and the Russian Science Foundation, project 16-15-10101.

Conflict of interest

The authors declare no conflict of interest.

References

- Bienskowska-Szewczyk K., Szewczyk B. Expression of genes coding for viral glycoproteins in heterologous systems. *Acta Biochim. Pol.* 1999;2:325-339.
- Chiocca E.A., Aguilar L.K., Bell S.D., Kaur B., Hardcastle J., Cavaliere R., McGregor J., Lo S., Ray-Chaudhuri A., Chakravarti A., Grecula J., Newton H., Harris K.S., Grossman R.G., Trask T.W., Baskin D.S., Monterroso C., Manzanera A.G., Aguilar-Cordova E., New P.Z. Phase IB study of gene-mediated cytotoxic immunotherapy adjuvant to up-front surgery and intensive timing radiation for malignant glioma. *J. Clin. Oncol.* 2011;29(27):3611-3619. DOI 10.1200/JCO.2011.35.5222.
- Ding Z.Y., Wei Y.Q. Cancer biotherapy: Progress in China. *Recent Advances Cancer Res. Ther.* 2012;1-31. DOI 10.1016/B978-0-12-397833-2.00001-7.
- Doniça S., Strêle I., Proboka G., Auziņš J., Alberts P., Jonsson B., Venskus D., Muceniece A. Adapted ECHO-7 virus Rignvir immunotherapy (oncolytic virotherapy) prolongs survival in melanoma patients after surgical excision of the tumour in a retrospective study. *Melanoma Res.* 2015;25(5):421-426. DOI 10.1097/CMR.0000000000000180.
- Hulou M.M., Cho C.F., Chiocca E.A., Bjerkvig R. Experimental therapies: gene therapies and oncolytic viruses. *Handb. Clin. Neurol.* 2016;134:183-197. DOI 10.1016/B978-0-12-802997-8.00011-6.
- Jackson C., Westphal M., Quiñones-Hinojosa A. Complications of glioma surgery. *Handb. Clin. Neurol.* 2016;134:201-218. DOI 10.1016/B978-0-12-802997-8.00012-8.
- Karim R., Palazzo C., Evrard B., Piel G. Nanocarriers for the treatment of glioblastoma multiforme: Current state-of-the-art. *J. Control. Release.* 2016;227:23-37. DOI 10.1016/j.jconrel.2016.02.026.
- Liang M. Clinical development of oncolytic viruses in China. *Curr. Pharm. Biotechnol.* 2012;13(9):1852-1857.
- Louis D.N., Ohgaki H., Wiestler O.D., Cavenee W.K., Burger P.C., Jouvet A., Scheithauer B.W., Kleihues P. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 2007;114(2):97-109. DOI 10.1007/s00401-007-0243-4.
- Maksyutov R.A., Tregubchak T.V., Denisova N.I., Maksyutov A.Z., Gavrilova E.V. Creating a modern platform comprising a set of oncolytic viruses with immune-stimulating properties. *Rossiyskiy immunologicheskii zhurnal = Russian Immunological Journal.* 2013;7(4):456-459. (in Russian)
- Moolten F.L. Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy. *Cancer Res.* 1986;46(10):5276-5281.
- Perry A., Wesseling P. Histologic classification of gliomas. *Handb. Clin. Neurol.* 2016;134:71-95. DOI 10.1016/B978-0-12-802997-8.00005-0.
- Petrov I.S., Goncharova E.P., Kolosova I.V., Pozdnyakov S.G., Schelkunov S.N., Zenkova M.A., Vlasov V.V. Antitumor effect of the LIVP-GFP recombinant vaccinia virus. *Doklady Akademii nauk = Reports of the Academy of Sciences.* 2013;451(5):592-597. (in Russian)
- Pol J., Kroemer G., Galluzzi L. First oncolytic virus approved for melanoma immunotherapy. *Oncoimmunol.* 2015;5(1):e1115641. DOI 10.1080/2162402X.2015.1115641.
- Snider J.W., Mehta M. Principles of radiation therapy. *Handb. Clin. Neurol.* 2016;134:131-147. DOI 10.1016/B978-0-12-802997-8.00008-6.
- Stupp R., Mason W.P., van den Bent M.J., Weller M., Fisher B., Taphoorn M.J., Belanger K., Brandes A.A., Marosi C., Bogdahn U., Curschmann J., Janzer R.C., Ludwin S.K., Gorlia T., Allgeier A., Lacombe D., Cairncross J.G., Eisenhauer E., Mirimanoff R. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* 2005;352:987-996.
- Urazova L.N., Kuznetsova T.I. Oncolytic viruses in oncology. *Sibirskiy onkologicheskii zhurnal = Siberian Journal of Oncology.* 2013;4:28-35. (in Russian)
- Weibel S., Raab V., Yu Y.A., Worschech A., Wang E., Marincola F.M., Szalay A.A. Viral-mediated oncolysis is the most critical factor in the late-phase of the tumor regression process upon vaccinia virus infection. *BMC Cancer.* 2011;11:68. DOI 10.1186/1471-2407-11-68.
- Westphal M., Ylä-Herttuala S., Martin J., Warnke P., Menei P., Eckland D., Kinley J., Kay R., Ram Z.; ASPECT Study Group. Adenovirus-mediated gene therapy with sitimagene ceradenovec followed by intravenous ganciclovir for patients with operable high-grade glioma (ASPECT): a randomised, open-label, phase 3 trial. *Lancet Oncol.* 2013;14(9):823-833. DOI 10.1016/S1470-2045(13)70274-2.
- Yakubitskiy S.N., Kolosova I.V., Maksyutov R.A., Shchelkunov S.N. Attenuation of vaccinia virus. *Acta Naturae.* 2015;7:108-117. (in Russian)
- Zav'yalov E.L., Razumov I.A., Gerlinskaya L.A., Romashchenko A.V. In vivo MRI visualization of growth and morphology in the orthotopic xenotransplantation U87 glioblastoma mouse SCID model. *Vavilovskii Zhurnal Genetiki i Selektzii = Vavilov Journal of Genetics and Breeding.* 2015;19(4):460-465. DOI 10.18699/VJ15.061. (in Russian)

Ренин-ангиотензин-альдостероновая система у крыс линии НИСАГ (ISIAH) со стресс-индуцированной артериальной гипертензией

А.Д. Дубинина¹✉, Е.В. Антонов¹, Л.А. Федосеева¹, Е.Н. Пивоварова¹, А.А. Маркель^{1, 2}, Л.Н. Иванова^{1, 2}

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Благодаря широкому спектру возможностей ренин-ангиотензиновой системы (РАС) в регуляции водно-солевого баланса и артериального давления изменение ее активности на системном или локальном (тканевом) уровнях в настоящее время рассматривается как один из наиболее значимых патогенетических факторов развития эссенциальной гипертензии. Цель работы – исследование циркуляторной и тканевой РАС у крыс НИСАГ со стресс-индуцированной артериальной гипертензией. Была измерена концентрация ренина, ангиотензин-превращающего фермента, ангиотензина II и альдостерона в сыворотке крови методом иммуноферментного анализа, а также исследована экспрессия основных генов РАС в почке, надпочечнике и отделах мозга с помощью полимеразной цепной реакции в реальном времени. Обнаружено, что в почке крыс НИСАГ снижена экспрессия мРНК гена ренина *Ren* по сравнению с нормотензивными крысами WAG, при этом концентрация ренина в сыворотке крови крыс данных линий не отличается. В то же время у крыс этой гипертензивной линии наблюдалось достоверное увеличение концентраций ангиотензина II и альдостерона в сыворотке крови, что может свидетельствовать о наличии «эктопического» очага синтеза ангиотензина II. В надпочечниках крыс НИСАГ экспрессия генов компонентов РАС не изменена. В структурах мозга показано увеличение экспрессии генов компонентов данной системы: *Ren* – в гипоталамусе, *Ace* – в стволе мозга, что подтверждает повышение базальной активности центральной РАС у крыс НИСАГ. Тем не менее проведенное исследование позволяет идентифицировать крыс линии НИСАГ как модель низкорениновой формы гипертонической болезни человека.

Ключевые слова: крысы линии НИСАГ; ренин-ангиотензин-альдостероновая система; низкорениновая гипертензия.

Renin-angiotensin-aldosterone system in ISIAH rats with stress-induced arterial hypertension

A.D. Dubinina¹✉, E.V. Antonov¹, L.A. Fedoseeva¹, E.N. Pivovarova¹, A.L. Markel^{1, 2}, L.N. Ivanova^{1, 2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

Because the renin-angiotensin system (RAS) has a wide range of opportunities in the regulation of fluid and electrolyte balance and arterial pressure, it is currently hypothesized that alterations in systemic circulating or local tissue RAS are some of the most important pathogenetic factors in the development of essential hypertension. The aim of the study was to investigate circulating and local tissue RAS activities in ISIAH rats with stress-induced arterial hypertension. We estimated the serum levels of renin, the angiotensin-converting enzyme, angiotensin II and aldosterone by an enzyme-linked immunosorbent assay, and mRNA expression of RAS genes in kidney, adrenals and brain tissues was measured by the real-time polymerase chain reaction. The mRNA expression of the renin gene (*Ren*) in the ISIAH rats was significantly decreased as compared to the normotensive WAG rats, but plasma renin concentrations had no difference. At the same time, the serum levels of angiotensin II and aldosterone in the ISIAH rats were enhanced, which suggests the existence of an ectopic site of angiotensin synthesis. Expression of RAS genes in the adrenals of hypertensive rats was unchanged. By contrast, a significant increase of RAS genes expression was found in the brain tissues. The mRNA of the *Ren* gene was increased in the hypothalamus, and the mRNA of *Ace* gene was increased in the brain stem of the ISIAH rats. This may be indicative of a local increase of RAS activity in the brain tissues of ISIAH rats. Nevertheless, the results of the study define ISIAH rat strain as a model of human low-renin hypertension.

Key words: ISIAH rat strain; renin-angiotensin-aldosterone system; low-renin hypertension.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Дубинина А.Д., Антонов Е.В., Федосеева Л.А., Пивоварова Е.Н., Маркель А.Л., Иванова Л.Н. Ренин-ангиотензин-альдостероновая система у крыс линии НИСАГ (ISIAH) со стресс-индуцированной артериальной гипертензией. Вавиловский журнал генетики и селекции. 2016;20(6):954-958. DOI 10.18699/VJ16.216

HOW TO CITE THIS ARTICLE:

Dubinina A.D., Antonov E.V., Fedoseeva L.A., Pivovarova E.N., Markel A.L., Ivanova L.N. Renin-angiotensin-aldosterone system in ISIAH rats with stress-induced arterial hypertension. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):954-958. DOI 10.18699/VJ16.216

Среди систем регуляции водно-электролитного баланса и артериального давления ключевую роль играет ренин-ангиотензиновая система (РАС), которая обеспечивает координацию работы сердечно-сосудистой системы, почек и надпочечников (Carey, Sigary, 2003; Coffman, Crowley, 2008). Снижение перфузионного давления в почечной артерии и увеличение активности симпатических нервов почки в результате снижения артериального давления стимулируют синтез и секрецию ренина в клетках юктагломерулярного аппарата почки, в то время как увеличение концентрации NaCl в канальцевой жидкости в области плотного пятна вследствие увеличения потребления соли с пищей оказывает ингибирующее действие на синтез и секрецию ренина (Montani, Van Vliet, 2004; Castrop et al., 2010). Ранее было показано, что повышение стресс-реактивности у крыс НИСАГ со стресс-чувствительной артериальной гипертензией проявляется в усилении стрессового ответа гипоталамо-гипофизарно-адренокортикальной и симпатoadреналовой систем (Markel et al., 2007). Со стороны симпатoadреналовой системы у крыс НИСАГ в надпочечниках наблюдаются повышенные содержание адреналина и экспрессия генов, кодирующих ключевые ферменты его биосинтеза, – тирозингидроксилазу, дофамин-β-гидроксилазу и фенилэтанол-амин-N-метилтрансферазу (Markel et al., 2007; Бузуева и др., 2010). Морфологические исследования мозгового вещества надпочечников крыс НИСАГ выявили структурные изменения, свидетельствующие о повышенной активности хромаффинных клеток уже в препубертатном периоде онтогенеза (Бузуева и др., 2010). Увеличение концентрации адреналина в крови при легком эмоциональном стрессе также более выражено у крыс НИСАГ, чем у крыс WAG (Markel et al., 2007). Предполагалось, что симпатoadреналовая гиперреактивность приводит к повышению активности синтеза ренина в почке крыс линии НИСАГ, что лежит в основе развивающейся гипертензии. Однако первые попытки оценить активность данной системы у крыс НИСАГ не подтвердили первоначальную гипотезу: было обнаружено, что активность ренина в плазме крови крыс НИСАГ не отличается от активности ренина в плазме нормотензивных крыс Вистар и WAG, а в ткани мозгового вещества почки даже понижена (Черкасова, Федоров, 2006; Amstislavsky et al., 2006). Цель настоящего исследования – охарактеризовать РАС крыс НИСАГ на системном и тканевом уровнях.

Материал и методы

Экспериментальные животные. Работа выполнена на взрослых (5 мес) самцах крыс двух инбредных линий – гипертензивной линии НИСАГ (ISIAH) и нормотензивной линии WAG. Все эксперименты проведены в соответствии с правилами Совета европейского сообщества (Директива 86/609/ЕЕС от 24 ноября 1986 г.) и с одобрения Комиссии по биоэтике Института цитологии и генетики СО РАН. Концентрацию ренина, ангиотензин-превращающего фермента, ангиотензина II и альдостерона в сыворотке крови определяли методом иммуноферментного анализа с помощью коммерческих наборов Renin (REN) BioAssay™ ELISA Kit (United States Biological, США), ELISA-ACE (Cloud-Clone Corp., США), Angiotensin II EIA Kit (Sigma-

Aldrich, США) и Aldosterone ELISA Kit (MyBioSource, США) соответственно.

