2019 • 23 • 2

Print ISSN 2500-0462
Online ISSN 2500-3259

Научный рецензируемый журнал

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.
Периодичность 8 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

2019 • 23 • 2

Print ISSN 2500-0462
Online ISSN 2500-3259

Scientific Peer Reviewed Journal

VAVILOV JOURNAL OF GENETICS AND BREEDING

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

Founded in 1997
Published 8 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

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

СОДЕРЖАНИЕ • 2019 • 23 • 2

Эволюция генома

Медицинская генетика

VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2019 • 23 • 2

Genome evolution

Medical genetics

ХРОМОСОМА 2018

В августе 2018 г. в новосибирском Академгородке прошла организованная Институтом молекулярной и клеточной биологии СО РАН и Новосибирским государственным университетом международная конференция «Хромосома 2018», посвященная вопросам организации, функционирования и эволюции хромосом и геномов. Она продолжила серию конференций, начало которым было положено еще в 1968 г. Александрой Алексеевной Прокофьевой-Бельговской и которые стали своеобразной площадкой для общения цитогенетиков и молекулярных биологов. В 1990-е годы традиция хромосомных конференций была прервана, и возобновились они лишь в 2009 г., когда в Новосибирске была проведена «Хромосома 2009», а затем «Хромосома 2012» и «Хромосома 2015». Нынешняя конференция «Хромосома 2018» вновь собрала множество заинтересованных участников, что подтвердило необходимость научных встреч такого формата.

В работе конференции приняли участие 207 исследователей из 42 организаций одиннадцати стран (Россия, Казахстан, США, Великобритания, Италия, Германия, Бельгия, Швеция, Франция, и впервые ученые из Бразилии и Японии). География российских участников охватывала всю страну – от Калининграда до Владивостока. Заслушано 82 доклада, некоторые из них вошли в настоящий номер «Вавиловского журнала генетики и селекции». Для публикации были отобраны статьи по основным проблемам, обсуждавшимся на конференции.

Первый раздел посвящен проблематике организации интерфазных хромосом, хроматина и организации генома в составе хромосом других типов. В обзоре И.О. Боголюбовой с соавт. кратко обсуждаются особенности молекулярного состава гетерохроматина, ассоциированного с проядрышками эмбрионов мыши. Актуальным вопросам организации центромерного гистона H3 на примере ржи, которая входит в группу экономически важных злаков, посвящено оригинальное исследование С.С. Гацкой и Е.В. Евтушенко «Паттерны нуклеотидного разнообразия различных доменов гена центромерного гистона H3 (CENH3) у *Secale* L.». В обзорной статье Т.Д. Колесниковой с соавт. «Пространственно-временная организация репликации у дрозофилы и ее особенности в политенных хромосомах» анализируются последние данные о процессах репликации в интерфазных политенных хромосомах, а также рассмотрены закономерности в организации процесса репликации и состояний хроматина. Из двух докладов об организации политенных хромосом (И.Ф. Жимулев и Т.Ю. Зыкова), а также постера Д.С. Сидоренко сформирован один общий обзор Д.С. Сидоренко с соавт. («Политенные хромосомы отражают функциональную организацию генома *Drosophila*»), в котором сведены в единую модель данные об организации генома на генном и хромосомном уровне. В статье М.Б. Шварц (Беркаевой) с соавт. «Хроматиновые белки ADF1 и BEAF-32 влияют на позиционирование нуклеосом и упаковку ДНК междиска 61C7/C8 политенных хромосом *Drosophila melanogaster*» продолжены исследования механизмов дифференцированности структуры хроматина в интерфазных хромосомах.

В следующем разделе представлены сведения, относящиеся к наиболее активно разрабатываемой в настоящее время области эпигенетических исследований. Обзорная статья Ю.А. Ильиной и А.Ю. Конева посвящена последним данным о функциях АТФ-зависимых хроматин-ремоделирующих факторов в сборке нуклеосом, ремоделировании хроматина и обмене гистонов *in vivo*. Обсуждается роль этих белков (совместно с гипероновыми шаперонами и модифицирующими хроматин ферментами) в образовании сети факторов, обеспечивающих поддержание целостности хроматина. В обзоре Л.С. Мельниковой с соавт. «Функциональные свойства Su(Hw)-зависимого комплекса определяются его регуляторным окружением и множественными взаимодействиями на белковой платформе Su(Hw)» установлено, как множественные взаимодействия между основными компонентами Su(Hw)-зависимого комплекса (Su(Hw)/Mod(mdg4)-67.2/CP190) влияют на его активность. И.С. Осадчий с соавт. в статье «Идентификация белков, участвующих в привлечении Ttk69 к геномным сайтам *Drosophila melanogaster*» устанавливают, что белок Ttk69 способен одновременно взаимодействовать с MEP1 и белками с «цинковыми пальцами». Вопросы регуляции генной активности в составе хроматина рассмотрены в работах Е.Н. Набирочкиной с соавт. («Роль SAGA комплекса в транскрипции и экспорте мРНК») и А.А. Шейнова с соавт. («Функции изоформ PHF10 – субъединицы PBAF-комплекса, ремоделирующего хроматин»).

Серия статей посвящена вопросам генетической регуляции на уровне генов и геномов. В частности, Е.Н. Андреевой с соавт. впервые охарактеризована функция гена *Non3* (*Novel nucleolar protein 3*) дрозофилы, который кодирует гомолог Brix домен-содержащего белка человека Rpf2, участвующего в процессинге рибосомной РНК. В обзоре О.В. Андреенкова и др. «Исследования регуляторной зоны гена *Notch* у *Drosophila melanogaster* с использованием новых подходов направленного геномного редактирования» предложены первые данные о получении хромосомных перестроек в регуляторной области гена *Notch* и обратной трансформации отредактированного фрагмента в геном организма. М.О. Лебедев с соавт. представили работу «Создание штрихкодированных плазмидных библиотек для множественного одновременного анализа эффекта по-

ложения гена», в которой они сконструировали и протестировали качество штрихкодированных плазмидных библиотек, которые можно использовать для высокопроизводительного анализа эффекта положения промоторов разных генов в клетках дрозофилы. М.Ю. Мазина и Н.Е. Воробьева подготовили обзор на тему «Механизмы регуляции транскрипции под действием экдизона». Е.С. Омелина и А.В. Пиндюрин посвятили свой обзор «Оптогенетическая регуляция транскрипции эндогенных генов млекопитающих» обсуждению преимуществ и недостатков систем, основанных на использовании белков типа «цинковый палец», TALEs и технологии CRISPR/Cas9, для оптогенетической регуляции активности генов.

В двух последних небольших разделах представлены работы, в которых затронуты отдельные аспекты эволюции геномов и медицинской генетики. В статьях Н.Г. Андреенковой с соавт. «Проблемы генетического обоснования выделения подвидов черного коршуна (*Milvus migrans*)» и А.П. Крюкова «Филогеография и гибридизация врановых птиц Палеарктики» рассмотрены возможности использования маркёров митохондриальной ДНК в систематике и генетике популяций у птиц. В.Г. Тамбовцева с соавт. изучили особенности мейоза у экспериментальных гибридов слепушонок *Ellobius tancrei* (Mammalia, Rodentia) и сделали предположение, что мейотический драйв является основным механизмом, обеспечивающим диверсификацию и быструю фиксацию возникающих хромосомных форм в природе.

Н.А. Скрябин с соавт. в своем оригинальном исследовании «Районы гомозиготности в тканях абортусов из семей с привычным невынашиванием беременности» рассмотрели вопросы о связи существования протяженных гомозиготных районов с этиологией привычной невынашиваемости беременности у пациенток с нормальным кариотипом и сделали предположение, что одним из механизмов реализации патологического проявления могут являться аномалии импринтинга определенных генов.

И.Ф. Жимулёв, д-р биол. наук, академик РАН,
С.А. Демаков, д-р биол. наук, директор
Института молекулярной и клеточной биологии СО РАН, Новосибирск

Вавиловский журнал генетики и селекции. 2019;23(2):129-134
DOI 10.18699/VJ19.471

Peculiarities of the molecular composition of heterochromatin associated with pronucleoli in mouse embryos

I.O. Bogolyubova[1] ✉, Z.K. Sailau[2], D.S. Bogolyubov[1]

[1] Institute of Cytology, RAS, St. Petersburg, Russia
[2] St. Petersburg State University, Russia
✉ e-mail: ibogol@mail.ru

The nucleus of pre-implantation mammalian embryos is characterized by peculiar structural organization. At the initial stages of cleavage, the nucleus of the embryo contains the so-called nucleolus precursor bodies (NPBs) or pronucleoli rather than functionally active nucleoli. The NPBs are fibrillar electron-dense structures inactive in RNA synthesis. The vast majority of NPBs are surrounded by a ring-shaped zone of transcriptionally inactive heterochromatin. Intriguingly, these zones contain not only tri-methylated histone H3K9me3 as an epigenetic mark of repressed chromatin but also acetylated histone H4K5ac, a well-known marker of active chromatin. Immunocytochemical data suggest that the molecular composition of this 'ring heterochromatin' in mouse embryos changes during the realization of embryonic genome activation events, as well as during artificial suppression of transcription. In zygotes, some factors of mRNA biogenesis including splicing factor SC35 (SRSF2) and basal transcription factor TFIID are detectable in the ring chromatin. At later stages of development, other nuclear proteins such as Y14, a core component of the exon-exon junction complex (EJC), as well as the proteins involved in chromatin remodeling (ATRX, Daxx) are also detectable in this area. A typical component of the 'ring heterochromatin' is actin. Anti-actin immunocytochemical labeling is most expressed at the two-cell cleavage stage after activation of the embryonic genome. Indicatively, the molecular composition of the 'ring heterochromatin' associated with different NPBs may differ significantly even in the same nucleus. This seems to reflect the functional heterogeneity of morphologically similar NPBs according to their competence to the process of nucleologenesis. Here, we discuss briefly some peculiarities of the molecular composition and possible functions of the NPB-associated heterochromatin in mouse early embryos.
Key words: heterochromatin; pre-implantation mouse embryos; immunocytochemistry.

For citation: Bogolyubova I.O., Sailau Z.K., Bogolyubov D.S. Peculiarities of the molecular composition of heterochromatin associated with pronucleoli in mouse embryos. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):129-134. DOI 10.18699/VJ19.471

Особенности молекулярного состава гетерохроматина, ассоциированного с проядрышками в эмбрионах мыши

И.О. Боголюбова[1] ✉, Ж.К. Сайлау[2], Д.С. Боголюбов[1]

[1] Институт цитологии Российской академии наук, Санкт-Петербург, Россия
[2] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
✉ e-mail: ibogol@mail.ru

Ядро доимплантационных эмбрионов млекопитающих характеризуется своеобразной структурной организацией. На начальных стадиях дробления вместо функционально активных ядрышек в ядре эмбриона присутствуют так называемые проядрышки – фибриллярные электронно-плотные структуры, неактивные в отношении синтеза РНК. Подавляющее большинство проядрышек окружено кольцеобразной зоной транскрипционно неактивного гетерохроматина, который, однако, содержит не только эпигенетическую метку репрессированного хроматина – триметилированный гистон H3K9me3, но и метку активного хроматина – ацетилированный гистон H4K5ac. По результатам непрямого иммуномечения, молекулярный состав кольцевого гетерохроматина в эмбрионах мыши изменяется в ходе реализации процессов активации эмбрионального генома, а также при искусственном подавлении транскрипционной активности. На стадии зиготы в составе кольцевого гетерохроматина выявляются некоторые факторы метаболизма мРНК, например SR-белок фактор сплайсинга SC35 (SRSF2) и базальный фактор транскрипции TFIID. На более поздних стадиях развития в этой области начинают выявляться другие ядерные белки, например Y14 – коровый компонент комплекса связи экзонов (EJC), а также белки, вовлеченные в ремоделинг хроматина – ATRX и Daxx. Типичным компонентом кольцевого гетерохроматина является актин, иммуноцитохимическое мечение которого наиболее выражено на двухклеточной стадии дробления, после активации эмбрионального генома. Характерно, что молекулярный состав гетерохроматина, ассоциированного с разными проядрыш-

ками в одном ядре, может различаться, что, возможно, отражает функциональную гетерогенность морфологически сходных проядрышек по их компетентности к процессу нуклеологенеза. В настоящем обзоре кратко обсуждаются некоторые особенности молекулярного состава гетерохроматина, ассоциированного с NPBs, и его возможные функции.
Ключевые слова: гетерохроматин; доимплантационные эмбрионы мыши; иммуноцитохимия.

Introduction

The combination of maternal and paternal gametes upon fertilization forms a totipotent embryo that will give rise to over 200 different cell types in a fully differentiated organism. The cardinal changes in the potency of cells to differentiate, as well as the integration of the parental genomes, are accompanied by pronounced rearrangements of the 3D-organization of the nucleus in zygotes and early embryos. Correspondingly, the structure and organization of chromatin regions during early embryonic development differs fundamentally compared to somatic cells, reflecting the peculiar plasticity and potency of the genomes (Burton, Torres-Padilla, 2010). Global rearrangements of chromatin and major specific changes in DNA methylation, histone modifications as well as the incorporation of histone variants (Li, 2002; reviewed in Burton, Torres-Padilla, 2010; Mason et al., 2012) are the crucial reprogramming processes during the early steps of mammalian pre-implantation development (Lanctôt et al., 2007). In this review, some aspects of heterochromatin organization during the beginning stages of mouse development are briefly discussed.

Nucleolus precursor bodies and surrounding heterochromatin are the unique nuclear structures of mammalian early embryos

The unique feature of the nucleus of early mammalian embryos is the absence of functionally active nucleoli at the initial stages of embryonic development. Instead, the nucleus contains prominent electron-dense structures of perfectly round shape, called the nucleolus precursor bodies (NPBs) or pronucleoli (Fig. 1). A number of early autoradiographical studies (Geuskens, Alexandre, 1984; Tesařík et al., 1986a, b; Kopecný et al., 1989; etc.) has shown that NPBs are functionally linked with transcriptionally active nucleoli that are gradually formed later at species-specific embryonic stages. Unlike typical nucleoli of somatic cells, the pronucleoli (NPBs) of embryos consist only of a finely fibrillar and densely packed material of still unknown nature and do not contain typical nucleolar structural constituents such as dense fibrillar and granular components.

The vast majority of NPBs are surrounded by a characteristic ring-like zone of heterochromatin that is stained intensely with DAPI (see Fig. 1, *a′*, *b′*). Significantly, these peculiar zones are enriched in centromeric and pericentric satellite DNA in male and female pronuclei of zygotes. Later, this centromeric and pericentric heterochromatin re-localizes from the NPB periphery to the nucleoplasm at the two-cell stage, forming new structures called the ‘pro-chromocenters’ (Martin et al., 2006; Probst et al., 2007). The pericentric heterochromatin is enriched in HP1 (Heterochromatin Protein-1), and the HP1β isoform is prevalent in the NPB-surrounding heterochromatin in both male and female pronuclei of zygotes. HP1β is constitutively found in these heterochromatin areas until the blastocyst stage (Meglicki et al., 2012). In the mouse embryo, NPB-associated heterochromatin formation was found to require the histone variant H3.3, in particular di- and tri-methylation of lysine 27 residues (Santenard et al., 2010).

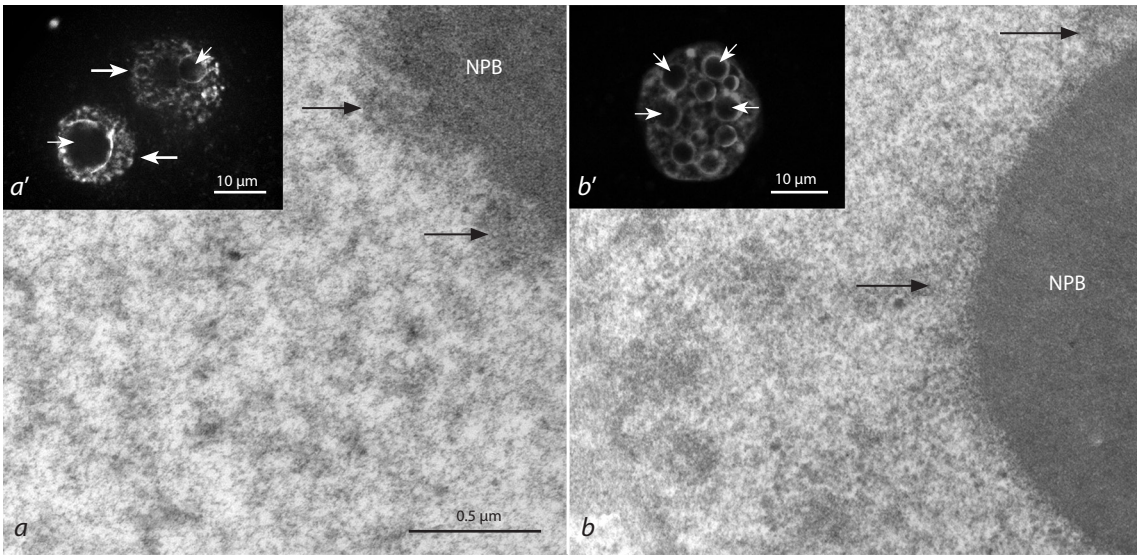


Fig. 1. Morphology of mouse zygotic pronuclei (*a*) and two-cell embryo nucleus (*b*).

a, *b* – electron microscopy; *a′*, *b′* – DAPI staining; NPB – pronucleolus. Large white arrows in *a′* indicate male (on the right) and female (on the left) pronuclei; small white arrows in *a′* and *b′* indicate some NPBs. Black arrows in *a* and *b* show heterochromatin clumps at the periphery of NPBs.

a, *b* – from (Bogolyubova, Bogolyubov, 2013), open access; *a′*, *b′* – from (Sailau et al., 2017), with permission from Elsevier.

H3.3 is associated with paternal pericentric heterochromatin during the first S-phase and plays a role in the transcription of pericentric repeats and subsequent tethering of HP1β.

NPB-surrounding heterochromatin areas contains the epigenetic marks of both transcriptionally repressed and active chromatin

The NPB-associated ring-like heterochromatin structures do not incorporate BrUTP (Fig. 2, *a*), suggesting their transcriptionally inert state (Bogolyubova, 2011; Bogolyubova, Bogolyubov, 2018). Surprisingly, these NPB-associated heterochromatin areas contain the markers of both repressed and active chromatin, e. g., the modified histones H3K9me3 and H4K5ac (see Fig. 2, *b*, *c*).

H3K9me3 domains are linked with transcriptional regulation and intracellular functions in pre-implantation mouse embryos. These domains are formed temporarily after fertilization, then increase dramatically in the number at the two-cell stage and diminish after the morula stage (Wang et al., 2018). A comprehensive analysis of H3K9me3-dependent heterochromatin dynamics in pre-implantation mouse embryos has shown that this heterochromatin undergoes the dramatic reprogramming during early embryonic development. Intriguingly, H3K9me3 domains are highly enriched in long terminal repeats (LTRs) that are hypomethylated and transcribed (Wang et al., 2018). Besides, RNA sequencing throughout early mouse embryogenesis revealed that expression of the repetitive-elements including LINE-1 and IAP retrotransposons is dynamic and stage specific, with most repetitive elements becoming repressed before implantation (Fadloun et al., 2013).