Выделение мРНК. мРНК выделяли из гомогенатов тканей животных (почки, надпочечника, гипоталамуса и ствола мозга) с использованием TRI-REAGENT (IBCO/Life Technologies, США) согласно рекомендациям производителя. Примеси геномной ДНК удаляли обработкой ДНКазой I (Promega, США) согласно рекомендациям производителя. Количество РНК определяли спектрофотометрически.

Получение кДНК. Обратную транскрипцию проводили в растворе объемом 50 мкл, содержащем 1 мкг выделенной РНК, 0.25 нмоль праймеров (N_9 – случайные нонануклеотидные праймеры (ЗАО «Биосан», Новосибирск, Россия)), 36 мкл буфера для обратной транскрипции, 40 ед. акт. обратной транскриптазы MoMLV (все – ЗАО «Вектор-Бест», Новосибирск, Россия) и 0.4 мМ dNTP. Синтез кДНК проводили 1 ч при 37 °С, затем 30 мин при 42 °С и 10 мин при 50 °С. Фермент инактивировали прогреванием смеси 5 мин при 75 °С.

Приготовление «стандартной» кДНК. Из всех полученных образцов кДНК одной ткани отбирали аликвоты (по 3 мкл) и смешивали. Полученный «усредненный» раствор использовали в полимеразной цепной реакции (ПЦР) для построения калибровочных кривых.

ПЦР в реальном времени. Для определения уровня экспрессии генов проводили ПЦР в реальном времени на амплификаторе iQ5 (Bio-Rad Laboratories, США) с использованием SYBR Green I. Последовательности олигонуклеотидов подбирали с помощью программы PrimerQuest (Integrated DNA Technologies, <http://eu.idtdna.com/PrimerQuest/>). Степень гомологии праймеров проверяли с применением алгоритма BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>). Список праймеров приведен в табл. 1. ПЦР в реальном времени проводили при следующих условиях: прогрев 1 мин при 95 °С; далее – 35–40 основных циклов: 15 с при 95 °С, 20 с при температуре отжига праймеров $T_{отж}$ (см. табл. 1), 20 с при 72 °С, регистрация флуоресценции продуктов ПЦР 10 с при $T_{рег}$ (см. табл. 1); запись кривой плавления от 65 до 94 °С. По калибровочным кривым, полученным из разведений «стандартной» кДНК, определяли относительное содержание исследуемых кДНК. Межлинейные различия в уровнях экспрессии изучаемых генов определяли путем сравнения отношений количества кДНК целевых генов к количеству кДНК гена сравнения у животных разных линий.

Статистический анализ. Статистические расчеты были выполнены с использованием программного пакета STATISTICA 6.0. Данные представлены в виде $M \pm m$ (среднее значение ± стандартная ошибка среднего). Достоверность различий между показателями у крыс двух генотипов устанавливали с помощью двухстороннего критерия Стьюдента для независимых выборок. Статистически значимыми считали различия при значениях $p < 0.05$.

Результаты и обсуждение

Исследование показало, что концентрации ренина и ангиотензин-превращающего фермента в сыворотке крови гипертензивных крыс НИСАГ и нормотензивных

Table 1. Primers used in real-time PCR with SYBR Green I

Gene	Nucleotide sequence of primers	T _{anneal}	T _{acq}
<i>Ren</i>	F: 5'– CCT GGG AGT CAA AGA GAA GA – 3' R: 5'– ACA GGT CAT CGT TCC TGA AG – 3'	64 °C	84 °C
<i>Agt</i>	F: 5'– CCT CGC TCT CTG GAC TTA TC – 3' R: 5'– CAG ACA CTG AGG TGC TGT TG – 3'	63 °C	87 °C
<i>Ace</i>	F: 5'– ATG GTA CAG AAG GGC TGG AA – 3' R: 5'– TTG TAG AAG TCC CAC GCA GA – 3'	62 °C	88 °C
<i>Agtr1a</i>	F: 5'– AAA TGA GCA CGC TTT CTT ACC G – 3' R: 5'– TGA GGC AGG GTG AAT GGT CC – 3'	63 °C	86 °C
<i>Rpl30</i>	F: 5'– CAT CTT GGC GTC TGA TCT TG –3' R: 5'– TCA GAG TCT GTT TGT ACC CC –3'	61–64 °C*	84 °C

T_{anneal} – annealing temperature, T_{acq} – temperature of fluorescence signal acquisition.
* Primer annealing for the reference gene was performed at the same temperature as for the target gene.

Table 2. Concentrations of the renin-angiotensin-aldosterone axis components in the blood serum of ISIAH and WAG rat strains

Serum concentration of	WAG (n = 5)	НИСАГ (n = 5)	p
renin, pg/ml	87.407 ± 10.600	91.852 ± 26.424	NS
ACE, ng/ml	11.693 ± 1.858	12.404 ± 1.959	NS
angiotensin-II, pg/ml	16.96 ± 1.94	27.2 ± 2.6	p < 0.05 ↑
aldosterone, pg/ml	780 ± 73.64	1113 ± 56.74	p < 0.01 ↑

ACE, angiotensin-converting enzyme.

Table 3. mRNA expression of the renin-angiotensin axis genes in the kidney, adrenal, and brain tissues of the ISIAH and WAG rats

Genes	Kidney		Adrenals		Hypothalamus		Brain stem	
	WAG (n = 7)	НИСАГ (n = 7)	WAG (n = 7)	НИСАГ (n = 7)	WAG (n = 7)	НИСАГ (n = 7)	WAG (n = 7)	НИСАГ (n = 7)
<i>Ren</i>	1.10 ± 0.12	0.61 ± 0.07***	–	–	0.45 ± 0.058	0.63 ± 0.065*	1.15 ± 0.08	1.27 ± 0.071
<i>Agt</i>	1.075 ± 0.123	0.885 ± 0.129	0.71 ± 0.15	0.77 ± 0.16	1.30 ± 0.13	1.34 ± 0.16	1.00 ± 0.02	1.11 ± 0.013
<i>Ace</i>	1.23 ± 0.13	0.70 ± 0.08***	–	–	0.87 ± 0.07	0.74 ± 0.13	0.54 ± 0.04	0.73 ± 0.11***
<i>Agtr1a</i>	1.27 ± 0.14	0.84 ± 0.08**	0.94 ± 0.041	0.94 ± 0.043	0.34 ± 0.039	0.35 ± 0.042	0.74 ± 0.079	0.80 ± 0.037

* p < 0.05; ** p < 0.01; *** p < 0.001, interstrain differences; –, negligible expression

крыс WAG достоверно не различаются (табл. 2). Тем не менее у крыс НИСАГ обнаружено достоверное увеличение концентраций ангиотензина II ($p < 0.05$) и альдостерона ($p < 0.01$) в сыворотке крови.

В почке гипертензивных крыс снижена экспрессия мРНК генов ренина *Ren* ($p < 0.001$), ангиотензин-превращающего фермента *Ace* ($p < 0.001$) и рецептора ангиотензина II первого типа *Agtr1a* ($p < 0.01$) (табл. 3). В надпочечниках крыс НИСАГ, впрочем, как и контрольных крыс WAG, количество мРНК *Ren* и *Ace* было чрезвычайно низким, фактически на уровне чувствительности метода. Межлинейных различий экспрессии других исследуемых генов в надпочечниках обнаружено не было. В гипоталамусе крыс НИСАГ отмечено увеличение экспрессии ренинового гена *Ren* ($p < 0.05$), а в стволе мозга наблюдается увеличение концентрации мРНК гена *Ace* ($p < 0.001$).

Анализ активности системной РАС показал, что в почке гипертензивных крыс НИСАГ снижена экспрессия мРНК гена ренина *Ren* по сравнению с крысами WAG, в то же время концентрация ренина в сыворотке крови крыс данных линий не отличается. Эти данные совпадают с полученными нами ранее результатами по иммуногистохимической оценке содержания ренина и мРНК гена ренина в прегломерулярных афферентных артериолах в почках крыс НИСАГ и измерению активности этого фермента в крови гипертензивных крыс (Amstislavsky et al., 2006), что в совокупности является убедительным доказательством пониженной активности РАС у крыс линии НИСАГ. Активность почечной РАС у крыс НИСАГ также снижена, о чем свидетельствует уменьшение экспрессии других компонентов данной системы: ангиотензин-превращающего фермента и ангиотензинового рецептора

типа 1А (см. табл. 2). Возникает вопрос, почему в таком случае наблюдается повышение концентраций ангиотензина II и альдостерона в плазме крови гипертензивных крыс (см. табл. 2). На первый взгляд эта ситуация напоминает картину, наблюдаемую у трансгенных крыс линии TGR(mREN2)27 (Langheinrich et al., 1996), полученной путем искусственного переноса в геном крыс дополнительного мышинного гена ренина *Ren2* (Mullins et al., 1990). Для крыс данной линии характерно наличие тяжелой формы гипертонии, однако активность ренина в плазме периферической крови остается либо нормальной, либо даже пониженной; в ткани почек содержание ренина также снижено. Более того, отмечается явное подавление активности РАС почки, о чем свидетельствует сниженный уровень экспрессии генов данной системы в почке. Было показано, что развитие гипертонии у трансгенных крыс, по-видимому, связано с изменением активности локальной РАС главным образом в надпочечнике, где была зарегистрирована повышенная экспрессия ренинового трансгена. Гиперактивация синтеза ренина в надпочечниках, вероятно, служила причиной повышенной секреции альдостерона. В свою очередь, снижение активности секреции ренина клетками юкстагломерулярного аппарата было связано с повышением перфузионного давления в почечных клубочках. Следовательно, причиной развития гипертонии в данном случае стало появление «эктопического» очага синтеза ренина в железе, синтезирующей альдостерон.

Таким образом, появилась необходимость исследовать активность экспрессии генов РАС в надпочечниках крыс НИСАГ, которая могла бы быть увеличена в результате длительной селекции этих крыс на повышенную стрессреактивность. Результаты исследования полностью исключили данное предположение. Количество мРНК ренинового гена и гена ангиотензин-превращающего фермента в надпочечниках крыс НИСАГ так же, как и у контрольных крыс WAG, было чрезвычайно низким, фактически на уровне чувствительности метода (см. табл. 2). Таким образом, исследование надпочечников не дало ответа на вопрос о причине повышения уровня ангиотензина II и альдостерона в плазме крыс НИСАГ.

В последние десятилетия все больше внимания уделяется РАС мозга и ее участию в регуляции симпатической активности и артериального давления (Coble et al., 2015). Выполненное нами исследование активности генов данной системы в отделах мозга – гипоталамусе и стволе мозга – выявило наличие существенных межлинейных различий: экспрессия ренинового гена была повышена в гипоталамусе крыс НИСАГ, а в стволе мозга отмечено достоверное увеличение концентрации мРНК гена *Ace*. На основании этих данных можно предположить повышение синтеза ангиотензина II в отделах мозга крыс НИСАГ. Гипотеза о центральном происхождении части циркулирующего ангиотензина II, на первый взгляд, кажется невозможной (Pardridge, 1983). Однако литературные данные о проницаемости гематоэнцефалического барьера для некоторых пептидов, в частности для ангиотензина II, дают определенную надежду на объяснение повышенного содержания «тандема» ангиотензин–альдостерон в крови крыс НИСАГ выходом ангиотензина II

из некоторых отделов мозга. Известно, что в области циркумвентрикулярных органов, к которым относятся субфорникальный орган, сосудистый орган терминальной пластинки и *area postrema* (самое заднее поле гипоталамуса), гематоэнцефалический барьер является наиболее проницаемым, что позволяет таким низкомолекулярным пептидам крови, как ангиотензин II, взаимодействовать с данными областями мозга (Coble et al., 2015). При гипертонии показано нарушение целостности гематоэнцефалического барьера (Pelisch et al., 2011; Biancardi et al., 2014; Biancardi, Stern, 2016), благодаря чему ангиотензин II из циркулирующей крови может также проникать в паравентрикулярное ядро гипоталамуса и ростральный отдел вентролатерального продолговатого мозга, вызывая повышение активности симпатической нервной системы. Нужно отметить, что нарушение гематоэнцефалического барьера отчасти может быть результатом воспалительных процессов в стенках микрососудов головного мозга, индуцируемых ангиотензином II (Zhang et al., 2010). Можно предположить, что дополнительным источником циркулирующего ангиотензина II у крыс НИСАГ являются некоторые отделы головного мозга, чему способствует повышение транскрипционной активности генов РАС в этих отделах.