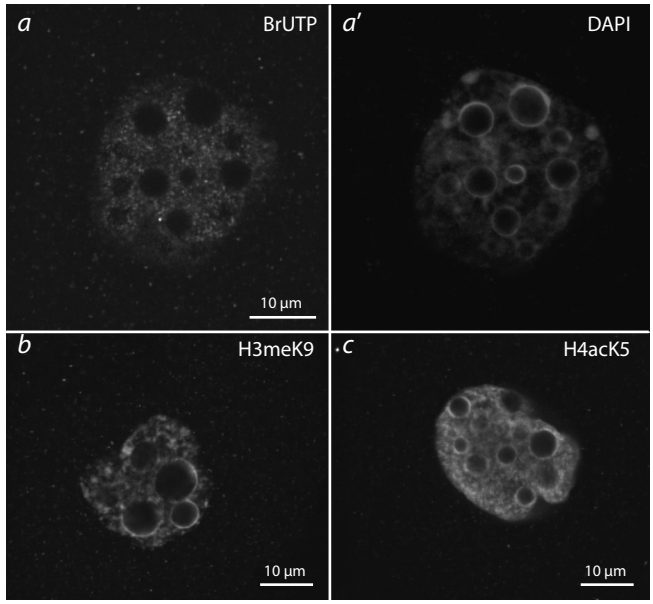


Fig. 2. Localization of BrUTP incorporation sites (*a*), H3K9me3 (*b*) and H4K5ac (*c*) in the nuclei of late two-cell mouse embryos.

Note: the ring-like chromatin surrounding NPBs does not incorporate BrUTP but contains both H3K9me3 and H4K5ac. Original.

Molecular composition dynamics of NPB-surrounding heterochromatin

The data obtain by indirect immunofluorescent microscopy (Table) suggest that the molecular composition of the NPB-surrounding heterochromatin including some chromatin-associated proteins, changes significantly during realization of the major events of embryonic genome activation as well as after artificial suppression of transcription. For example, some factors involved in mRNA metabolism including the spicing factor SC35 and basal transcription factor TFIID are revealed in the NPB-associated chromatin at transcriptionally inert stages (Bogolyubova, Bogolyubov, 2013) (Fig. 3). At later stages, however, other nuclear proteins (e. g., Y14, a core component of the exon-exon junction complex (EJC), the

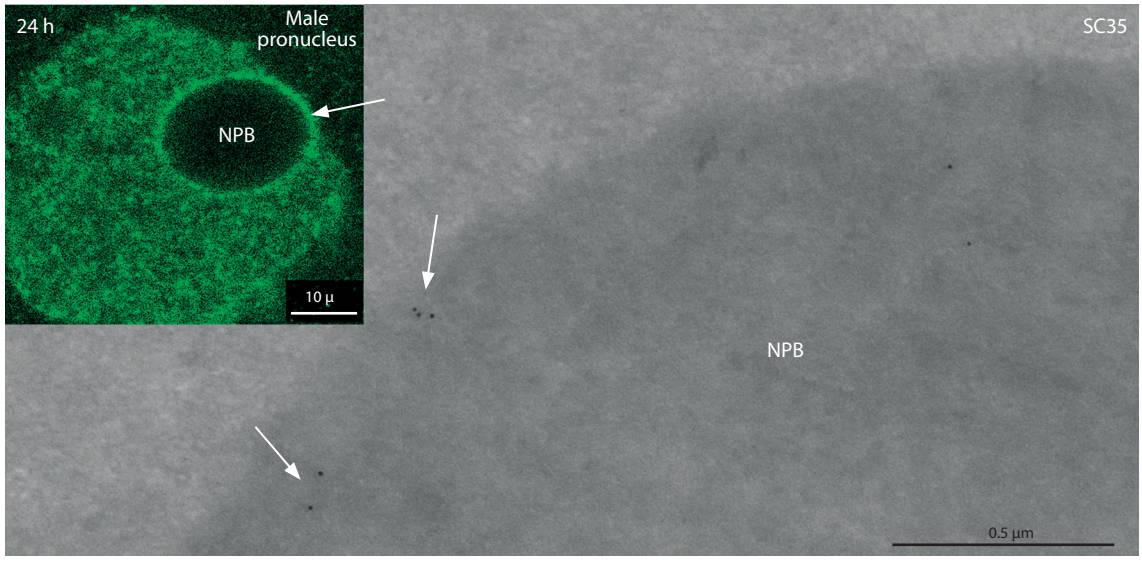


Fig. 3. Immunolocalization of the S/R-rich splicing factor SC35/SRSF2 in NPB-associated chromatin at an early stage (transcriptionally inert, 24 h post-hGG) of mouse development.

Arrows indicate anti-SC35 labeling in the vicinity of NPBs. Original; inset is from (Bogolyubova, Bogolyubov, 2013), open access.

Immunocytochemical localization of some nuclear antigens in the NPB-associated heterochromatin in early mouse embryos

Protein studied	Early zygote (20–24 h)		Late zygote (27–28 h)		Late two-cell stage (46–48 h)	Artificial inhibition of transcription
	f	m	f	m		
H3K9me3	++	–/+	++	–/+	+/++	N/A
H4K5ac	–	+	–	+	+/++	N/A
Actin	–	+	–	+	++	↑
SC35 splicing factor	+	+	–	–	–	–
Basal transcription factor TFIID	+	+	–	–	–	–
Chromatin-remodeling protein ATRX	–	–	+	+	++	+
Daxx, a chaperone of the modified histone H3K9me3	+	+	++	++	++	N/A
Y14, an EJC core component	–	–	–	–	+	↑

Note: "–" – no labeling; "+" – weak labeling; "++" – strong labeling, "↑"– increased labeling intensity and/or the appearance of additional zones of labeling; N/A – not available; f – female pronucleus; m – male pronucleus.

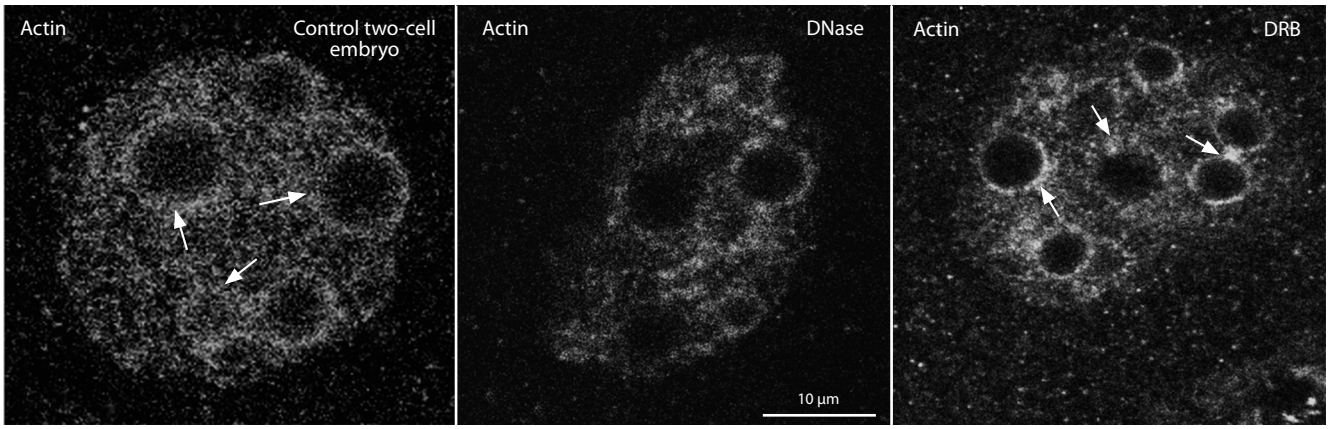


Fig. 4. Immunofluorescent localization of actin in the nuclei of two-cell mouse embryos.

Note: the prominent staining around NPBs (*arrows*), which becomes weaker after DNase treatment and enhances after inhibition of transcription by DRB. From (Bogolyubova, 2013), with permission from Tsitologiya.

chromatin-remodeling protein ATRX and Daxx, a chaperone of H3K9me3) begin to appear in the given area (see Table).

Also intriguing is the fact that the NPB-associated chromatin co-localizes with nuclear actin, and the anti-actin staining is most expressed at the transcriptionally active two-cell stage (Bogolyubova, Bogolyubova, 2009). When transcription was suppressed pharmacologically, actin continues to be revealed around the NPBs. Moreover, additional zones of its localization appear. The intensity of labeling decreases significantly after DNase treatment. This may suggest that actin is able to interact with DNA directly (Bogolyubova, 2013) (Fig. 4).

Molecular composition of different NPB-associated heterochromatin areas may differ one from another within the same nucleus

This point is well illustrated by the peculiarities of the distribution of ATRX protein being detectable at the periphery of only several NPBs (Sailau et al., 2017) (Fig. 5). The peculiarities of the molecular composition of NBP-associated chromatin may reflect the functional heterogeneity of NPBs themselves. The change in the number of NPBs during embryo development, initially much higher than the number of active nucleoli that function at later stages of cleavage, allows us to believe that this heterogeneity is the reflection of the competence of individual NPBs to nucleologenesis.

The ideas concerning the heterogeneity of NPBs in relation to nucleologenesis were previously expressed by other authors. For example, L. Romanova et al. (2006) using the method of highly sensitive FISH, have shown that one or more NPBs are not associated with rDNA in the nucleus of early mouse embryos. In this regard, the authors speak about the heterogeneity of NPBs with the respect to their ability to bind rDNA and, therefore, about their different contributions to the formation of functionally mature nucleoli.