Для крыс линии НИСАГ помимо гипертонии характерны дислипидемия (повышенное содержание триглицеридов и холестерина липопротеинов очень низкой и низкой плотности, сниженное содержание холестерина липопротеинов высокой плотности в сыворотке крови), повышенные содержание глюкозы в крови и масса тела по сравнению с крысами WAG (Пивоварова и др., 2011). Таким образом, крысы НИСАГ демонстрируют весь симптомокомплекс, характерный для метаболического синдрома. В настоящее время в качестве одной из причин развития метаболического синдрома рассматривают нарушение работы РАС в жировой ткани (Yasue et al., 2010). Согласно данной гипотезе, увеличение синтеза ангиотензиногена и ангиотензина II в адипоцитах способствует повышению концентрации этих белков в циркуляторном русле, что приводит к развитию гипертонии у человека. Более того, ангиотензин II, помимо влияния на сердечно-сосудистую систему и почки, вызывает инсулинорезистентность, нарушение толерантности к глюкозе, подавляет дифференцировку адипоцитов, способствуя их гипертрофии, стимулирует липогенез и секрецию провоспалительных адипокинов адипоцитами (Frigolet et al., 2013), в результате чего развивается метаболический синдром. При этом активность РАС в жировой ткани не зависит от активности системной РАС: экспрессия ренина в жировой ткани не коррелирует с концентрацией этого фермента в крови (Marcus et al., 2013). На основании этих данных можно предположить, что повышенная концентрация ангиотензина II при сниженной концентрации ренина в крови НИСАГ может быть обусловлена гиперактивацией РАС жировой ткани. Однако эта гипотеза требует экспериментального подтверждения.

Тем не менее необходимо подчеркнуть, что общая картина гипертонии у крыс линии НИСАГ соответствует низкорениновой форме артериальной гипертонии, которая встречается у людей не менее чем в 25 % случаев (Sahay,

Sahay, 2012). Для данной формы артериальной гипертензии характерны понижение синтеза ренина в почке и повышение минералокортикоидной активности, что приводит к повышенной солевой чувствительности. Нами показано, что содержание крыс НИСАГ в условиях повышенной солевой нагрузки (питье 2 % раствора NaCl) в течение недели приводит к повышению артериального давления до 230 мм рт. ст. (Неопубл. данные).

Таким образом, крысы линии НИСАГ представляют собой модель низкорениновой формы гипертонической болезни человека. Наличие такой модели позволит провести целенаправленные исследования, посвященные раскрытию механизмов формирования низкорениновой гипертонической болезни, которые до сих пор остаются мало изученными.

Acknowledgments

This work was supported by State Budgeted Projects 0324-2015-0004 and 0324-2015-0022 and by the Russian Foundation for Basic Research, project 16-04-00763.

Conflict of interest

The authors declare no conflict of interest.

References

- Amstislavsky S., Welker P., Frühauf J.H., Maslova L., Ivanova L., Jensen B., Markel A.L., Bachman S. Renal and endocrine changes in rats with inherited stress-induced arterial hypertension (ISIAH). *Histochem. Cell Biol.* 2006;125(6):651-659. DOI 10.1007/s00418-005-0118-5.
- Biancardi V.C., Son S.J., Ahmadi S., Filosa J.A., Stern J.E. Circulating angiotensin II gains access to the hypothalamus and brain stem during hypertension via breakdown of the blood-brain barrier. *Hypertension.* 2014;63(3):572-579. DOI 10.1161/HYPERTENSIONAHA.113.01743.
- Biancardi V.C., Stern J.E. Compromised blood-brain barrier permeability: novel mechanism by which circulating angiotensin II signals to sympathoexcitatory centres during hypertension. *J. Physiol.* 2016;594(6):1591-1600. DOI 10.1113/JP271584.
- Buzueva I.I., Filyushina E.E., Shmerling M.D., Markel A.L., Jacobson G.S. The chronic stress influence on the adrenal glands structure in hypertensive ISIAH rats after preventive treatment with terazosin. *Byulleten SO RAMN = Bulletin of SB RAMS.* 2010;30(4):56-61. (in Russian)
- Carey R.M., Siragy H.M. Newly recognized components of the renin-angiotensin system: potential roles in cardiovascular and renal regulation. *Endocr. Rev.* 2003;24:261-271. DOI 10.1210/er.2003-0001.
- Castrop H., Höcherl K., Kurtz A., Schweda F., Todorov V., Wagner C. Physiology of kidney renin. *Physiol. Rev.* 2010;90(2):607-673. DOI 10.1152/physrev.00011.2009.
- Cherkasova O.P., Fedorov V.I. Renin activity in kidney and plasma in hereditary stress-induced arterial hypertension. *Vestnik TGU = Tomsk State University Journal.* 2006;21:167-168. (in Russian)
- Coble J.P., Grobe J.L., Johnson A.K., Sigmund C.D. Mechanisms of brain renin angiotensin system-induced drinking and blood pressure: importance of the subfornical organ. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 2015;308:R238-R249. DOI 10.1152/ajpregu.00486.2014.
- Coffman T.M., Crowley S.D. Kidney in hypertension: guyton redux. *Hypertension.* 2008;51(4):811-816. DOI 10.1161/HYPERTENSIONAHA.105.063636.
- Frigolet M.E., Torres N., Tovar A.R. The renin-angiotensin system in adipose tissue and its metabolic consequences during obesity. *J. Nutr. Biochem.* 2013;24:2003-2015. DOI 10.1016/j.jnutbio.2013.07.002.
- Langheinrich M., Lee M.A., Böhm M., Pinto Y.M., Ganten D., Paul M. The hypertensive Ren-2 transgenic rat TGR (mIXEN) 27 in hypertension research – characteristics and functional aspects. *Am. J. Hypertens.* 1996;9:506-512.
- Marcus Y., Shefer G., Stern N. Adipose tissue renin-angiotensin-aldosterone system (RAAS) and progression of insulin resistance. *Mol. Cell. Endocrinol.* 2013;378:1-14.
- Markel A.L., Redina O.E., Gilinsky M.A., Dymshits G.M., Kalashnikova E.V., Khvorostova Y.V., Fedoseeva L.A., Jacobson G.S. Neuroendocrine profiling in inherited stress-induced arterial hypertension rat strain with stress-sensitive arterial hypertension. *J. Endocrinol.* 2007;195(3):439-450. DOI 10.1677/JOE-07-0254.
- Montani J.-P., Van Vliet B.N. General physiology and pathophysiology of the renin-angiotensin system. *Handbook of Experimental Pharmacology.* Ed. T. Unger, B.A. Scholkens. Berlin: Springer Verlag, 2004;163(1):3-29.
- Mullins J.J., Peters J., Garnten D. Fulminant hypertension in transgenic rats harboring the mouse Ren-2 gene. *Nature.* 1990;344:541-544. DOI 10.1038/344541a0.
- Pardridge W.M. Neuropeptides and the blood-brain barrier. *Ann. Rev. Physiol.* 1983;45:73-82. DOI 10.1146/annurev.ph.45.030183.000445.
- Pelisch N., Hosomi N., Ueno M., Nakano D., Hitomi H., Mogi M., Shimada K., Kobori H., Horiuchi M., Sakamoto H., Matsumoto M., Kohno M., Nishiyama A. Blockade of AT1 receptors protects the blood-brain barrier and improves cognition in Dahl salt-sensitive hypertensive rats. *Am. J. Hypertens.* 2011;24(3):362-368. DOI 10.1038/ajh.2010.241.
- Pivovarova E.N., Dushkin M.I., Perepechaeva M.L., Kobzev V.F., Trufakin V.A., Markel A.L. All signs of metabolic syndrome in hypertensive ISIAH rats are associated with elevated activity of transcription factors PPAR, LXR, PXR, and CAR in the liver. *Biomeditsinskaya khimiya = Biomedical Chemistry.* 2011;57(4):435-445. (in Russian)
- Sahay M., Sahay R.K. Low renin hypertension. *Indian J. Endocrinol. Metab.* 2012;16(5):728-740. DOI 10.4103/2230-8210.100665.
- Yasue S., Masuzaki H., Okada S., Ishii T., Kozuka C., Tanaka T., Fujikura J., Ebihara K., Hosoda K., Katsurada A., Ohashi N., Urushihara M., Kobori H., Morimoto N., Kawazoe T., Naitoh M., Okada M., Sakaue H., Suzuki S., Nakao K. Adipose tissue-specific regulation of angiotensinogen in obese humans and mice: impact of nutritional status and adipocyte hypertrophy. *Am. J. Hypertens.* 2010;23:425-431. DOI 10.1038/ajh.2009.263.
- Zhang M., Mao Y., Ramirez S.H., Tuma R.F., Chabrashvili T. Angiotensin II induced cerebral microvascular inflammation and increased blood-brain barrier permeability via oxidative stress. *Neuroscience.* 2010;171(3):852-858. DOI 10.1016/j.neuroscience.2010.09.029.



Аутизм как проявление нарушения молекулярных механизмов регуляции развития и функций синапсов

Е.А. Трифонова , Т.М. Хлебодарова, Н.Е. Груntenко

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

Расстройства аутистического спектра – отдельная группа дефектов развития с очень высокой генетической компонентой. Генетический скрининг выявил множество различных генетических вариаций, связанных с аутизмом, а биоинформатический анализ сигнальных путей и генов сетей привел к пониманию, что многие из этих мутационных изменений вовлечены в функционирование синапсов. Синапс является местом электрохимической коммуникации между нейронами и необходимой субъединицей для обучения и формирования памяти. Межнейронная коммуникативная связь пластична, и наиболее стойкие формы синаптической пластичности сопровождаются изменениями в биосинтезе белка как в теле нейронов, так и локально в дендритах. Локальная трансляция – это тонко регулируемый процесс, центральную роль в регуляции инициации которого играет сигнальный путь mTOR (mammalian or mechanistic target of rapamycin). Мутационное повреждение хотя бы одного из звеньев этого сигнального пути приводит к нарушениям синаптической пластичности и поведения. С нарушениями регуляции локальной трансляции в дендритах связаны моногенные синдромы Нунана, Костелло, Каудена, Ретта, туберозный склероз, нейрофиброматоз I типа и синдром ломкой X-хромосомы, у части носителей которых также диагностируются расстройства аутистического спектра. Данный обзор посвящен молекулярным механизмам синдромного аутизма, обусловленного моногенными мутациями, а также механизм-обоснованной терапии некоторых расстройств аутистического спектра.

Ключевые слова: расстройства аутистического спектра (РАС); синапс; mTOR; механизм-обоснованная терапия; синдромный аутизм.

Molecular mechanisms of autism as a form of synaptic dysfunction

E.A. Trifonova , T.M. Khlebarova, N.E. Gruntenko

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Autism spectrum disorders are a separate group of defects with a very high genetic component. Genetic screening has identified hundreds of mutations and other genetic variations associated with autism, and bioinformatic analysis of signaling pathways and gene networks has led to understanding that many of these mutational changes are involved in the functioning of synapses. A synapse is a site of electrochemical communication between neurons and an essential subunit for learning and memory. Interneuronal communicative relationships are plastic. The most prominent forms of synaptic plasticity are accompanied by changes in protein biosynthesis, both in neuron body and in dendrites. Protein biosynthesis or translation is a carefully regulated process, with a central role played by mTOR (mammalian or mechanistic target of rapamycin). Normally mTOR-regulated translation is slightly inhibited, and in most cases mutational damage to at least one of the links of the mTOR signaling pathway, increases translation and leads to impaired synaptic plasticity and behavior. Deregulation of the local translation in dendrites is connected with the following monogenic autism spectrum disorders: neurofibromatosis type 1, Noonan syndrome, Costello syndrome, Cowden syndrome, tuberous sclerosis, fragile X chromosome, syndrome, and Rett syndrome. The review considers the most important mutations leading to monogenic autism, as well as the possibility of a mechanism-based treatment of certain disorders of the autism spectrum.

Key words: autism spectrum disorders (ASD); the synapse; mechanistic or mammalian target of rapamycin (mTOR); mechanism-based therapy; syndromic autism.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Трифопова Е.А., Хлебодарова Т.М., Груntenко Н.Е. Аутизм как проявление нарушения молекулярных механизмов регуляции развития и функций синапсов. Вавиловский журнал генетики и селекции. 2016; 20(6):959-967. DOI 10.18699/VJ16.217

HOW TO CITE THIS ARTICLE:

Trifonova E.A., Khlebarova T.M., Gruntenko N.E. Molecular mechanisms of autism as a form of synaptic dysfunction. Vavilovskii Zhurnal Genetiki i Selektcii = Vavilov Journal of Genetics and Breeding. 2016;20(6):959-967. DOI 10.18699/VJ16.217

REVIEW

Received 17.11.2016

Accepted for publication 22.12.2016

© AUTHORS, 2016



e-mail: trifonova.k@rambler.ru

Расстройства аутистического спектра (РАС) – это комплексные нарушения психического развития, процент диагностирования которых непрерывно растет в последние 40 лет и в настоящее время составляет более 1 % (1 из 68) детского населения. Основные симптомы аутизма можно разделить на: 1) значительное снижение коммуникативных и социальных функций, в том числе речевой коммуникации и 2) значительное повышение стереотипии в поведении и речи, крайняя ограниченность интересов. Имеющиеся в настоящее время лекарственные препараты не способны скорректировать основные симптомы и нацелены лишь на облегчение некоторых, часто связанных с аутизмом, состояний, таких как эпилепсия, депрессия, гневливость и тревожность (Veenstra-VanderWeele, Blakely, 2012).

Изучение семейных историй и тесты на однояйцевых близнецах показали крайне высокую наследуемость аутизма (в 73–95 % диагностировались оба близнеца), а генетический скрининг выявил сотни генов предрасположенности к РАС (Ebrahimi-Fakhari, Sahin, 2015).

Дальнейшие исследования позволили выделить формы расстройства, наблюдаемые у пациентов с моногенными синдромами, фенотипические проявления которых часто включают аутизм. Самые известные из этих синдромов – синдром ломкой X-хромосомы, нейрофиброматоз, синдром Ретта и туберозный склероз. Ранее считалось, что редкие формы синдромного аутизма ответственны за 10–20 % случаев аутизма, причем каждый отдельный моногенный синдром встречается не более чем у 1 % пациентов с РАС (El-Fishawy, State, 2010). Но с развитием полноэкзомного секвенирования и его применения для анализа геномов аутистов стало ясно, что более 25 % случаев РАС может быть объяснено моногенными мутациями (Yoo, 2015). Следовательно, около 70 % случаев так называемого типичного аутизма имеют другую генетическую природу (El-Fishawy, State, 2010), возможно, связанную с мутациями в некодирующих последовательностях ДНК.

Моногенные синдромы сыграли огромную роль в изучении молекулярных механизмов РАС. С помощью модельных животных, несущих мутации, приводящие к разным типам синдромного аутизма, было продемонстрировано, что нарушения в структуре, функциях или формировании синапсов являются основной причиной возникновения фенотипа РАС (Singh, Eroglu, 2013). Благодаря полногеномным ассоциативным исследованиям обнаружены сотни генетических вариаций, связанных с аутизмом, а биоинформатический анализ сигнальных путей и генных сетей привел к пониманию, что эти мутационные изменения вовлечены в функционирование синапсов (рис. 1). В последнее время для расстройств, связанных с нарушениями синаптической передачи, стали использовать термин «синапсопатия» (synapsopathy). Успешный поиск лекарств для лечения различных синапсопатий – прекрасная иллюстрация использования методов молекулярной биологии в медицине. Типичный путь от открытия нового гена, мутация в котором вызывает некий известный синдром, до создания механизма-обоснованной терапии представлен на рис. 2.

Для некоторых мутаций, приводящих к синдромному аутизму, вышеописанный путь пройден уже более чем на

половину, и в данном обзоре мы попытаемся обобщить все доступные на сегодняшний день экспериментальные факты, связывающие РАС с некоторыми синапсопатиями, и возможности фармакологической коррекции этих патологий.

Синдром Фелан–МакДемид и связанные с геном *SHANK3* заболевания аутистического спектра

Синдром Фелан–МакДемид (СФМ) является результатом микроделеции 22q13 и характеризуется снижением интеллекта, задержкой или отсутствием речи, гипотонией, РАС и часто эпилепсией (Phelan, 2001). Ключевой ген, мутация в котором ответственна за формирование СФМ, – *SHANK3*, – кодирует структурный белок, отвечающий за фиксацию рецепторов глутамата и нейролиггин-нейрексиновых комплексов в постсинаптическом пространстве (Sheng, Kim, 2000). Чаще всего СФМ встречается у гетерозигот по делеции *SHANK3*, причем протяженность этой делеции может существенно варьировать и иногда сводиться к точечной мутации. Кроме того, было зафиксировано несколько случаев дупликаций данного гена, приводящих к синдрому Аспергера или снижению интеллекта (Durand et al., 2007; Okamoto et al., 2007).

Белок Shank3 входит в семейство, состоящее из Shank 1, 2 и 3 белков, являющихся структурными компонентами постсинаптического пространства, которые также называются обогащенными пролином синапс-связанными белками (ProSAPs). Shank-белки содержат несколько доменов белок-белкового взаимодействия, позволяющих им связываться со множеством белков в постсинаптическом пространстве, в том числе с рецепторами глутамата, нейролиггин-нейрексиновыми комплексами и компонентами актинового цитоскелета (Böckers et al., 2001). В результате этих взаимодействий Shank-белки модулируют морфологию дендритных шипиков и синаптическую передачу.

Эксперименты подтвердили важную роль Shank3-белка для функционирования синапсов. Так, в культуре нейронов гиппокампа нокаун гена *SHANK3* привел к увеличению длины и уменьшению количества дендритных шипиков (Roussignol et al., 2005). У модельных мышей с мутантным вариантом гена *SHANK3* была снижена АМРА-опосредованная синаптическая передача, нарушена длительная потенция в гиппокампе (Bozdagi et al., 2010), изменены соотношения белков в постсинаптическом пространстве и морфология дендритов (Kouser et al., 2013; Peça et al., 2011).

Для изучения механизмов болезни были созданы линии плюрипотентных стволовых клеток от пациентов с СФМ, на которых было показано, что потеря функциональности Shank3 приводит в нейронах к снижению амплитуды и частоты постсинаптического возбуждения и уменьшает количество возбуждающих синапсов. Обработка инсулиноподобным фактором роста 1 (IGF-1) восстанавливала нормальную функциональность синапсов. В настоящее время IGF-1 проходит вторую фазу клинических испытаний для лечения СФМ и РАС (Shcheglovitov et al., 2013), хотя молекулярные механизмы действия IGF-1 не ясны.

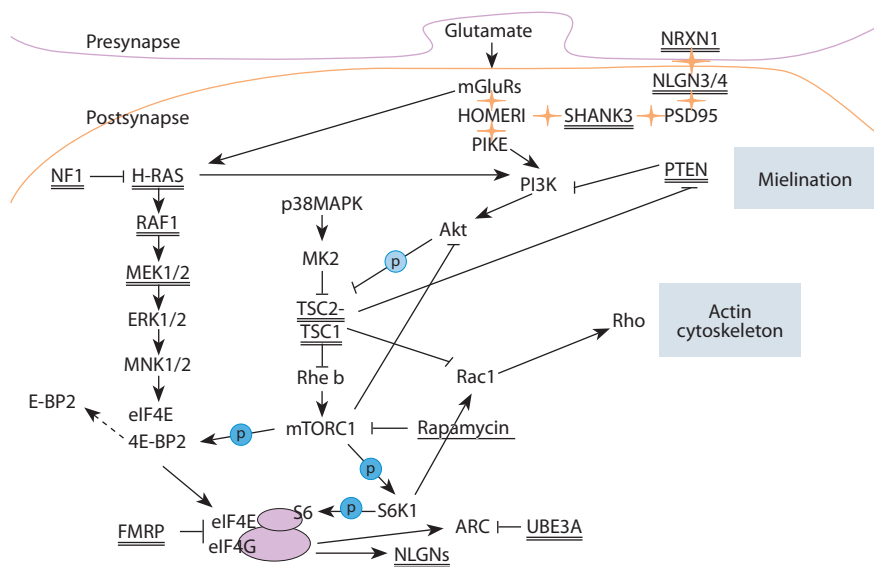


Fig. 1. The signaling pathway in an axodendritic glutamergic synapse, examples of the cell location of gene products associated with different types of typical and syndromic autism, and examples of signaling pathways involved in the formation of the ASD phenotype.

Products of mutant genes increasing ASD probability are shown with double underlining. SHANK3, Phelan–McDermid syndrome; PTEN, Cowden syndrome; NF1, type I neurofibromatosis; H-RAS, RAF1, and MEK1, Costello–Noonan syndrome; TSC2–TSC1, tuberous sclerosis; FMRP, fragile X syndrome; UBE3A, Angelman syndrome; NLGN3/4 and NRXN1, typical autism (from (Zoghbi, Bear, 2012), with modifications).

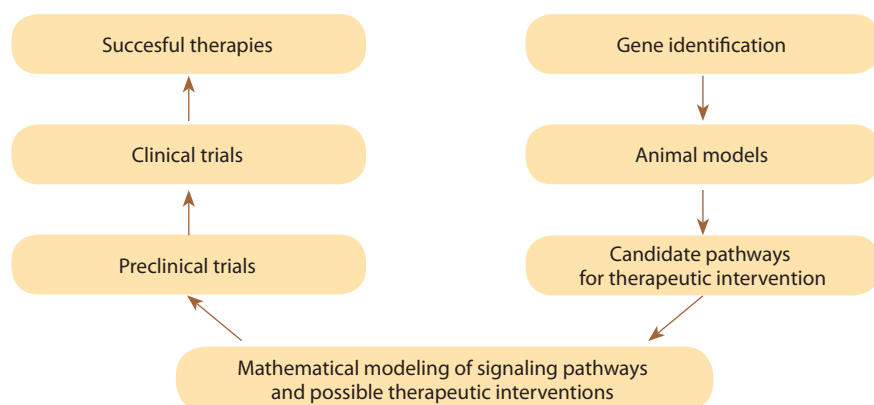


Fig. 2. Approach to mechanism-based treatment.

РАС, связанные с дисфункциями нейролигинов и нейрексинов

Нейролигины и нейрексины являются молекулами клеточной адгезии, необходимыми для функционирования синапсов. Нейролигины расположены на постсинаптической мембране, а нейрексины – на пресинаптической мембране (см. рис. 1). Мутации в генах *NLGN3* и *NLGN4* были описаны в числе первых, связанных с проявлениями РАС. Эти гены локализованы в X-хромосоме, их мутации имеют рецессивный характер, поэтому фенотипические проявления аутизма заметны только у мужчин. Для носителей мутаций в генах нейролигинов характерны различные проявления – от синдрома Аспергера до аутизма со значительно сниженным интеллектом, судорожными приступами и тиками, что связано, по-видимому, с различной функциональностью мутантных белков (Jamain et al., 2003). Мутации в генах нейрексинов также приводят к фенотипу РАС (Zweiger et al., 2009). Три гена нейрексинов, *NRXN1*, *NRXN2* и *NRXN3*, кодируют по две изоформы белков каждый, которые счита-

ются с независимых промоторов, а сложный альтернативный сплайсинг генерирует тысячи вариантов белков (Missler, Südhof, 1998). Нейролигины являются эндогенными лигандами нейрексинов и кодируются четырьмя генами: *NLGN1*, *NLGN2*, *NLGN3* и *NLGN4*, мРНК которых также подвергаются альтернативному сплайсингу. Как нейрексины, так и нейролигины имеют по одному трансмембранному и короткому цитоплазматическому домену, последний содержит PDZ-мотив на карбоксильном конце (Hata et al., 1996; Irie et al., 1997). В норме нейролигины и нейрексины формируют транс-синаптический комплекс, организующий постсинаптический и пресинаптический компартменты путем связывания с белками CASK, MAGUK и PSD95 (Südhof, 2008).

Эксперименты с модельными животными показали, что отсутствие трех нейролигинов, *Nlgn1*, *Nlgn2* и *Nlgn3*, приводит к перинатальной гибели эмбрионов из-за респираторной недостаточности (Varghaeas et al., 2006). Полное отсутствие α-нейрексинов также приводит к гибели эмбрионов из-за проблем с дыханием, но уже пренатально (Missler et al., 2003). У мышей, несущих аллель *NLGN3* с мутацией R451C, впервые обнаруженной у одного из пациентов с РАС, было зафиксировано поведение, аналогичное аутизму без снижения обучаемости, но с нарушением социального взаимодействия (Tabuchi et al., 2007). Делеция ортолога *NLGN4* гена человека у мышей вызывала нарушение социального взаимодействия и уменьшение ультразвуковых сигналов, что является аналогией РАС у человека (Jamain et al., 2008).

Синдром Энгельмана (синдром счастливой куклы)

Синдром Энгельмана (СЭ) характеризуется задержкой развития, сниженным интеллектом, задержкой речевого развития, проблемами со сном и уникальным типом постоянно счастливой и дружелюбного поведения, которое и объясняет второе название расстройства – «синдром счастливой куклы». В большинстве случаев СЭ связывают с делецией 15q11-q13, в районе расположения которой локализуется ген убиквитин-лигазы *E3* (*UBE3A*). Именно его считают от-

ветственным за появление данного синдрома (Kishino et al., 1997; Matsuura et al., 1997).

У модельных мышей, нокауты по гену *UBE3A*, снижены плотность и сила возбуждающих синапсов, нарушены длительная потенция в гиппокампе и способность к обучению (Jiang et al., 1998).

Очевидно, что проблемы в СЭ возникают в результате избытка белков-мишеней убиквитин-лигазы E3 в пост-синаптическом пространстве, одним из которых является ассоциированный с цитоскелетом белок Arc. У мышей, мутантных по *UBE3A*, экспрессия *Arc* повышена. Функции белка Arc связаны с интернализацией AMPA подтипа рецепторов глутамата в возбуждающих синапсах (Greer et al., 2010). Трансляция Arc регулируется FMRP, белком, ответственным за формирование синдрома ломкой X-хромосомы.

Предполагается, что гиперэкспрессия Arc является точкой пересечения в патофизиологии СЭ и синдрома ломкой X-хромосомы (Kelleher, Bear, 2008) и ингибирование mGluR5 может скорректировать фенотип СЭ так же, как и синдром ломкой X-хромосомы (Zoghbi, Bear, 2012).

Для коррекции некоторых черт мутантного фенотипа СЭ был использован препарат L-ДОФА (L-3,4-дигидроксифенилаланин, L-DOPA), предшественник катехоламинов. Он прошел клинические испытания, хотя молекулярные механизмы, обуславливающие его эффективность, еще не вполне понятны. Показано только, что мутация или делеция гена *Ube3a* у мышей нарушает функционирование дофаминэргической системы (Riday et al., 2012).

РАС, связанные с нарушением регуляции сигнального пути mTOR

Механистическая или типичная для млекопитающих мишень для рапамицина mTOR (mammalian or mechanistic target of rapamycin) – серин/треонин киназа. Киназа mTOR является ключевым регулятором инициации трансляции (Sato, 2016) и действует как центральный компонент двух мультибелковых комплексов, mTORC1 и mTORC2, которые различаются композицией белков и субстратами (Lipton, Sahin, 2014).