What stage of embryogenesis this heterogeneity is formed at? It is difficult to answer this question at the present time, since the molecular mechanisms of the assembly of NPBs remain virtually unexplored. There are only a few data concerning the molecular mechanisms of assembly and/or maintenance of the NPB structure. For example, both NPBs and NPB-associated heterochromatin were found to exhibit some abnormalities in the embryos of mice, knocked-out in the gene encoding the nucleolar transcription factor UBF (Hamdane et al., 2017). At the moment there are no data on

differences in the morphology and/or molecular composition of the NPBs, associated and non-associated with rDNA. Therefore, it cannot be excluded that the formation of NPBs after fertilization and first division of cleavage occurs according to the mechanism of macromolecular crowding, and the local concentration of the constituent molecules plays a crucial role in this process. According to modern concepts, the macromolecular crowding is a general principle of the formation of nuclear organelles not surrounded by physical boundaries, such as the membranes (Richter et al., 2008; Rajendra et al., 2010; Cho, Kim, 2012). These non-membrane organelles, also mentioned as the biomolecular condensates (Banani et al., 2017), are formed according to the laws of colloidal physical chemistry due to the concentration of the molecules in the process of liquid-liquid phase separation (LLPS) (Zhu, Brangwynne, 2015; Courchaine et al., 2016; Staněk, Fox, 2017; Gomes, Shorter, 2018; Sawyer et al., 2018a, b). It is possible that the association of individual NPBs with rDNA may be random at the initial stages of the formation of pronuclei.

From our point of view, despite the possible role of stochastic processes at the beginning steps of the formation of the NPB-heterochromatin complex, this complex can be regarded to as a multifunctional provisional domain of early embryos during embryonic genome activation, both components of which largely function as a whole. However, the mechanisms leading to the formation of this complex structure as well as the functional heterogeneity of NPBs require further studies, which have to combine classical morphological methods with modern molecular biological approaches.

Conclusions

Our observations generally show that the NPB-associated heterochromatin in one- to two-cell mouse embryos has a peculiar molecular composition, different from that of canonical peripheral heterochromatin. Since the nucleus of mammalian blastomeres represents a unique dynamic system at the early stages of cleavage (Bogolyubova, Bogolyubov, 2014), one can assume that in early embryos, a number of functional nuclear domains including the peculiar heterochromatin rims around NPBs have a wider range of functions than in somatic cells. For instance, a close relationship between the NPBs and NBP-associated heterochromatin with newly assembling Cajal bodies has already been shown (Ferreira, Carmo-Fonseca, 1995; Zatsepina et al., 2003). At the initial stages of mouse embryogenesis, pericentric and centromeric heterochromatin was shown to localize around the NPBs (Probst et al., 2007). All these data allows assuming that the NPB-associated heterochromatin represents not just an area of repressed chromatin at the beginning of mouse development, but could be a structural scaffold to form a definitive 3D architectonics of the cell nucleus.

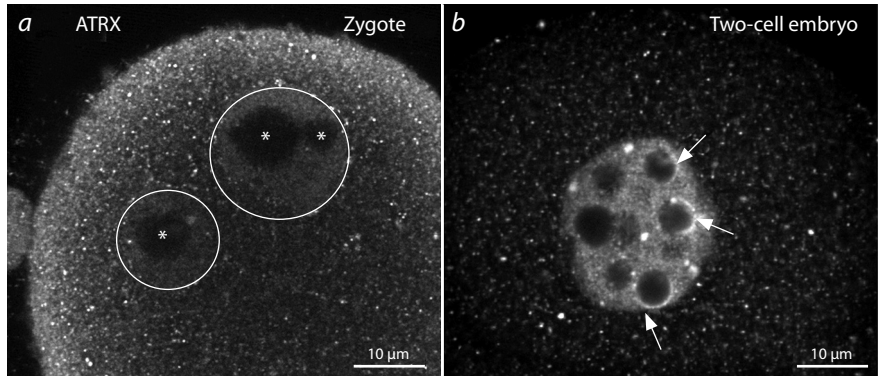


Fig. 5. Immunolocalization of the chromatin-remodeling protein ATRX in the nuclei of zygote (*a*) and two-cell embryo (*b*).

a – pronuclei are encircled, *asterisks* indicate NPBs. Note that ATRX is detectable in NPB-associated chromatin (arrows) only at the two-cell stage, and anti-ATRX staining is associated not with all NPBs. From (Sailau et al., 2017), with permission from Elsevier.

References

Banani S.F., Lee H.O., Hyman A.A., Rosen M.K. Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. 2017;18(5):285-298. DOI 10.1038/nrm.2017.7.

Bogolyubova I.O. Transcriptional activity of nuclei in 2-cell blocked mouse embryos. Tissue Cell. 2011;43(4):262-265. DOI 10.1016/j.tice.2011.03.005.

Bogolyubova I.O. Comparative analysis of the fluorescent labeling pattern of nuclei of early mouse embryos by using antibodies to various actin molecule domains. Cell Tiss. Biol. 2013;7(1):37-42. DOI 10.1134/S1990519X13010033.

Bogolyubova I.O., Bogolyubov D.S. An immunocytochemical study of interchromatin granule clusters in early mouse embryos. Biomed. Res. Int. 2013;2013:931564. DOI 10.1155/2013/931564.

Bogolyubova I.O., Bogolyubov D.S. Nuclear distribution of RNA polymerase II and mRNA processing machinery in early mammalian embryos. Biomed. Res. Int. 2014;2014:681596. DOI 10.1155/2014/681596.

Bogolyubova I.O., Bogolyubov D.S. Combined detection of newly synthesized RNA and nuclear proteins at the ultrastructural level: a modification of the protocol for immunoelectron microscopy. Cell Tiss. Biol. 2018;12(6):517-522. DOI 10.1134/S1990519X18060020.

Bogolyubova N.A., Bogolyubova I.O. Actin localization in nuclei of two-cell mouse embryos. Cell Tiss. Biol. 2009;3(5):417-422. DOI 10.1134/S1990519X09050034.

Burton A., Torres-Padilla M.E. Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo. Brief. Funct. Genomics. 2010; 9(5-6):444-454. DOI 10.1093/bfgp/elq027.

Cho E.J., Kim J.S. Crowding effects on the formation and maintenance of nuclear bodies: insights from molecular-dynamics simulations of simple spherical model particles. Biophys. J. 2012;103(3):424-433. DOI 10.1016/j.bpj.2012.07.007.

Courchaine E.M., Lu A., Neugebauer K.M. Droplet organelles? EMBO J. 2016;35(15):1603-1612. DOI 10.15252/embj.201593517.

Fadloun A., Le Gras S., Jost B., Ziegler-Birling C., Takahashi H., Gorab E., Carninci P., Torres-Padilla M.E. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat. Struct. Mol. Biol. 2013;20(3):332-338. DOI 10.1038/nsmb.2495.

Ferreira J., Carmo-Fonseca M. The biogenesis of the coiled body during early mouse development. Development. 1995;121(2):601-612.

Geuskens M., Alexandre H. Ultrastructural and autoradiographic studies of nucleolar development and rDNA transcription in preimplantation mouse embryos. Cell Differ. 1984;14(2): 125-134.

Gomes E., Shorter J. The molecular language of membraneless organelles. J. Biol. Chem. 2018. DOI 10.1074/jbc.TM118.001192.

Hamdane N., Tremblay M.G., Dillinger S., Stefanovsky V.Y., Németh A., Moss T. Disruption of the UBF gene induces aberrant somatic nucleolar bodies and disrupts embryo nucleolar precursor bodies. Gene. 2017;612:5-11. DOI 10.1016/j.gene.2016.09.013.

Kopecný V., Flechon J.-E., Camous S., Fulka J.Jr. Nucleologenesis and the onset of transcription in the eight-cell bovine embryo: fine structural autoradiographic study. Mol. Reprod. Dev. 1989;1(2):79-90. DOI 10.1002/mrd.1080010202.

Lanctôt C., Cheutin T., Cremer M., Cavalli G., Cremer T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat. Rev. Genet. 2007;8(2):104-115. DOI 10.1038/nrg2041.

Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nat. Rev. Genet. 2002;3(9):662-673. DOI 10.1038/nrg887.

Martin C., Beaujean N., Brochard V., Audouard C., Zink D., Debey P. Genome restructuring in mouse embryos during reprogramming and early development. Dev. Biol. 2006;292(2):317-332. DOI 10.1016/j.ydbio.2006.01.009.

Mason K., Liu Z., Aguirre-Lavin T., Beaujean N. Chromatin and epigenetic modifications during early mammalian development. Anim. Reprod. Sci. 2012;134(1-2):45-55. DOI 10.1016/j.anireprosci.2012.08.010.

Meglicki M., Teperek-Tkacz M., Borsuk E. Appearance and heterochromatin localization of HP1α in early mouse embryos depends on cytoplasmic clock and H3S10 phosphorylation. Cell Cycle. 2012;11(11):2189-2205. DOI 10.4161/cc.20705.

Probst A.V., Santos F., Reik W., Almouzni G., Dean W. Structural differences in centromeric heterochromatin are spatially reconciled on fertilization in the mouse zygote. Chromosoma. 2007;116(4):403-415. DOI 10.1007/s00412-007-0106-8.

Rajendra T.K., Praveen K., Matera A.G. Genetic analysis of nuclear bodies: from nondeterministic chaos to deterministic order. Cold Spring Harb. Symp. Quant. Biol. 2010;75:365-374. DOI 10.1101/sqb.2010.75.043.

Richter K., Nessling M., Lichter P. Macromolecular crowding and its potential impact on nuclear function. Biochim. Biophys. Acta. 2008;1783(11):2100-2107. DOI 10.1016/j.bbamcr.2008.07.017.

Romanova L., Korobova F., Noniashvilli E., Dyban A., Zatsepina O. High resolution mapping of ribosomal DNA in early mouse embryos by fluorescence in situ hybridization. Biol. Reprod. 2006;74(5):807-815. DOI 10.1095/biolreprod.105.047340.

Sailau Z.K., Bogolyubov D.S., Bogolyubova I.O. Nuclear distribution of the chromatin-remodeling protein ATRX in mouse early embryogenesis. Acta Histochem. 2017;119(1):18-25. DOI 10.1016/j.acthis.2016.11.001.

Santenard A., Ziegler-Birling C., Koch M., Tora L., Bannister A.J., Torres-Padilla M.E. Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat. Cell Biol. 2010;12(9):853-862. DOI 10.1038/ncb2089.

Sawyer I.A., Bartek J., Dundr M. Phase separated microenvironments inside the cell nucleus are linked to disease and regulate epigenetic state, transcription and RNA processing. Semin. Cell Dev. Biol. 2018a. DOI 10.1016/j.semcdb.2018.07.001.

Sawyer I.A., Sturgill D., Dundr M. Membraneless nuclear organelles and the search for phases within phases. Wiley Interdiscip. Rev. RNA. 2018b;e1514. DOI 10.1002/wrna.1514.

Staněk D., Fox A.H. Nuclear bodies: new insights into structure and function. Curr. Opin. Cell Biol. 2017;46:94-101. DOI 10.1016/j.ceb.2017.05.001.

Tesařík J., Kopecný V., Plachot M., Mandelbaum J. Activation of nucleolar and extranucleolar RNA synthesis and changes in the ribosomal content of human embryos developing *in vitro*. J. Reprod. Fertil. 1986a;78(2):463-470.

Tesařík J., Kopecný V., Plachot M., Mandelbaum J., Dalage C., Flechon J.-E. Nucleologenesis in the human embryo developing *in vitro*: ultrastructural and autoradiographic analysis. Dev. Biol. 1986b;115(1):193-203.

Wang C., Liu X., Gao Y., Yang L., Li C., Liu W., Chen C., Kou X., Zhao Y., Chen J., Wang Y., Le R., Wang H., Duan T., Zhang Y., Gao S. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat. Cell Biol. 2018;20(5):620-631. DOI 10.1038/s41556-018-0093-4.

Zatsepina O., Baly C., Chebrou M., Debey P. The step-wise assembly of a functional nucleolus in preimplantation mouse embryos involves the Cajal (coiled) body. Dev. Biol. 2003;253(1):66-83.

Zhu L., Brangwynne C.P. Nuclear bodies: the emerging biophysics of nucleoplasmic phases. Curr. Opin. Cell Biol. 2015;34:23-30. DOI 10.1016/j.ceb.2015.04.003.

Acknowledgements. The work is supported by the granting program of the Presidium of RAS "Molecular and Cell Biology".

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised January 18, 2019. Accepted January 24, 2019.

Вавиловский журнал генетики и селекции. 2019;23(2):135-139
DOI 10.18699/VJ19.472

Patterns of nucleotide diversity for different domains of centromeric histone H3 (CENH3) gene in *Secale* L.

S.S. Gatzkaya, E.V. Evtushenko

Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
e-mail: evt@mcb.nsc.ru

Rye (*Secale*) is among staple cereals along with other members of the Triticeae tribe: wheat and barley. The genus *Secale* includes perennial and annual, cross-pollinating and self-pollinating species, and they can be donors of valuable genes in wheat and rye breeding programs. Studies of the structure of the gene for centromeric histone H3 (CENH3), essential for centromere functions, are relevant to the breeding of agronomically important crops. We have investigated the nucleotide diversity of sequences of two variants of the rye *CENH3* gene inside the N-terminal tail (NTT) and the conservative HFD (histone fold domain) domain in the genus *Secale*. The mean values of nucleotide diversity in the NTT and HFD of wild cross- and self-pollinating taxa are close in α*CENH3*: π_{tot} = 0.0176–0.0090 and 0.0136–0. 0052, respectively. In the case of β*CENH3*, the mean values for NTT (π_{tot} = 0.0168–0.0062) are lower than for HFD (π_{tot} = 0.0259–0.084). The estimates of nucleotide and haplotype diversity per site for the *CENH3* domains are considerably lower in taxa with narrow geographic ranges: *S. cereale* subsp. *dighoricum* and *S. strictum* subsp. *kuprijanovii*. Commercial breeding reduces the nucleotide sequence variability in α*CENH3* and β*CENH3*. Cultivated rye varieties have π values within 0.0122–0.0014. The nucleotide and haplotype diversity values in α*CENH3* and β*CENH3* are close in *S. sylvestre*, which is believed to be the oldest rye species. The results of this study prove that the frequency of single nucleotide polymorphisms and nucleotide diversity of sequences in genes for CENH3 in *Secale* species are influenced by numerous factors, including reproduction habits, the geographic isolation of taxa, breeding, and the evolutionary age of species.
Key words: *Secale* L.; centromeric histone CENH3; nucleotide diversity; single nucleotide polymorphism.

For citation: Gatzkaya S.S., Evtushenko E.V. Patterns of nucleotide diversity for different domains of centromeric histone H3 (CENH3) gene in *Secale* L. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):135-139. DOI 10.18699/VJ19.472

Паттерны нуклеотидного разнообразия различных доменов гена центромерного гистона H3 (CENH3) у *Secale* L.

С.С. Гацкая, Е.В. Евтушенко

Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
e-mail: evt@mcb.nsc.ru

Рожь (*Secale*) входит в группу экономически важных злаков наряду с такими представителями трибы Triticeae, как пшеница и ячмень. Род *Secale* включает многолетние, однолетние, перекрестноопыляющиеся и самоопыляющиеся виды, которые используются как источник ценных генов для улучшения существующих сортов пшеницы и ржи. Исследования структуры гена центромерного гистона H3 (CENH3), определяющего функциональную центромеру, сейчас становится актуальным для агрономически важных растений. Мы изучили нуклеотидное разнообразие последовательностей двух вариантов гена *CENH3* ржи внутри N-терминального района (NTT) и консервативного домена (HFD) гена в роде *Secale*. Средние значения нуклеотидного разнообразия у диких перекрестно- и самоопыляющихся таксонов для доменов α*CENH3* были близки для NTT (π_{tot} = 0.0176–0.0090) и HFD (π_{tot} = 0.0136–0.0052), а для β*CENH3* средние значения были меньше в NTT (π_{tot} = 0.0168–0.0062), чем в HFD (π_{tot} = 0.0259–0.084). Значения нуклеотидного и гаплотипного разнообразия для доменов *CENH3* были существенно меньше у таксонов, занимающих узкую географическую нишу, *S. cereale* subsp. *dighoricum* и *S. strictum* subsp. *kuprijanovii*. К снижению изменчивости нуклеотидных последовательностей доменов α*CENH3* и β*CENH3* приводит действие селекции: у сортов культивируемой ржи значения π варьируют от 0.0122 до 0. 0014. Значения нуклеотидного и гаплотипного разнообразия поддерживаются на одном уровне в последовательностях α*CENH3* и β*CENH3* у *S. sylvestre*, считающегося наиболее древним видом ржи. Полученные результаты подтверждают, что на частоту однонуклеотидных полиморфизмов и нуклеотидное разнообразие последо-

вательностей вариантов CENH3 у видов *Secale* влияет ряд факторов, включая способы размножения, степень географической изоляции таксона, действие селекции, эволюционный возраст видов.
Ключевые слова: *Secale* L.; центромерный гистон H3 (CENH3); нуклеотидное разнообразие; однонуклеотидный полиморфизм.

Introduction

Rye (*Secale*) belongs to the tribe Triticeae along with other grain crops, such as wheat (*Triticum* spp.) and barley (*Hordeum* spp.) According to the taxonomy proposed by S. Frederiksen and G. Petersen (1998) on the base of morphometrical analysis of rye species, the genus consists of three botanical species: the cross-pollinating perennial species *Secale strictum* Presl., cross-pollinating annual species *S. cereale* L., and self-pollinating annual species *S. sylvestre* Host. The cross-pollinating species *S. cereale* L. and *S. strictum* Presl. include self-pollinating subspecies *S. vavilovii* and *S. africanum*, respectively.

Cultivated and wild rye varieties can be donors of valuable traits, such as winter hardiness, high protein content, and disease resistance. They are used for improvement of existing rye and wheat cultivars and in interspecies crosses of rye and wheat (Tang et al., 2011). Interspecies hybridization is often accompanied by elimination of whole chromosomes or their parts. The incompatibility between centromeres and the centromere-specific histone H3 variant (CENH3) of parents may be one of the causes of chromosome elimination from interspecies hybrids (Sanei et al., 2011). Conversely, close similarity between CENH3 proteins of distant parents can secure the normal function of centromeres and formation of true hybrid plants bearing genomes of both parents (Ishii et al., 2015). The structure of CENH3 includes two domains: the variable N-terminal tail (NTT) and the more conservative C-terminal histone fold domain (HFD) (Roach et al., 2012). The latter interacts with centromeric DNA, whereas NTT is not required for localization on centromeric DNA but is essential for correct chromosome segregation in mitosis and meiosis (Maheshwari et al., 2015). Nevertheless, the segregating polymorphism of *CENH3* genes in grass species is poorly investigated. The multitude of rye species, including annual, perennial, self-pollinating, and cross-pollinating forms, allows assessment of the action of various factors on the genetic variation of *CENH3*.

The objectives of this study were: (1) estimation of the nucleotide and haplotype diversity of *CENH3* domains in three *Secale* species and (2) assessment of the influence of taxonomic and geographic factors and mating systems on nucleotide polymorphisms within the domains in two variants of the *CENH3* gene.