Сигнальный путь mTOR интегрирует множество внутриклеточных и экстраклеточных сигналов, включая факторы роста, нутриенты, стресс, инфекции (Lisse, Hewison, 2011). Он вовлечен в регуляцию иммунного ответа (Liu et al., 2015), клеточного и опухолевого роста (Zhou, Parada, 2012), а также в формирование памяти и долговременной синаптической пластичности. Большая часть доказательств связи функций mTOR с синаптической пластичностью получена с использованием рапамицина (сиролимуса), ингибитора mTOR. Исследования на генетически модифицированных животных подтвердили вовлеченность этого сигнального пути в регуляцию пластичности межнейронных связей (Ehninger et al., 2009).

Пластичность межнейронной коммуникативной связи означает, что реактивность, т.е. сила синаптического соединения, изменяется с течением времени, причем некоторые изменения происходят за секунды, а другие могут поддерживаться в течение всей жизни организма. У позвоночных долговременные изменения в силе синап-

тического соединения часто описываются как длительная потенция (LTP) в случае усиления синапса и длительная депрессия (LTD) при его ослаблении.

Наиболее стойкие формы синаптической пластичности сопровождаются изменениями в биосинтезе белка как в теле нейронов, так и локально в дендритах (Lipton, Sahin, 2014). Мутационное повреждение любого компонента сигнального пути mTOR приводит к нарушениям локальной трансляции, и как следствие, синаптической пластичности и поведения.

С нарушениями регуляции локальной трансляции в дендритах связаны следующие моногенные расстройства аутистического спектра: нейрофиброматоз I типа, синдром Нунана, синдром Костелло, синдром Каудена, туберозный склероз, синдром ломкой X-хромосомы и синдром Ретта (Zoghbi, Bear, 2012).

Синдром ломкой X-хромосомы

Синдром ломкой X-хромосомы (СХХ) диагностируется у 5–8 % детей с РАС и обычно характеризуется глубоким аутизмом со сниженным интеллектом. Подавляющее большинство пациентов с СХХ мужского пола. Фенотипические проявления СХХ у них весьма значительны и включают макроорхидизм, увеличенный размер ушей, асимметричные черты лица, увеличенные руки, часто аномалии соединительной ткани, митрального клапана и аорты. У трети женщин, носителей этой мутации, при отсутствии каких-либо иных фенотипических проявлений снижен интеллект.

Фенотип СХХ вызывается транскрипционным сайленсингом гена *FMR1*, в 5'-нетранслируемом районе которого содержится нестабильный при передаче в поколениях CGG-повтор. В норме количество CGG-повторов не превышает 55, при СХХ количество повторов достигает 200 и более, что приводит к метилированию всего промоторного района гена *FMR1*, замолчанию гена, и, следовательно, отсутствию его белкового продукта FMRP. FMRP является РНК-связывающим белком, и его функции включают транспорт большого количества мРНК в дендриты, прямую регуляцию локальной трансляции и стабильности мРНК в синапсе (Bagni, Oostra, 2013).

В экспериментах с модельными мышами, у которых нокаутирован ген, гомологичный *FMR1* (*Fmr1* KO), показано повышение базального уровня белкового синтеза в гиппокампе. Попытки скорректировать избыточный биосинтез белка в синапсах *Fmr1* KO мышей были предприняты с использованием негативных регуляторов метаболитных рецепторов глутамата (mGluR1 и mGluR5). В настоящее время четыре различных препарата, являющихся негативными регуляторами mGluR5 рецепторов, находятся на различных стадиях клинических испытаний для коррекции аберрантного поведения при СХХ (Ebrahimi-Fakhari, Sahin, 2015). Другой перспективной возможностью коррекции СХХ фенотипа оказалась активация рецепторов GABA, которая смягчала проявление синдрома как у модельных мышей (Pacey et al., 2009), так и у мух дрозофил (Chang et al., 2008). R-баклофен, селективный агонист GABA-рецепторов, успешно прошел третью фазу клинических испытаний (Ebrahimi-Fakhari, Sahin, 2015). Альтернативный подход к коррекции фенотипа СХХ был

использован В. Доланом с коллегами (Dolan et al., 2013), они предположили, что коррекция структурных дефектов дендритов приведет к улучшению поведенческих аномалий у модельных животных. В экспериментах были также использованы мыши *Fmr1 KO*, а в качестве белковой мишени была выбрана p21-активируемая киназа (PAK), которая модулирует динамику актинового цитоскелета. Продemonстрировано, что даже однократного применения ингибитора PAK, небольшой молекулы, названной авторами FRAX486, оказалось достаточно для полной коррекции CXX-фенотипа у взрослых *Fmr1 KO* мышей. Улучшение поведения наблюдалось на фоне полной нормализации морфологии дендритных шипиков, на основании чего авторы сделали вывод, что их подход доказывает возможность быстрой коррекции CXX даже у взрослых пациентов (Dolan et al., 2013).

Туберозный склероз

Туберозный склероз – мультисистемное расстройство, характерной чертой которого являются доброкачественные опухоли (гамартомы), расположенные во многих органах и тканях, чаще всего в мозге, на коже, глазах, в почках и сердце (Curatolo et al., 2008). Неврологические проявления туберозного склероза у пациентов очень варьируют – от полного отсутствия до значительного снижения интеллекта и появления судорог. Около половины пациентов с туберозным склерозом соответствуют критериям PAC (Allingham-Hawkins et al., 1999).

Туберозный склероз вызывают мутации в двух генах, *TSC1* и *TSC2*, причем значительная доля случаев приходится на спорадические мутации (Curatolo et al., 2008). Продукты генов *TSC1* и *TSC2*, гамартин и туберин соответственно, формируют гетеродимерный комплекс, который является негативным регулятором mTOR. В отсутствие ингибирования со стороны TSC1/2 комплекса mTOR чрезмерно стимулирует клеточный рост и пролиферацию. Мутации как *TSC1*, так и *TSC2* гена в гомозиготном состоянии летальны уже на эмбриональной стадии развития организма.

В экспериментах на модельных животных было показано, что PAC и снижение интеллекта у больных туберозным склерозом, скорее всего, не связаны с эпилепсией и формированием доброкачественных опухолей. Особенно интересный фенотип был обнаружен у мышей *Tsc2^{+/-}*, которые продемонстрировали нарушения синаптической пластичности и поведения при отсутствии судорог и избыточного разрастания нервной ткани (Ehninger et al., 2009). У данных мышей была усилена поздняя фаза длительной потенциации, как известно, требующая интенсивного белкового синтеза, но именно повышение уровня биосинтеза белка является отражением повышенной активности mTOR.

У всех модельных линий мышей, мутантных по *TSC1* или *TSC2*, поведенческие и молекулярные нарушения устранялись после применения mTOR-ингибиторов, таких как эверолимус и сиролимус (рапамицин) (Meikle et al., 2008; Ehninger, Silva, 2011; Tsai, Sahin, 2011). Примечательно, что даже у взрослых мышей *Tsc2^{+/-}* после лечения рапамицином были отмечены как восстановление длительной потенциации, так и значительное улучшение

поведения и обучаемости (Ehninger et al., 2009), что предполагает обратимость поведенческих нарушений при аутизме. В настоящее время эверолимус одобрен Food and Drug Administration для лечения не-неврологических проявлений туберозного склероза, субэпендимальных гигантоклеточных астроцитом при невозможности выполнения хирургической резекции опухоли и ангиомиолипомы почки, не требующей срочного хирургического вмешательства. Проводятся испытания этого препарата для лечения нейрокогнитивных нарушений у детей (Ebrahimi-Fakhari, Sahin, 2015).

Синдром Каудена и PAC, связанные с мутациями PTEN

PTEN (phosphatase and tensin homolog deleted on chromosome 10) – липидная фосфатаза – негативный регулятор сигнального пути mTOR, расположенный выше комплекса туберозного склероза TSC1/2 (см. рис. 1). В настоящее время известна функция PTEN как супрессора опухолевого роста и показана роль инактивации этого фермента в канцерогенезе.

С мутациями в гене *PTEN* связывают синдром Каудена, иначе называемый синдромом множественных гамартом. Также такие мутации часто сопровождаются неврологическими проявлениями – макроцефалией, эпилепсией, снижением интеллекта и аутизмом.

Считается, что доля мутации *PTEN* составляет до 5 % в PAC, но она может быть и выше, поскольку, как правило, анализируются образцы крови пациентов методом полноэкзомного секвенирования, а при таком варианте анализа не учитываются мутации в некодирующих участках, включая промоторную область и районы ДНК, фланкирующие экзоны.

Интересно, что связанные с аутизмом мутации гена *PTEN* обычно не снижают липидно-фосфатазную активность фермента (Zhou, Parada, 2012). Такие мутационные изменения не приводят к формированию гамартом, и, как правило, это мутации *de novo*. Корреляция между *PTEN*-связанным аутизмом и канцерогенезом отсутствует (Zhou, Parada, 2012).

Эксперименты на модельных животных подтвердили роль PTEN в формировании PAC. Делеция гена *PTEN* приводила в нейронах мышей к увеличению размеров тела нейронов, гипертрофии аксонов и дендритов, и, как следствие, размеров синапсов. Социальное взаимодействие таких животных также было нарушено (Kwon et al., 2006).

В экспериментах с избирательно нокаутированным геном *PTEN* в нейронах слухового пути наблюдалось значительное усиление возбуждающего входного синаптического сигнала по сравнению с нейронами с нормальной экспрессией *PTEN* (Xiong et al., 2012). Дендриты *PTEN*-нокаутированных нейронов были длиннее нормы, и плотность расположения шипиков на них была повышена. Как морфологические, так и функциональные нарушения данных нейронов блокировались рапамицином, поэтому авторы сделали вывод, что все фенотипические проявления делеции *PTEN* в нейронах происходят в результате гиперактивации сигнального пути mTOR (Xiong et al., 2012).

Синдром Ретта и РАС, связанные с мутациями MECP2

Синдром Ретта обычно диагностируется у девочек после 6–18 мес нормального развития. Первыми проявлениями становятся бесцельные стереотипные движения руками, потеря речи, глазного контакта и выразительности лица. Практически все пациенты соответствуют критериям РАС, с отягощениями в виде нарушения дыхания, судорог, тахикардии, брадикардии и сколиоза.

Синдром Ретта (СР) связан с мутацией в гене *MECP2* (methyl-CpG-binding protein 2), локализованном на X-хромосоме. Для мальчиков null-мутация *MECP2* летальна, ее носители обычно не доживают до двух лет. В настоящее время описано достаточно много различных мутаций *MECP2*, в том числе и синдром Хq28, причиной которого является дупликация гена *MECP2*. Последствиями этих мутаций являются самые разнообразные нейropsychические заболевания: аутизм, эпилепсия, шизофрения, биполярное расстройство, болезнь Паркинсона, тремор и снижение интеллекта (Zoghbi, Bear, 2012).

MeCP2 – это белок, который локализован в ядре клеток и связывается с метилированным цитозином. Взаимодействуя с промоторными областями некоторых генов, включая *BDNF* (brain-derived neurotrophic factor), MeCP2 активирует их транскрипцию. Было продемонстрировано также, что MeCP2 конкурирует с гистон H1 за сайты связывания с ДНК, но молекулярные механизмы, приводящие к формированию фенотипов СР и Хq28, пока не выяснены (Ghosh et al., 2010).

Мыши, лишенные функционального MeCP2-белка, воспроизводили фенотип СР (Chen et al., 2001; Guy et al., 2001). Морфологически мозг таких мышей выглядел нормальным, за исключением микроцефалии, но дендритов у нейронов было меньше нормы и наблюдалось их распухание (Belichenko et al., 2009). На молекулярном уровне у модельных мышей MeCP2 было обнаружено чрезмерное ингибирование mTOR сигнального пути – гипофосфорилирование протеинкиназы В (Akt), mTOR киназы, рибосомного белка S6, p70S6K киназы и, как следствие, снижение биосинтеза белка (Ricciardi et al., 2011). Ген *BDNF* был среди первых изученных мишеней MeCP2. Белок BDNF связывается с рецептором тропомиозин-подобной киназы В (TrkB) и активирует целый ряд внутриклеточных сигнальных путей, включая mTOR (Segal, Greenberg, 1996). Считается, что снижение уровня белка и mРНК BDNF вносит основной вклад в патофизиологию синдрома Ретта (Katz, 2014).

Использование рекомбинантного BDNF в терапии СР осложнено низкой проницаемостью гематоэнцефалического барьера для этого белка, поэтому исследования сосредоточены на бустерах и миметиках BDNF. В настоящее время на стадии клинических испытаний находятся два бустера BDNF – финголимод и копаксон, оба препарата ранее были одобрены для лечения рассеянного склероза. Финголимод является модулятором сфингозин-1-фосфатных рецепторов, который приводит к увеличению экспрессии BDNF и активации TrkB-зависимых сигнальных путей (Deogracias et al., 2012). Копаксон известен как иммуномодулятор. Один из предполагаемых механизмов его действия – увеличение экспрессии и высво-

бождения BDNF из аутореактивных Т-клеток (Ziemssen et al., 2002).

Нарушение регуляции сигнального пути mTOR как связующее звено между поведенческими и иммунными проблемами, наблюдаемыми у пациентов с РАС

Центральная роль mTOR в интегрировании множества экстраклеточных и внутриклеточных сигналов объясняет сложную картину симптоматики аутизма у человека, которая часто включает аллергии и иммунные нарушения (Angelidou et al., 2011). Многие родители детей-аутистов замечают улучшения в поведении и развитии своих детей после перехода на диеты, исключающие основной белок многих злаков – глютен и молочный белок – казеин. Долгое время эти диеты не находили поддержки у врачей и исследователей аутизма, но в последнее время положение изменилось. Было показано, что сигнальный путь mTOR также связан с индуцируемыми пищевой аллергией поведенческими и иммунологическими нарушениями (Wu et al., 2015). У орально сенсibilизированных модельных мышей была вызвана аллергическая реакция на молочную сыворотку, которая приводила к снижению социального взаимодействия и нарастанию повторяющегося поведения. Активация mTOR была выявлена в кишечнике и в мозге данных модельных животных и, как и в случае с аутизмом, рапамицин снимал как поведенческие, так и иммунологические проявления аллергии на молочный белок (Wu et al., 2015). Таким образом, возможно, mTOR является тем сигнальным путем, который объединяет поведенческие и иммунологические проблемы РАС и, следовательно, может оказаться оптимальной мишенью для коррекции фенотипа РАС и всех коморбидных расстройств, например пищеварительных и иммунных.

Среди факторов среды, влияющих на формирование фенотипа РАС, часто упоминается витамин D3. Известно, что дефицит этого витамина приводит к увеличению размеров головного мозга и изменению его формы, подобным тем, которые наблюдаются у некоторых аутистов, а у детей с РАС часто диагностируется пониженный уровень витамина D3.

Предположение, что добавление в диету витамина D3 может улучшить состояние детей с РАС, было проверено в исследованиях F. Jia с коллегами (2015). У 32-месячного ребенка, принимавшего этот витамин в терапевтических дозах, было отмечено ослабление основных симптомов аутизма (Jia et al., 2015).

Механизм действия витамина D3 на симптоматику РАС также может быть связан с его способностью ингибировать mTOR посредством усиления экспрессии активируемого повреждениями ДНК транскрипта 4 (DNA damage-inducible transcript 4 (DDIT4)), который является сильным супрессором mTOR (Lisse et al., 2011). В целом ингибирование mTOR играет ключевую роль в адаптивном иммунитете, так как способствует формированию автофагосом, уничтожающих бактерии, и влияет на метаболизм и функции Т-лимфоцитов (Lisse, Hewison, 2011). Витамин D3 и рапамицин имеют общую мишень – mTOR, следовательно, необходимы более широкие исследования витамина D3 как, возможно, менее эффек-

тивного, но гораздо более доступного препарата для коррекции PAC.

Предположение о том, что прививки, в особенности АКДС (против коклюша, дифтерии и столбняка) в России и MMR (против кори, свинки и краснухи) в США вызывают аутизм, постоянно обсуждается на форумах родителей детей-аутистов. Многократные статистические исследования и даже эксперименты на приматах не подтвердили эти предположения (Gadad et al., 2015). Однако новые данные о мутациях в сигнальном пути mTOR, приводящих к PAC, и связь этого сигнального пути с иммунитетом выводят эту проблему на совершенно новый уровень. Действительно, иммунологи, занимающиеся проблемой Т-лимфоцит-обусловленного иммунитета, продемонстрировали (Liu et al., 2015), что в норме слегка ингибированный mTOR поддерживает гомеостаз Т-лимфоцитов. Однако распознавание антигена «необученным» Т-лимфоцитом приводит к немедленной активации mTOR и последующему дифференцированию этих клеток в определенный тип лимфоцитов-эффекторов (Liu et al., 2015). Возникает очевидный вопрос: как отразится на ребенке с генетически обусловленным гиперактивированным mTOR поливалентная иммунизация или тяжелое инфекционное заболевание в критические периоды развития ЦНС? Следовательно, необходимы дополнительные исследования о взаимосвязи иммунной системы, социального поведения и адаптивности животных с гиперактивированным mTOR.

Заключение

Синапсопатии, и в частности гиперактивация сигнального пути mTOR, являются общей характеристикой для постоянно растущего числа психических и неврологических заболеваний, таких как болезнь Альцгеймера (Pei, Hugon, 2008), эпилепсия (Wong, 2010), расстройства аутистического спектра и даже синдром Дауна (Trosa-Marín et al., 2012). Гипоактивация mTOR также приводит к снижению интеллекта и невропатологии, как было показано для синдрома Ретта. Успешное применение таких ингибиторов mTOR, как сиролimus и эверолимус, приводящее к значительному улучшению клинической картины на модельных животных и в ходе клинических испытаний на человеке, вселяет надежду на разработку адекватной механизм-обоснованной терапии расстройств аутистического спектра как связанных с нарушениями активности сигнального пути mTOR, так и обусловленных другими нарушениями структуры и функции синапсов.

Acknowledgments

This work was supported by State Budgeted Project 0324-2015-0004.

Conflict of interest

The authors declare no conflict of interest.

References

Allingham-Hawkins D.J., Babul-Hirji R., Chitayat D., Holden J.J., Yang K.T., Lee C., ... Costa S.S., Otto P.A., Mingroni-Netto R.C., Murray A., Webb J., Vieri F. Fragile X premutation is a significant risk factor for premature ovarian failure: The International Collaborative POF in Fragile X study-preliminary data. *Am. J. Med. Genet.*

1999;83(4):322-325. DOI 10.1002/(SICI)1096-8628(19990402)83:4<322::AID-AJMG17>3.0.CO;2-B [pii].

Angelidou A., Alysandratos K.-D., Asadi S., Zhang B., Francis K., Vasiadi M., ... Theoharides T.C. Brief Report: "Allergic Symptoms" in Children with Autism Spectrum Disorders. More than Meets the Eye? *J. Autism Dev. Disord.* 2011;41(11):1579-1585. DOI 10.1007/s10803-010-1171-z.

Bagni C., Oostra B.A. Fragile X syndrome: from protein function to therapy. *Am. J. Med. Genet. Part A.* 2013;161(11):2809-2821. DOI 10.1002/ajmg.a.36241.

Belichenko P.V., Wright E.E., Belichenko N.P., Masliah E., Li H.H., Mobley W.C., Francke U. Widespread changes in dendritic and axonal morphology in Mecp2-mutant mouse models of Rett syndrome: Evidence for disruption of neuronal networks. *J. Comp. Neurol.* 2009;514(3):240-258. DOI 10.1002/cne.22009.

Böckers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C., Buck F., ... Kreienkamp H.J. Synaptic scaffolding proteins in rat brain: Ankyrin repeats of the multidomain Shank protein family interact with the cytoskeletal protein α -fodrin. *J. Biol. Chem.* 2001;276(43):40104-40112. DOI 10.1074/jbc.M102454200.

Bozdagi O., Sakurai T., Papapetrou D., Wang X., Dickstein D.L., Takahashi N., ... Buxbaum J.D. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. *Mol. Autism.* 2010;1(1):15. DOI 10.1007/s10863-010-9115-1 [pii]. DOI 10.1007/s10863-010-9115-1.

Chang S., Bray S.M., Li Z., Zarnescu D.C., He C., Jin P., Warren S.T. Identification of small molecules rescuing fragile X syndrome phenotypes in *Drosophila*. *Nat. Chem. Biol.* 2008;4(4):256-263. DOI 10.1038/nchembio.78.

Chen R.Z., Akbarian S., Tudor M., Jaenisch R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* 2001;27(3):327-331. DOI 10.1038/85906.

Curatolo P., Bombardieri R., Jozwiak S. Tuberous sclerosis. *Lancet.* 2008;372(9639):657-668. DOI 10.1016/S0140-6736(08)61279-9.

Deogracias R., Yazdani M., Dekkers M.P.J., Guy J., Ionescu M.C.S., Vogt K.E., Barde Y.-A. Fingolimod, a sphingosine-1 phosphate receptor modulator, increases BDNF levels and improves symptoms of a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. USA.* 2012;109(35):14230-14235. DOI 10.1073/pnas.1206093109.

Dolan B.M., Duron S.G., Campbell D.A., Vollrath B., Shankaranarayana Rao B.S., Ko H.-Y., ... Tonegawa S. Rescue of fragile X syndrome phenotypes in *Fmr1* KO mice by the small-molecule PAK inhibitor FRAX486. *Proc. Natl. Acad. Sci. USA.* 2013;110(14):5671-5676. DOI 10.1073/pnas.1219383110.

Durand C.M., Betancur C., Böckers T.M., Bockmann J., Chaste P., Fauchereau F., ... Bourgeron T. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* 2007;39(1):25-27. DOI 10.1038/ng1933.

Ebrahimi-Fakhari D., Sahin M. Autism and the synapse: emerging mechanisms and mechanism-based therapies. *Curr. Opin. Neurol.* 2015;1(617):1-12. DOI 10.1097/WCO.0000000000000186.

Ehninger D., Han S., Shilyansky C., Zhou Y., Li W., David J. Reversal of learning deficits in a Tsc2^{-/-} mouse model of tuberous sclerosis. *Nat. Med.* 2009;14(8):843-848. DOI 10.1038/nm1788.Reversal.

Ehninger D., Silva A.J. Rapamycin for treating Tuberous sclerosis and Autism Spectrum disorders. *Trends Mol. Med.* 2011;17(2):78-87. DOI 10.1016/j.molmed.2010.10.002.

El-Fishawy P., State M.W. The genetics of autism: key issues, recent findings, and clinical implications. *Psychiatr. Clin. North Am.* 2010;33(1):83-105. DOI 10.1016/j.psc.2009.12.002.

Gadad B.S., Li W., Yazdani U., Grady S., Johnson T., Hammond J., ... German D.C. Administration of thimerosal-containing vaccines to infant rhesus macaques does not result in autism-like behavior or neuropathology. *Proc. Natl. Acad. Sci. USA.* 2015;112(40):12498-12503. DOI 10.1073/pnas.1500968112.

Ghosh R.P., Horowitz-Scherer R.A., Nikitina T., Shlyakhtenko L.S., Woodcock C.L. MeCP2 binds cooperatively to its substrate and