Materials and methods

Plant material and RNA isolation. Experiments were done with ten wild and cultivated rye accessions of three *Secale* species. Seeds of *Secale cereale* subsp. *cereale* (cvs. Otello, Imperial), *S. cereale* subsp. *vavilovii*, *S. cereale* subsp. *dighoricum*, *S. cereale* subsp. *afghanicum*, *S. strictum* subsp. *kuprijanovii*, *S. strictum* subsp. *strictum*, *S. strictum* subsp. *anatolicum*, *S. strictum* subsp. *africanum*, and *S. sylvestre* were supplied by the Leibniz Institute of Plant Genetics and Crop Plant Research (Germany), the US Department of Agriculture (United States), and the N.I. Vavilov Research Institute of Plant Industry (Russia) from their germplasm collections. Total RNA isolation, synthesis of first-strand cDNA, and PCR were conducted as in (Evtushenko et al., 2017). Amplification primers specific to NTTs and HFDs of α*CENH3* and β*CENH3* genes from rye cDNA had been chosen in (Evtushenko et al., 2017). The amplification products were cloned and sequenced using BigDye Terminator Cycle Sequencing chemistry (v. 3.1) on an ABI3100 Genetic Analyzer (Applied Biosystems, CA, USA).

Sequence analysis. Alignments of *CENH3* coding sequences were performed using online Clustal Omega (Sievers et al., 2011) at http://www.ebi.ac.uk/Tools/msa/clustalo. The DnaSP version 5.10.01 (Librado, Rozas, 2009) was used to estimate the levels of nucleotide diversity for each domain individually in all subspecies with regard to different functions of CENH3 domains. The levels of genetic variation within *CENH3* were estimated as nucleotide diversity π, haplotype diversity H_d, and θ_W, the last index being the relationship between segregating sites and alleles. The Watterson estimator θ_W is based on the number of polymorphic sites in a sample of sequences drawn at random from a population (Watterson, 1975), whereas nucleotide diversity π represents the average sequence divergence of all homologous sequences among all individuals in a given set for comparison (Nei, Li, 1979).

Results

We sequenced the NTT and HFD domains of *CENH3* from 10 to 25 samples per domain for each accession. Formerly, we had shown that the main forms of rye α*CENH3* are β*CENH3* were 501- and 456-bp long, respectively, in all *Secale* species and subspecies (Evtushenko et al., 2017). The lengths of NTTs and HFDs in α*CENH3* and β*CENH3* analyzed in this study are shown in Table 1.

In α*CENH3*, 1 to 21 single-nucleotide polymorphisms (SNPs), or segregating sites, were found inside the domains (Table 2). In the perennial *S. strictum* Presl. subspecies, the number of SNPs in HFD was greater than in NTT, whereas in annual *S. cereale* L. subspecies NTTs had more segregating sites. Among perennial subspecies, the highest genetic variation in the *CENH3* domains was detected in wild cross-pollinating forms: *S. strictum* ssp. *strictum*, hereafter referred to as *S. strictum* and *S. strictum* ssp. *anatolicum* (hereafter *S. anatolicum*). There, π ranged from 0.0131 to 0.0140 in NTT

Table 1. Lengths of coding sequences (CDS) of *CENH3* genes in *Secale*

Full gene/domain	α*CENH3*	β*CENH3*
Full-length CDS (bp)	501	456
N-terminal tail	213	165
HFD	288	291

Table 2. Estimates of nucleotide diversity in the NTT and HFD of α*CENH3*

Species and subspecies	Growth habit	*CENH3* domain	S_n	$H_d \pm SD$	Polymorphism ($\times 10^{-2}$)			
					$\theta_w \pm SD$	$\pi_{tot} \pm SD$	π_{syn}	π_{nonsyn}
S. strictum ssp. *strictum*	O, P, W	NTT	10	0.972 ± 0.064	1.70 ± 0.85	1.31 ± 0.22	0.985	1.43
		HFD	14	0.805 ± 0.069	1.87 ± 0.78	1.16 ± 0.26	1.80	2.35
S. strictum ssp. *anatolicum*	O, P, W	NTT	15	0.879 ± 0.052	1.95 ± 0.81	1.40 ± 0.26	2.41	1.02
		HFD	21	0.942 ± 0.048	2.84 ± 1.13	1.36 ± 0.21	2.59	1.01
S. strictum ssp. *africanum*	S, P, W	NTT	8	0.867 ± 0.107	1.31 ± 0.67	1.28 ± 0.25	3.21	0.53
		HFD	13	0.881 ± 0.047	1.71 ± 0.27	1.01 ± 0.16	1.98	0.73
S. strictum ssp. *kuprijanovii*	O, P, W	NTT	3	0.700 ± 0.218	0.66 ± 0.47	0.65 ± 0.26	1.69	0.25
		HFD	5	0.380 ± 0.125	0.63 ± 0.33	0.19 ± 0.07	0.35	0.15
S. cereale ssp. *cereale* (Otello)	O, A, Cv	NTT	7	0.736 ± 0.109	1.01 ± 0.51	0.81 ± 0.21	1.66	0.49
		HFD	4	0.476 ± 0.155	0.58 ± 0.34	0.25 ± 0.09	0.84	0.08
S. cereale ssp. *cereale* (Imperial)	O, A, Cv	NTT	14	0.801 ± 0.088	1.82 ± 0.75	0.79 ± 0.16	0.95	0.76
		HFD	4	0.806 ± 0.120	0.64 ± 0.39	0.51 ± 0.12	0.97	0.37
S. cereale ssp. *afghanicum*	O, A, W	NTT	5	0.833 ± 0.222	1.26 ± 0.82	1.23 ± 0.40	2.50	0.74
		HFD	6	0.727 ± 0.144	0.97 ± 0.52	0.52 ± 0.15	1.14	0.34
S. cereale ssp. *vavilovii*	S, A, W	NTT	14	0.924 ± 0.057	2.14 ± 0.97	1.76 ± 0.47	3.56	1.08
		HFD	13	0.758 ± 0.106	1.79 ± 0.77	0.96 ± 0.26	3.55	0.20
S. cereale ssp. *dighoricum*	O, A, W	NTT	3	0.524 ± 0.209	0.56 ± 0.38	0.39 ± 0.19	0.48	0.36
		HFD	1	0.111 ± 0.096	0.14 ± 0.13	0.053 ± 0.046	0.0000	0.068
S. sylvestre	S, A, W	NTT	9	0.879 ± 0.079	1.31 ± 0.62	0.90 ± 0.16	1.87	0.53
		HFD	16	0.859 ± 0.066	2.03 ± 0.81	0.86 ± 0.14	2.24	0.46

Note: O – open-pollinated; S – self-pollinated; A – annual; P – perennial; W – wild/weedy; Cv – cultivar; S_n – number of segregating sites; H_d – haplotype diversity; θ_w – Watterson estimator; SD – standard deviation.
Average nucleotide pairwise diversity: π_{tot} – total (synonymous, nonsynonymous); π_{syn} – synonymous; π_{nonsyn} – nonsynonymous.

and from 0.0116 to 0.0136 in HFD. Similarly, the estimates of nucleotide diversity for wild annual cross-pollinating *S. cereale* ssp. *afghanicum* (hereafter *S. afghanicum*) in both α*CENH3* domains were higher than in annual rye cultivars. High nucleotide diversity levels were found in all the three self-pollinating rye accessions studied: *S. cereale* ssp. *africanum* (hereafter *S. africanum*), *S. cereale* ssp. *vavilovii* (hereafter *S. vavilovii*), and *S. sylvestre* Host., regardless of their belonging to annual or perennial *Secale* species. The overall number of segregating sites in both α*CENH3* domains of self-pollinating subspecies varied from 21 in in *S. africanum* to 27 in *S. vavilovii*, and the π values for these subspecies ranged from 0.0128 to 0.0176 in NTT (see Table 2). The estimates of nucleotide diversity for the self-pollinating rye species *S. sylvestre* were somewhat lower, but the overall number of segregating sites was 25 for both domains. The θ_W values in *S. sylvestre* were also higher than in some cross-pollinating subspecies: 1.31 (NTT) and 2.03 (HFD), and the π values for NTT and HFD were close: 0.0090 and 0.0086. Thus, the nucleotide diversity of *CENH3* is equally high in rye accessions from wild cross- and self-pollinating populations.

Perennial *S. strictum* ssp. *kuprijanovii* (*S. kuprijanovii*) and annual *S. cereale* ssp. *dighoricum* (*S. dighoricum*) showed the lowest nucleotide and haplotype diversities of α*CENH3* among the subspecies: 0.0065 to 0.0039 in NTT and 0.0019 to 0.005 in HFD. Accessions of these subspecies may have originated from a small geographic range, where they had higher inbreeding coefficients to limit geneflow within the taxa (Hagenblad et al., 2016). The comparison of nucleotide diversity values (π_{tot}) for α*CENH3* domains of two rye cultivars and wild rye subspecies showed that the estimates of diversity were low in cultivated rye, which might have resulted from specific breeding features. Haplotype diversity values (H_d) were uniform across all domains of rye α*CENH3*, ranging from 0.972 (*S. strictum*) to 0.736 (*S. cereale* cv. Otello) in NTT and from 0.942 (*S. anatolicum*) to 0.727 (*S. afghanicum*) in HFD with the exception of low H_d values in *S. kuprijanovii* and *S. dighoricum*.

The average estimates of nucleotide diversity π for β*CENH3* were no lower than for α*CENH3* (Table 3). High π values were observed for the wild cross-pollinating subspecies *S. strictum* (NTT, 0.0137; HFD, 0.0120) and *S. afghanicum* (NTT, 0.0134; HFD, 0.0112). In the self-pollinating subspecies *S. africanum* and *S. vavilovii*, we also noted large numbers of segregating sites (21 and 15 in both *CENH3* domains) and high nucleotide diversity (0.0168 and 0.0095 in NTT, 0.0295 and 0.0084 in

Table 3. Estimates of nucleotide diversity in the NTT and HFD of β*CENH3*

Species and subspecies	Growth habit	*CENH3* domain	S_n	$H_d \pm SD$	Polymorphism ($\times 10^{-2}$)			
					$\theta_w \pm SD$	$\pi_{tot} \pm SD$	π_{syn}	π_{nonsyn}
S. strictum ssp. *strictum*	O, P, W	NTT	10	0.790 ± 0.105	1.83 ± 0.85	1.37 ± 0.36	2.22	1.06
		HFD	10	0.786 ± 0.151	1.35 ± 0.69	1.20 ± 0.35	3.86	0.34
S. strictum ssp. *africanum*	S, P, W	NTT	11	0.933 ± 0.077	2.35 ± 1.14	1.68 ± 0.35	2.09	1.53
		HFD	10	1.000 ± 0.045	3.69 ± 1.60	2.59 ± 0.59	6.11	1.39
S. strictum ssp. *kuprijanovii*	O, P, W	NTT	6	0.731 ± 0.133	1.15 ± 0.61	0.69 ± 0.17	0.69	0.68
		HFD	8	0.800 ± 0.010	1.01 ± 0.52	0.95 ± 0.17	1.67	0.74
S. cereale ssp. *cereale* (Otello)	O, A, Cw	NTT	11	0.956 ± 0.045	2.05 ± 0.95	1.22 ± 0.22	1.25	1.20
		HFD	9	0.933 ± 0.077	1.11 ± 0.56	0.87 ± 0.08	1.09	0.80
S. cereale ssp. *cereale* (Imperial)	O, A, Cw	NTT	7	0.657 ± 0.138	1.59 ± 0.63	0.54 ± 0.16	0.89	0.42
		HFD	3	0.257 ± 0.142	0.31 ± 0.20	0.14 ± 0.08	0.00	0.18
S. cereale ssp. *afghanicum*	O, A, W	NTT	15	0.875 ± 0.081	2.05 ± 0.88	1.34 ± 0.27	1.67	1.22
		HFD	8	0.900 ± 0.161	1.34 ± 0.78	1.12 ± 0.33	2.89	0.55
S. cereale ssp. *vavilovii*	S, A, W	NTT	9	0.699 ± 0.117	1.55 ± 0.72	0.95 ± 0.28	1.16	0.87
		HFD	11	0.694 ± 0.021	1.41 ± 0.60	0.84 ± 0.17	1.96	0.46
S. sylvestre	S, A, W	NTT	6	0.571 ± 0.014	1.11 ± 0.58	0.62 ± 0.21	0.56	0.64
		HFD	15	0.857 ± 0.090	1.60 ± 0.69	0.90 ± 0.16	2.61	0.37

Note: Designations follow Table 2.

HFD). In most accessions, the levels of nucleotide polymorphism were high in both β*CENH3* domains. In the HFD domains of *S. africanum* and *S. sylvestre*, the indices S_n (the number of segregating sites), θ_W, π, and H_d were higher than in NTTs. Of the two paralogous *CENH3* genes of rye, β*CENH3* appears to be the younger (Evtushenko et al., 2017), and this younger age may be responsible for the high genetic variation in HFD, the more conservative *CENH3* region. Both α*CENH3* and β*CENH3* show higher nucleotide diversities at synonymous sites than at nonsynonymous except for the nucleotide diversity values in NTT and HFD of α*CENH3* in *S. strictum* and two cases with $\pi_{syn} = 0.0000$: α*CENH3* of *S. dighoricum* and β*CENH3* of *S. cereale* cv. Imperial. These estimates confirm the effect of purifying selection on rye CENH3 and the possibility of adaptive selection for individual codons, formerly demonstrated by E.V. Evtushenko et al. (2017).

Discussion

We compare nucleotide diversity patterns in domains of the coding sequence of the gene encoding centromeric histone H3 (CENH3), which is one of the epigenetic tags of an active centromere. Centromeres, to which microtubules are attached, define the proper cell segregation in mitosis and meiosis (Comai et al., 2017). Nucleotide diversity comparisons are performed within the genus *Secale*, whose accessions represent annual, perennial, cross-pollinating, and self-pollinating forms; subspecies that experienced geographic isolation; and cultivated varieties. Levels of genetic diversity assessed as the numbers of segregating sites S_n, the Watterson estimator θ_W, nucleotide diversity π, and haplotype diversity H_d in the sequences of the α*CENH3* and β*CENH3* domains are higher in wild cross-pollinating subspecies, be they perennial or annual. However, the cross-pollinating reproductive habit is by no means the only factor increasing nucleotide diversity in rye *CENH3*. The same trend is observed in self-pollinating perennial *S. africanum* and in the annual subspecies *S. vavilovii* and *S. sylvestre*. In contrast, significantly lower nucleotide diversities π and numbers of segregating sites S_n are found in cross-pollinating subspecies populating small geographic ranges: *S. dighoricum* and *S. kuprijanovii*. The lower *CENH3* nucleotide diversity in rye cultivars in comparison to wild accessions may be related to the effects of different genotypes involved in breeding and the breeding process itself. The estimates of nucleotide diversity for rye *CENH3* are higher than for the *ScVrn1* gene, which controls vernalization sensitivity in rye (Li et al., 2011), but are close to estimates for *HTR12* (*CENH3* analog) in *Arabidopsis lyrata* and *Arabidopsis lyrata* ssp. *petraea* (Kawabe et al., 2006).

Conclusion

The nucleotide variation, or sequence diversity, in rye *CENH3* is elevated by both cross-pollination and self-pollination in large natural populations. Geographic isolation of cross-pollinating rye forms, as well as breeding processes in case of cultivated rye reduce the nucleotide diversity of *CENH3*.

Nowadays, operations with the CENH3 structure are in broad use in breeding programs for production of haploid lines in important crops, capture of heterosis (Karimi-Ashtiyani et al., 2015). Further advance in this field demands identification and knowledge of CENH3 features in commercially important

species. Data on the nucleotide diversity of rye *CENH3* may be useful in choosing parents for crosses between wild and cultivated species.

References

Comai L., Maheshwari S., Marimuthu M.P.A. Plant centromeres. Curr. Opin. Plant Biol. 2017;36:158-167. DOI 10.1016/j.pbi.2017.03.003.

Evtushenko E.V., Elisafenko E.A., Gatzkaya S.S., Lipikhina Y.A., Houben A., Vershinin A.V. Conserved molecular structure of the centromeric histone CENH3 in *Secale* and its phylogenetic relationships. Sci. Rep. 2017;7:17628. DOI 10.1038/s41598-017-17932-8.

Frederiksen S., Petersen G. A taxonomic revision of Secale (Triticeae, Poaceae). Nord. J. Bot. 1998;18:399-420. DOI 10.1111/j.1756-1051.1998.tb01517.x.

Hagenblad J., Oliveira H.R., Forsberg N.E.G., Leino M.W. Geographical distribution of genetic diversity in *Secale landrace* and wild accessions. BMC Plant Biol. 2016;16:23. DOI 10.1186/s12870-016-0710-y.

Ishii T., Sunamura N., Matsumoto A., Eltayeb A.E., Tsujimoto H. Preferential recruitment of the maternal centromere-specific histone H3 (CENH3) in oat (*Avena sativa* L.)×pearl millet (*Pennisetum glaucum* L.) hybrid embryos. Chromosome Res. 2015;23:709-718. DOI 10.1007/s10577-015-9477-5.

Karimi-Ashtiyani R., Ishii T., Niessen M., Stein N., Heckmann S., Gurushidze M., Banaei-Moghaddam A.M., Fuchs J., Schubert V., Koch K., Weiss O., Demidov D., Schmidt K., Kumlehn J., Houben A. Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proc. Nat. Acad. Sci. USA. 2015;112:11211-11216. DOI 10.1073/PNAS.1504333112.

Kawabe A., Shuhei Nasuda S., Charlesworth D. Duplication of centromeric histone H3 (*HTR12*) gene in *Arabidopsis halleri* and *A. lyrata*, plant species with multiple centromeric satellite sequences. Genetics. 2006;174(4):2021-2032. DOI 10.1534/genetics.106.063628.

Li Y., Haseneyer G., Schön C.-C., Ankerst D., Korzun V., Wilde P., Bauer E. High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. BMC Plant Biol. 2011;11:6. DOI 10.1186/1471-2229-11-6.

Librado P., Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451-1452. DOI 10.1093/bioinformatics/btp187.

Maheshwari S., Tan E.H., West A., Franklin F.C.H., Comai L., Chan S.W.L. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet. 2015;11(2):e1004970. DOI 10.1371/journal.pgen.1004970.

Nei M., Li W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA. 1979;76(4):5269-5273. DOI 10.1073/pnas.76.10.5269.

Roach K.C., Ross B.D., Malik H.S. Rapid evolution of centromeres and centromeric/kinetochore proteins. Eds. R. Singh, J. Xu, R. Kulathinal. Evolution in the Fast Lane: Rapidly Evolving Genes and Genetic Systems. Oxford University Press, 2012;83-93.

Sanei M., Pickering R., Kumke K., Nasuda S., Houben A. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proc. Natl. Acad. Sci. USA. 2011;108(33):498-505. DOI 10.1073/pnas.1103190108.

Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J., Higgins D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 2011;7:539. DOI 10.1038/msb.2011.75.