- competes with histone H1 for chromatin binding sites. *Mol. Cell Biol.* 2010;30(19):4656-4670. DOI 10.1128/MCB.00379-10.
- Greer P.L., Hanayama R., Bloodgood B.L., Mardinly A.R., Lipton D.M., Flavell S.W., ... Greenberg M.E. The Angelman syndrome protein Ube3A regulates synapse development by ubiquitinating arc. *Cell.* 2010;140(5):704-716. DOI 10.1016/j.cell.2010.01.026.
- Guy J., Hendrich B., Holmes M., Martin J.E., Bird A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* 2001;27(3):322-326. DOI 10.1038/85899.
- Hata Y., Butz S., Südhof T.C. CASK: a novel dlg/PSD95 homolog with an N-terminal calmodulin-dependent protein kinase domain identified by interaction with neuroligins. *J. Neurosci.* 1996;16(8):2488-2494. DOI papers2://publication/uuid/43B65FDC-BC39-4EC1-8178-47495149E5C1.
- Irie M., Hata Y., Takeuchi M., Ichtenko K., Toyoda A., Hirao K., ... Südhof T.C. Binding of neuroligins to PSD-95. *Science.* 1997;277:1511-1515. DOI 10.1126/science.277.5331.1511.
- Jamain S., Quach H., Betancur C., Råstam M., Colineaux C., Gillberg I.C., ... Bourgeron T. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.* 2003;34(1):27-29. DOI 10.1038/ng1136.
- Jamain S., Radyushkin K., Hammerschmidt K., Granon S., Boretius S., Varoqueaux F., ... Brose N. Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *Proc. Natl. Acad. Sci. USA.* 2008;105(5):1710-1715. DOI 10.1073/pnas.0711555105.
- Jia F., Wang B., Shan L., Xu Z., Staal W.G., Du L. Core symptoms autism improved after vitamin D supplementation. *Pediatrics.* 2015;135(1):e196-e198. DOI 10.1542/peds.2014.2121.
- Jiang Y.H., Armstrong D., Albrecht U., Atkins C.M., Noebels J.L., Eichele G., ... Beaudet A.L. Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. *Neuron.* 1998;21(4):799-811. DOI 10.1016/S0896-6273(00)80596-6.
- Katz D.M. Brain-derived neurotrophic factor and rett syndrome. *Handbook of Experimental Pharmacology.* Eds. G.R. Lewin, B.D. Carter. Berlin: Springer-Verlag, 2014;220:481-495. DOI 10.1007/978-3-642-45106-5_18.
- Kelleher R.J., Bear M.F. The autistic neuron: troubled translation? *Cell.* 2008;135(3):401-406. DOI 10.1016/j.cell.2008.10.017.
- Kishino T., Lalande M., Wagstaff J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat. Genet.* 1997;15(1):70-73. DOI 10.1038/ng0197-70.
- Kouser M., Speed H.E., Dewey C.M., Reimers J.M., Widman A.J., Gupta N., ... Powell C.M. Loss of predominant Shank3 isoforms results in hippocampus-dependent impairments in behavior and synaptic transmission. *J. Neurosci.* 2013;33(47):18448-18468. DOI 10.1523/JNEUROSCI.3017-13.2013.
- Kwon C.H., Luikart B.W., Powell C.M., Zhou J., Matheny S.A., Zhang W., ... Parada L.F. Pten regulates neuronal arborization social interaction in mice. *Neuron.* 2006;50(3):377-388. DOI 10.1016/j.neuron.2006.03.023.
- Lipton J.O., Sahin M. The neurology of mTOR. *Neuron.* 2014;84(2):275-291. DOI 10.1016/j.neuron.2014.09.034.
- Lisse T.S., Hewison M. Vitamin D: A new player in the world of mTOR signaling. *Cell Cycle.* 2011;10(12):1888-1889. DOI 10.4161/cc.10.12.15620.
- Lisse T.S., Liu T., Irmeler M., Beckers J., Chen H., Adams J.S., Hewison M. Gene targeting by the vitamin D response element binding protein reveals a role for vitamin D in osteoblast mTOR signaling. *FASEB J.* 2011;25(3):937-947. DOI 10.1096/fj.10-172577.
- Liu Y., Zhang D., Liu X. mTOR signaling in T cell immunity and autoimmunity. *Int. Rev. Immunol.* 2015;34(1):50-66. DOI 10.3109/08830185.2014.933957.
- Matsuura T., Sutcliffe J.S., Fang P., Galjaard R.J., Jiang Y.H., Benton C.S., ... Beaudet A.L. De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat. Genet.* 1997;15:74-77. DOI 10.1038/ng0197-74.
- Meikle L., Pollizzi K., Egnor A., Kramvis I., Lane H., Sahin M., Kwiatkowski D.J. Response of a neuronal model of tuberous sclerosis to mammalian target of rapamycin (mTOR) inhibitors: effects on mTORC1 and Akt signaling lead to improved survival and function. *J. Neurosci.* 2008;28(21):5422-5432. DOI 10.1523/JNEUROSCI.0955-08.2008.
- Missler M., Südhof T.C. Neuroligins: Three genes and 1001 products. *Trends Genet.* 1998;14(1):20-26. DOI 10.1016/S0168-9525(97)01324-3.
- Missler M., Zhang W., Rohlmann A., Kattenstroth G., Hammer R.E., Gottmann K., Südhof T.C. Alpha-neurexins couple Ca²⁺ channels to synaptic vesicle exocytosis. *Nature.* 2003;423:939-948. DOI 10.1038/nature01755.
- Okamoto N., Kubota T., Nakamura Y., Murakami R., Nishikubo T., Tanaka I., ... Uchino S. 22q13 microduplication in two patients with common clinical manifestations: A recognizable syndrome? *Am. J. Med. Genet. Part A.* 2007;143(23):2804-2809. DOI 10.1002/ajmg.a.31771.
- Pacey L.K.K., Heximer S.P., Hampson D.R. Increased GABA(B) receptor-mediated signaling reduces the susceptibility of fragile X knockout mice to audiogenic seizures. *Mol. Pharmacol.* 2009;76(1):18-24. DOI 10.1124/mol.109.056127.posits.
- Peça J., Feliciano C., Ting J.T., Wang W., Wells M.F., Venkattraman T.N., ... Feng G. Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature.* 2011;472(7344):437-442. DOI 10.1038/nature09965.
- Pei J.J., Hugon J. mTOR-dependent signalling in Alzheimer's disease. *J. Cell. Mol. Med.* 2008;12(6B):2525-2532. DOI 10.1111/j.1582-4934.2008.00509.x.
- Phelan M.C., Rogers R.C., Saul R.A., Stapleton G.A., Sweet K., McDermid H., ... Kelly D.P. 22Q13 deletion syndrome. *Am. J. Med. Genet.* 2001;101(2):91-99. DOI 10.1002/1096-8628(20010615)101:2<91::AID-AJMG1340>3.0.CO;2-C.
- Ricciardi S., Boggio E.M., Grosso S., Lonetti G., Forlani G., Stefanelli G., ... Broccoli V. Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. *Hum. Mol. Genet.* 2011;20(6):1182-1196. DOI 10.1093/hmg/ddq563.
- Riday T.T., Dankoski E.C., Krouse M.C., Fish E.W., Walsh P.L., Han J.E., ... Malanga C.J. Pathway-specific dopaminergic deficits in a mouse model of Angelman syndrome. *J. Clin. Invest.* 2012;122(12):4544-4554. DOI 10.1172/JCI61888.
- Roussignol G., Ango F., Romorini S., Tu J.C., Sala C., Worley P.F., ... Fagni L. Shank expression is sufficient to induce functional dendritic spine synapses in aspiny neurons. *J. Neurosci.* 2005;25(14):3560-3570. DOI 10.1523/JNEUROSCI.4354-04.2005.
- Sato A. mTOR, a potential target to treat autism spectrum disorder. *CNS Neurol. Disord. Drug Targets.* 2016;15(5):533-543. DOI 10.2174/1871527315666160413120638.
- Segal R.A., Greenberg M.E. Intracellular signaling pathways activated by neurotrophic factors. *Annu. Rev. Neurosci.* 1996;19:463-489. DOI 10.1146/annurev.ne.19.030196.002335.
- Shcheglovitov A., Shcheglovitova O., Yazawa M., Portmann T., Shu R., Sebastiano V., ... Dolmetsch R.E. SHANK3 and IGF1 restore synaptic deficits in neurons from 22q13 deletion syndrome patients. *Nature.* 2013;503(7475):267-271. DOI 10.1038/nature12618.
- Sheng M., Kim E. The Shank family of scaffold proteins. *J. Cell Sci.* 2000;113(1):1851-1856.
- Singh S.K., Eroglu C. Neuroligins provide molecular links between syndromic and nonsyndromic autism. *Sci. Signal.* 2013;6(283):re4. DOI 10.1126/scisignal.2004102.
- Südhof T.C. Neuroligins and neuroligins link synaptic function to cognitive disease. *Nature.* 2008;455(7215):903-911. DOI 10.1038/nature07456.
- Tabuchi K., Blundell J., Etherton M.R., Hammer R.E., Liu X., Powell C.M., Südhof T.C. A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science.* 2007;318(5847):71-76. DOI 10.1126/science.1146221.
- Troca-Marin J.A., Alves-Sampaio A., Montesinos M.L. Deregulated mTOR-mediated translation in intellectual disability. *Prog. Neurobiol.* 2012;96(2):268-282. DOI 10.1016/j.pneurobio.2012.01.005.

- Tsai P., Sahin M. Mechanisms of neurocognitive dysfunction and therapeutic considerations in tuberous sclerosis complex. *Curr. Opin. Neurol.* 2011;24(2):106-113. DOI 10.1097/WCO.0b013e32834451c4.
- Varoqueaux F., Aramuni G., Rawson R.L., Mohrmann R., Missler M., Gottmann K., ... Brose N. Neuroligins determine synapse maturation and function. *Neuron*. 2006;51(6):741-754. DOI 10.1016/j.neuron.2006.09.003.
- Veenstra-VanderWeele J., Blakely R.D. Networking in autism: leveraging genetic, biomarker and model system findings in the search for new treatments. *Neuropsychopharmacology*. 2012;37(1):196-212. DOI 10.1038/npp.2011.185.
- Wong M. Mammalian target of rapamycin (mTOR) inhibition as a potential antiepileptogenic therapy: From tuberous sclerosis to common acquired epilepsies. *Epilepsia*. 2010;51(1):27-36. DOI 10.1111/j.1528-1167.2009.02341.x.
- Wu J., De Theije C.M.G., Da Silva S.L., Van Der Horst H., Reiniers M.T.M., Broersen L.M., ... Kraneveld A.D. mTOR plays an important role in cow's milk allergy-associated behavioral and immunological deficits. *Neuropharmacology*. 2015;97:220-232. DOI 10.1016/j.neuropharm.2015.04.035.
- Xiong Q., Oviedo H.V., Trotman L.C., Zador A.M. PTEN regulation of local and long-range connections in mouse auditory cortex. *J. Neurosci.* 2012;32(5):1643-1652. DOI 10.1523/JNEUROSCI.4480-11.2012.
- Yoo H. Genetics of autism spectrum disorder: current status and possible clinical applications. *Exp. Neurobiol.* 2015;24(4):257-272. DOI 10.5607/en.2015.24.4.257.
- Zhou J., Parada L.F. PTEN signaling in autism spectrum disorders. *Curr. Opin. Neurobiol.* 2012;22(5):873-879. DOI 10.1016/j.conb.2012.05.004.
- Ziemssen T., Kämpfel T., Klinkert W.E.F., Neuhaus O., Hohlfeld R. Glatiramer acetate-specific T-helper 1- and 2-type cell lines produce BDNF: implications for multiple sclerosis therapy. Brain-derived neurotrophic factor. *Brain*. 2002;125(11):2381-2391. DOI 10.1093/brain/awf252.
- Zoghbi H.Y., Bear M.F. Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harb. Perspect. Biol.* 2012;4(3):a009886. DOI 10.1101/cshperspect.a009886.
- Zweier C., de Jong E.K., Zweier M., Orrico A., Ousager L.B., Collins A.L., ... Rauch A. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in *Drosophila*. *Am. J. Hum. Genet.* 2009;85(5):655-666. DOI 10.1016/j.ajhg.2009.10.004.

Алфавитный указатель авторов статей, опубликованных в журнале в 2016 г.

Абдуразакова З.Л. 5, 653
Абдурахмонов И.Ю. 5, 643
Абнизова И.И. 6, 746
Абугалиева А.И. 3, 303
Августинович Д.Ф. 2, 155
Адонина И.Г. 1, 44
Азарахш М. 4, 441
Айтназаров Р.Б. 6, 756
Алехина Т.А. 1, 28
Амелькина О.А. 2, 145
Андрушкевич О.М. 5, 723
Анисимова А.В. 3, 320
Анисимова М.В. 6, 918
Антонов Д.А. 6, 804
Антонов Е.В. 6, 954
Антонова О.Ю. 5, 596
Антонцева Е.В. 6, 894
Арбузова В.С. 3, 355
Афонников Д.А. 6, 863
Ахметова К.А. 1, 65
Ачилов С.Г. 5, 643

Бабенко В.Н. 6, 804
Бабина А.В. 2, 234
Бадаева Е.Д. 1, 44
Бады-Хоо М.С. 1, 7
Бажан Н.М. 2, 138
Базарова Н.Э. 1, 125
Байборodin С.И. 1, 108; 5, 716, 723
Банникова С.В. 6, 869
Баранова О.А. 3, 320
Барашков Н.А. 1, 7
Баричева Э.М. 6, 762
Баркова О.Ю. 1, 34
Баттулин Н.Р. 4, 493; 6, 909, 918, 925
Батурина Г.С. 2, 234
Башева Е.А. 4, 408
Белан И.А. 3, 370
Белогрудова И. 4, 537
Белоусова В.П. 3, 348
Белый А.А. 4, 426
Беньковская Г.В. 1, 78
Берман Д.И. 1, 125
Блинова Е.В. 3, 286; 5, 636
Бобохужаев Ш.У. 5, 643
Богачев С.С. 1, 96, 108; 5, 716, 723
Бойко Н.И. 3, 277
Болоболова Е.У. 1, 65
Боме Н.А. 5, 629, 666
Бондарева Э.А. 1, 23
Бондарь Н.П. 2, 155
Бородин П.М. 4, 408
Брусенцов И.И. 6, 762, 797
Брызгалов Л.О. 6, 797

Брянская А.В. 6, 869
Будашкина Е.Б. 3, 311
Булахова Н.А. 1, 125
Булыгина В.В. 2, 212
Бухарина Т.А. 6, 832
Быков Р.А. 6, 899

Ваганова О.Ф. 3, 320
Вакула С.И. 1, 51
Ватанабе Н. 1, 57
Ваулин О.В. 5, 695
Верещагин Е.И. 1, 96, 108; 5, 716, 723
Витяев Е.Е. 6, 770
Вишневская Г.Б. 2, 155
Воевода М.И. 6, 756, 804, 887
Волкова Г.В. 3, 320
Волкова О.А. 6, 779
Воробьева О.В. 2, 262
Воронина Е.Н. 1, 16

Габдуллин М.А. 5, 666
Гавриленко Т.А. 5, 596
Гаврилова О.П. 3, 286
Гагкаева Т.Ю. 3, 286
Гайнуллин Н.Р. 3, 320
Галиева Э.Р. 6, 746, 770
Галингер Д.Н. 3, 320
Генаев М.А. 6, 863
Гербек Ю.Э. 2, 145, 220
Герлинская Л.А. 6, 918, 925
Глаголева А.Ю. 6, 823
Гладкова Е.В. 3, 320
Глазова М.В. 2, 204
Година Е.З. 1, 23
Голованова Е.В. 1, 125
Гомес Фернандес М.Х. 4, 408
Гончаренко Г.М. 5, 584
Гончарова А.В. 3, 344
Горячковская Т.Н. 6, 869, 876
Грауда Д. 4, 537
Гриневич В.В. 2, 220
Гринкевич Л.Н. 2, 262
Грунтенко Н.Е. 6, 959
Гулевич Р.Г. 2, 145, 165, 220
Гуляева Н.В. 2, 191

Дейкин А.В. 5, 576
Деменков П.С. 6, 840
Дементьева Н.В. 1, 39
Демидов Е.А. 6, 869
Демидова Е.В. 6, 876
Денисова Г.А. 5, 571
Дергилев А.И. 6, 770
Деренко М.В. 5, 571

Дзюбенко Н.И. 5, 636
Дидоренко С.В. 3, 303
Добровольская О.Б. 3, 355
Додуева И.Е. 4, 441
Долгова Е.В. 1, 96, 108; 5, 716, 723
Дорогова Н.В. 1, 65
Дорофеева Н.А. 2, 204
Дорошков А.В. 6, 863
Дрозд У.С. 2, 255
Дубатолов В.В. 6, 899
Дубатолова Т.Д. 5, 723
Дубинина А.Д. 6, 954
Дубовец Н.И. 3, 378
Дыгало Н.Н. 2, 198, 212, 247, 255

Емцева М.В. 3, 295
Ермаков А.А. 3, 386
Ермолаев В.И. 5, 584
Ефимов В.М. 6, 816
Ефремов Я.Р. 1, 96, 108; 5, 716, 723
Ефремова Т.Т. 3, 355

Жагата К. 4, 537
Жечев Д.А. 6, 857

Забара Ю.М. 1, 51
Запара Т.А. 2, 255
Запорожец В.И. 5, 584
Захаренко Л.П. 1, 72
Захаров И.К. 1, 72
Зелена Д. 2, 228
Землянская Е.В. 3, 386; 4, 459
Зосен Д.В. 2, 204
Зыцарь М.В. 1, 7

Иванисенко В.А. 6, 840, 857
Иванисенко Н.В. 6, 840, 857
Иванисенко Т.В. 6, 840;
Иванова Л.Н. 2, 234; 6, 954
Иванощук Д.Е. 6, 840
Илинский Ю.Ю. 6, 899
Ильницкая Е.Т. 4, 528
Ильницкая С.И. 5, 708
Ильчибаева Т.В. 2, 172
Исакова Ж.Т. 6, 804

Каган А. 2, 165
Каледин В.И. 5, 708
Калинина Т.С. 2, 212
Карагодин Д.А. 6, 762
Каткова Л.Е. 2, 234
Катохин А.В. 2, 155
Кашина Е.В. 2, 155; 6, 894
Кильчевский А.В. 4, 482
Кирильчук Р.С. 6, 756
Китаев К.А. 1, 78
Клименко Н.С. 5, 596
Князев С.П. 5, 584; 6, 756
Кобзев В.Ф. 6, 887
Ковалев В.С. 3, 333

Ковалева В.Ю. 6, 816
Коваленко Д.В. 5, 576
Коваленкова М.В. 4, 403
Ковригин И.И. 6, 918
Кожемякина Р.В. 1, 28; 2, 145, 165
Козел А.В. 1, 96, 108
Колосова И.В. 6, 949
Колчанов Н.А. 1, 96, 108; 4, 459; 5, 716, 723; 6, 840
Конарев А.В. 5, 636
Кондратьева И.В. 3, 277
Кондрахин Ю.В. 6, 779
Коновалова Г.С. 5, 616
Коновалова Н.А. 6, 840
Коновалова О.С. 6, 840
Коношенко М.Ю. 2, 145
Константинова С.Г. 6, 869
Концевая Г.В. 5, 704; 6, 918, 925
Коняев С.В. 1, 125
Кораблев А.Н. 6, 930
Костина Л.И. 5, 596
Кочетов А.В. 4, 475; 6, 904
Кочнев Н.Н. 6, 756
Кравцова Л.А. 3, 370
Красневская Н. 4, 537
Криворучко А.Ю. 5, 576
Кругова Т.М. 1, 125
Кузмицкая П.В. 5, 673
Кулакова Е.В. 6, 770

Лаврик И.Н. 6, 840
Лазарева Е.Н. 3, 320
Лазарева Н.А. 2, 191
Ланка Г. 4, 537
Ланшаков Д.А. 2, 255
Лапочкина И.Ф. 1, 57; 3, 320
Лашин С.А. 1, 7
Лебедева М.А. 4, 441
Леберфарб Е.Ю. 6, 797
Леонова И.Н. 1, 44; 3, 311
Леонтьева Т.Л. 1, 78
Лепехов С.Б. 3, 337
Лисина Н. 4, 537
Литвинов А.Н. 5, 571
Литвинов Ю.Н. 6, 816
Литвинова Е.А. 5, 704
Личман Д.В. 1, 16
Логинова Д.Б. 1, 87
Лоскутов И.Г. 3, 286; 5, 636
Лукиянчикова В.А. 4, 493; 6, 930
Лутова Л.А. 4, 441
Лыжин А.С. 3, 329
Львова М.Н. 2, 155
Любечанский И.И. 1, 125

Майкова О.О. 4, 403
Макамов А.Х. 5, 643
Макаркина М.В. 4, 528
Макарова Е.Н. 2, 138
Максютов Р.А. 6, 949
Мальчевский В.А. 5, 666

Малярчук Б.А. 5, 571
Марданшин И.С. 1, 78
Маркель А.Л. 6, 954
Мартинек П. 3, 355
Мензоров А.Г. 6, 925, 930, 945
Меньшанов П.Н. 2, 247
Меркулова Т.И. 6, 762, 797
Мещерякова И.А. 6, 869
Микельсоне А. 4, 537
Минкевич А.М. 1, 96, 108; 5, 716, 723
Мироль П. 4, 408
Миронова В.В. 3, 386
Митрофанова О.В. 1, 39
Митрофанова О.П. 4, 545
Михальская В.Ю. 1, 7
Мордвинов В.А. 2, 155; 6, 894
Морозкова Т.С. 5, 708
Москалев А.А. 4, 426
Мошкин М.П. 5, 704; 6, 918

Нарушко М.В. 5, 666
Науменко В.С. 2, 180
Науменко Ф.М. 6, 746, 770
Недолужко А.И. 5, 658
Никитин С.В. 5, 584
Никитина Л.С. 2, 204
Николин В.П. 1, 96, 108; 5, 716, 723
Новиков Ю.М. 5, 695
Новикова Л.Ю. 5, 596
Новикова Т.И. 3, 348

Овчинникова Л.П. 5, 708
Омельянчук Н.А. 3, 386; 4, 459
Омигов В.В. 1, 96, 108
Онуфриев М.В. 2, 191
Орлов Ю.Л. 6, 746, 770, 804
Орлова Н.Г. 6, 746
Орловская И.А. 2, 155
Орницанс Р. 4, 537
Осадчая Т.С. 3, 370
Осипова Л.П. 1, 16
Останин А.А. 1, 96, 108; 5, 716, 723
Оськина И.Н. 2, 145
Отмахова Ю.С. 4, 511
Ощепков Д.Ю. 6, 869, 894
Ощепкова Е.А. 4, 459

Пальчикова Н.А. 1, 28
Пастернак Т. 4, 459
Пельтек С.Е. 1, 125; 6, 869, 876
Перегуд Д.И. 2, 191
Першина Л.А. 3, 370; 4, 416
Петров Д.Б. 1, 96, 108
Петров С.А. 5, 666
Петрова Д.Д. 5, 723
Печковская Т.В. 1, 51
Пивоварова Е.Н. 6, 954
Пилипенко А.С. 6, 887
Пилипенко И.В. 6, 887
Пискарев В.В. 3, 277

Плюснина И.З. 2, 145
Подколотная О.А. 6, 840, 848
Подколотный Н.Л. 6, 848
Понимаскин Е.Г. 2, 180
Пономаренко М.П. 6, 787
Пономаренко П.М. 6, 787
Попик В.М. 6, 869, 876
Попова Н.А. 1, 96, 108; 5, 708, 716; 723
Попова Н.К. 2, 172, 180
Посух О.Л. 1, 7
Поттер Е.А. 1, 96, 108; 5, 716, 723
Присяжнюк И.Е. 6, 925
Присяжнюк М.С. 6, 887
Причко Т.Г. 5, 607
Прокудина О.И. 1, 28
Проскурина А.С. 1, 96, 108; 5, 716, 723
Прошкина Е.Н. 4, 426
Пшеничникова Т.А. 1, 57; 4, 511

Радченко Е.Е. 5, 616
Разумов И.А. 6, 949
Рассказов Д.А. 6, 787
Рашаль И. 4, 537
Рипбергер Е.И. 5, 629
Рогачев В.А. 1, 96, 108; 5, 716, 723
Романов Г.П. 1, 7
Романова И.В. 2, 138
Росеева Л.П. 3, 370

Савельев Н.И. 3, 329; 4, 555
Савельева Н.Н. 3, 329; 4, 555
Савенко Е.Г. 3, 333
Савина М.С. 4, 459
Савинкова Л.К. 6, 787
Сайк О.В. 6, 840
Салина Е.А. 1, 44
Санамьян М.Ф. 5, 643
Свичкарев А.В. 6, 770
Селионова М.И. 5, 576
Семенов А.И. 6, 869
Серов О.Л. 6, 945
Серова И.А. 6, 918, 930
Сидорова К.К. 3, 344, 348
Силкова О.Г. 1, 87
Симонов А.В. 1, 57
Смарагдов М.Г. 1, 34
Смирнов А.В. 4, 493; 6, 918, 925
Смирнова О.Г. 6, 904
Соболева О.Н. 5, 616
Соловьев И.А. 4, 426
Сохраняева Л.С. 2, 191
Спицина А.М. 6, 746, 770
Степаничев М.Ю. 2, 191
Стёпочкин П.И. 3, 295
Страздиня В. 4, 537
Субботин А.М. 5, 666
Супрун И.И. 3, 333
Сурина Е.В. 1, 78
Сухарева Е.В. 2, 212
Сычева Е.А. 3, 378

Табиханова Л.Э. 1, 16
Таранов О.С. 1, 96, 108; 5, 716, 723
Твердохлеб Н.Н. 6, 848
Творогова В.Е. 4, 441
те Боекхорст Р. 6, 746
Терлецкий В.П. 1, 39
Тийс Р.П. 1, 16
Тишкина А.О. 2, 191
Топоркова Л.Б. 2, 155
Торгашева А.А. 4, 408
Траутц Д. 5, 629
Трегубчак Т.В. 6, 949
Трифонов Е.А. 6, 959
Трубачеева Н.В. 3, 370; 4, 416
Трухачев В.И. 5, 576
Тыщенко В.И. 1, 39

Удалов М.Б. 1, 78
Ульяновская Е.В. 5, 607
Урбанович О.Ю. 5, 673
Усенко Н.И. 4, 511

Федорова С.А. 1, 65
Федосеева Л.А. 6, 954
Феофанова Н.А. 2, 155; 6, 918, 925
Фетере В. 4, 537
Филипенко М.Л. 1, 16
Фишман В.С. 6, 930
Фокина О. 4, 537
Фотев Ю.В. 3, 348
Фрейман С.В. 2, 191
Фурман Д.П. 6, 832, 894

Хакимова А.Г. 4, 545;
Харченко Е.С. 3, 333
Хегай И.И. 2, 243
Хлебодарова Т.М. 6, 959
Хлесткина Е.К. 3, 364; 4, 511; 6, 823
Хотылева Л.В. 1, 51; 4, 482; 5, 683

Цыбко А.С. 2, 172

Чадаева И.В. 6, 746, 770, 787
Чалая Л.Д. 5, 607
Чен М. 6, 770
Черниговская Е.В. 2, 204
Черных Е.Р. 1, 96, 108; 5, 716, 723
Чесноков И.Н. 1, 65
Чуманова Е.В. 3, 355
Чуркина Т.В. 1, 16

Шаварда А.Л. 5, 636
Шаманин В.П. 3, 320
Шаманина М.Ю. 6, 894
Шаптуренко М.Н. 1, 51; 4, 482; 5, 683
Шарипов Р.Н. 6, 779
Шарыпова Е.Б. 6, 787
Шатохин К.С. 5, 584
Швачко Н.А. 5, 596
Шеленга Т.В. 5, 636
Шепелева Д.В. 2, 220
Шеховцов С.В. 1, 125
Шихевич С.Г. 2, 145, 165
Шишкина Г.Т. 2, 198
Шоева О.Ю. 6, 823
Шувалов О.Ю. 5, 596
Шувалова А.Р. 5, 596
Шумный В.К. 4, 475
Шуплецова О.Н. 5, 623

Щеглов М.А. 6, 869
Щелкунов С.Н. 6, 949
Щенникова И.Н. 5, 623
Щербаков Д.Ю. 4, 403

Юдин Н.С. 6, 756
Юдина М.А. 6, 899
Юдина Р.С. 3, 364
Юнусова А.М. 4, 493; 6, 909
Юнусханов Ш. 5, 653

Якимович А.В. 1, 51
Яковлев А.Ф. 1, 39

Прием статей через электронную редакцию на сайте <http://vavilov.elpub.ru/index.php/jour>
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции»/“Vavilov Journal of Genetics and Breeding”
до 2011 г. выходил под названием «Информационный вестник ВОГиС»/
“The Herald of Vavilov Society for Geneticists and Breeding Scientists”.

Регистрационное свидетельство ПИ № ФС77-45870 выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 20 июля 2011 г.

«Вавиловский журнал генетики и селекции» включен ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Российский индекс научного цитирования, ВИНТИ, базы данных Ebsco, DOAJ, Ulrich's Periodicals Directory, Google Scholar, Russian Science Citation Index на платформе Web of Science.

Открытый доступ к полным текстам:
на сайте ИЦИГ СО РАН – bionet.nsc.ru/vogis/
платформе Elpub – vavilov.elpub.ru/index.php/jour
платформе Научной электронной библиотеки – elibrary.ru/title_about.asp?id=32440

Подписку на «Вавиловский журнал генетики и селекции» можно оформить в любом почтовом отделении России. Индекс издания 42153 по каталогу «Пресса России».

При перепечатке материалов ссылка на журнал обязательна.

✉ e-mail: vavilov_journal@bionet.nsc.ru

Проспект Академика Лаврентьева, 10, Новосибирск, 630090.

Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.

Издание подготовлено информационно-издательским отделом ИЦИГ СО РАН. Тел.: (383)3634963*5218.

Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.

Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Подписано в печать 30.12.2016. Формат 60 × 84 ¹/₈. Усл. печ. л. 27,43. Уч.-изд. л. 30,8

Тираж 200 экз. Заказ № 359

Отпечатано в типографии ФГУП «Издательство СО РАН», Морской проспект, 2, Новосибирск, 630090.