Tang Z.X., Ross K., Ren Z.L., Yang Z.J., Zhang H.Y., Chikmawati T., Miftahudin M., Gustafson J.P. *Secale*. Ed. C. Kole. Wild Crop Relatives: Genomic and Breeding Resources. New York: Springer, 2011. DOI 10.1007/978-3-642-14228-4_8.

Watterson G.A. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 1975;7:256-276. DOI 10.1016/0040-5809(75)90020-9.

ORCID ID

E.V. Evtushenko orcid.org/0000-0003-2468-5924

Acknowledgements. This study was supported by the Russian Fundamental Scientific Research Program, project No. 0310-2019-0003, and the Russian Foundation for Basic Research, project No. 17-04-00748a. The authors acknowledge the resources provided by the Molecular and Cellular Biology Shared Access Center of the Institute of Molecular and Cellular Biology, Novosibirsk.

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised December 11, 2018. Accepted December 11, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):140-147
DOI 10.18699/VJ19.473

Replication timing in *Drosophila* and its peculiarities in polytene chromosomes

T.D. Kolesnikova[1, 2] ✉, O.V. Antonenko[1], I.V. Makunin[1, 3]

[1] Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Research Computing Centre, The University of Queensland, St Lucia, Queensland, Australia
✉ e-mail: kolesnikova@mcb.nsc.ru

Drosophila melanogaster is one of the popular model organisms in DNA replication studies. Since the 1960s, DNA replication of polytene chromosomes has been extensively studied by cytological methods. In the recent two decades, the progress in our understanding of DNA replication was associated with new techniques. Use of fluorescent dyes increased the resolution of cytological methods significantly. High-throughput methods allowed analysis of DNA replication on a genome scale, as well as its correlation with chromatin structure and gene activity. Precise mapping of the cytological structures of polytene chromosomes to the genome assembly allowed comparison of replication between polytene chromosomes and chromosomes of diploid cells. New features of replication characteristic for *D. melanogaster* were described for both diploid and polytene chromosomes. Comparison of genomic replication profiles revealed a significant similarity between *Drosophila* and other well-studied eukaryotic species, such as human. Early replication is often confined to intensely transcribed gene-dense regions characterized by multiple replication initiation sites. Features of DNA replication in *Drosophila* might be explained by a compact genome. The organization of replication in polytene chromosomes has much in common with the organization of replication in chromosomes in diploid cells. The most important feature of replication in polytene chromosomes is its low rate and the dependence of S-phase duration on many factors: external and internal, local and global. The speed of replication forks in *D. melanogaster* polytene chromosomes is affected by SUUR and Rif1 proteins. It is not known yet how universal the mechanisms associated with these factors are, but their study is very promising.
Key words: *Drosophila melanogaster*; replication timing; replication origins; replication initiation zone; polytene chromosome; endocycle; Supressor of UnderReplication; SUUR; Rif1.

For citation: Kolesnikova T.D., Antonenko O.V., Makunin I.V. Replication timing in *Drosophila* and its peculiarities in polytene chromosomes. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019; 23(2):140-147. DOI 10.18699/VJ19.473

Пространственно-временная организация репликации у дрозофилы и ее особенности в политенных хромосомах

Т.Д. Колесникова[1, 2] ✉, О.В. Антоненко[1], И.В. Макунин[1, 3]

[1] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Исследовательский вычислительный центр, Квинслендский университет, Сент-Люсия, Квинсленд, Австралия
✉ e-mail: kolesnikova@mcb.nsc.ru

Drosophila melanogaster – один из популярных модельных организмов для изучения репликации ДНК. Начиная с 1960-х годов репликацию политенных хромосом активно изучали с помощью цитологических методов. В последние двадцать лет прогресс в изучении репликации определялся применением новых методов. Использование флуоресцентных красителей значительно улучшило разрешение цитологических подходов. Наличие геномной последовательности позволило изучить и соотнести репликацию ДНК со структурой хроматина и активностью генов для эухроматиновых районов в масштабе генома. Картирование границ цитологических структур политенных хромосом на последовательности генома дало возможность сравнить временные характеристики репликации районов хромосом в клеточных культурах и клетках слюнной железы. Были описаны новые особенности репликации как для хромосом диплоидных клеток, так и для политенных хромосом дрозофилы. Анализ временных профилей репликации показал, что организация репликации имеет в своей основе те же закономерности, что и у других хорошо изученных с точки зрения репликации видов, в частности человека. Ранняя репликация, как правило, приурочена к районам, характеризующимся высоким уровнем транскрипции, высокой плотностью генов и присутствием множественных сайтов потенциальной инициации репликации. Компактность генома *D. melanogaster* вносит некоторые особенности в организацию ее репликации. Последовательность репликации генома в политенных хромосомах и хромосомах диплоидных клеток имеет много общего: инициация репликации приурочена к одним и тем же районам, между которыми лежат протяженные участки генома, где репликация происходит преимущественно от краев к середине. Важнейшими особенностями репликации в политенных хромосомах являются низкая

скорость репликационных вилок и зависимость протяженности S-фазы от множества как внутренних, так и внешних факторов. В политенных хромосомах *D. melanogaster* скорость движения репликационных вилок зависит от присутствия в хроматине белков SUUR и Rif1.
Ключевые слова: *Drosophila melanogaster*; репликационный тайминг; сайт начала репликации; политенные хромосомы; эндоцикл; Supressor of UnderReplication; SUUR; Rif1.

Introduction

The development of high-throughput methods facilitated the genome-scale analysis of many biological phenomena. By now, detailed replication timing profiles have been described for genomes of many model organisms. As a rule, these profiles were created by pulse labeling of fluorescent DNA precursors in nonsynchronized cell cultures followed by automatic cell sorting by amount of DNA, which reflects cell cycle stages. Generally, two or three fractions corresponding to the early, middle, and late S-phases were isolated (Gilbert, Cohen, 1987; Schübeler et al., 2002; Schwaiger et al., 2009). Replication profiles are calculated as log2-transformed ratios of probabilities of replication in the early and late S-phase (Schübeler et al., 2002). Negative values correspond to replication in early S-phase, and positive values point to late replication.

Replication timing profiles reflect nonuniform chromatin organization, which is in turn associated with the nonuniform organization of the genome. Generally, regions with transcriptionally active chromatin replicate in the first half of the S-phase, while regions with silenced chromatin replicate later (Gilbert, 2002; Hiratani, Gilbert, 2009; Schwaiger et al., 2009; Gilbert et al., 2010; Hansen et al., 2010). Both in mammals and *Drosophila* early replication correlates with high gene density and high transcription level. Borders of early replicating domains match borders of topological domains. Late replicating regions coincide with domains of silenced chromatin and lamina-associated domains (Pickersgill et al., 2006; Belyakin et al., 2010; Peric-Hupkes et al., 2010; Rhind, Gilbert, 2013; Pope et al., 2014; Boulos et al., 2015; Prioleau, MacAlpine, 2016). Replication profiles may vary among different cell types, reflecting different chromatin states in particular genomic regions (Hiratani et al., 2008, 2010; Schwaiger et al., 2009; Pope et al., 2010). Aberration of replication timing may cause genome instability, defects of chromatin condensation, and, as a consequence, oncogenic transformation of cells (Hiratani, Gilbert, 2009).

Replication timing is regulated at the level of extended chromosomal domains (Berezney et al., 2000; Hiratani, Gilbert, 2009; Gillespie, Blow, 2010). In human and mouse genomic regions of a megabase scale demonstrate nearly synchronous initiation of DNA replication and borders between early and late replicating regions coincide with borders of open and closed chromatin (Jackson, Pombo, 1998; Ma et al., 1998; Julienne et al., 2013). Activation of origins and initiation of replication play a key role in the determination of replication timing (Jackson et al., 2006; Gillespie, Blow, 2010). Recent studies demonstrated that topological domains played an important role in the establishment of replication domains, or regions with coordinated regulation of DNA replication (Pope et al., 2014). In addition, it was shown that some genomic regions have no active replication initiation sites. These regions replicate late by very long replicons initiated in adjacent early replicating regions (Norio et al., 2005; Durkin, Glover, 2007; Cadoret et al., 2008; Letessier et al., 2011; Debatisse et al., 2012).

Replication timing in *Drosophila* cell cultures

D. melanogaster was one of the first organisms for which whole-genome replication timing profiles were obtained and extended (>100 kb) domains of early and late replication were identified (Schübeler et al., 2002; MacAlpine D.M. et al., 2004; Schwaiger et al., 2009; MacAlpine H.K. et al., 2010; Lubelsky et al., 2014). According to D.M. MacAlpine et al. (2004) and M. Schwaiger et al. (2009), the average size of replication domains in *Drosophila* (i.e., extended zones of similar replication timings where replication initiation is coordinated) is estimated to be 180 kb, much smaller than megabase-sized domains in mammals (White et al., 2004; Woodfine et al., 2005; Hiratani et al., 2008). The distribution of the Origin Recognition Complex (ORC) is the most important determinant of the replication-timing program in *D. melanogaster* chromosomes (MacAlpine D.M. et al., 2004; MacAlpine H.K. et al., 2010; Lubelsky et al., 2014). Early replication domains are characterized by high density of ORC-binding sites, presence of active chromatin marks, high gene density, and elevated transcriptional activity. On the contrary, late replicating domains demonstrate the absence of active chromatin marks, low gene density, weak ORC binding, and presence of H3K27me2/3 or H3K9me2/3 (MacAlpine et al., 2010; Lubelsky et al., 2014).

Comparison of replication timing in two cell types revealed changes in 21 % of autosomal regions (Schwaiger et al., 2009). Changes in replication timing on autosomes correlate with difference in gene expression. An elevated level of H4K16ac was detected in early replicating regions, including nontranscribed regions on the male X chromosome (Schwaiger et al., 2009). Y. Lubelsky et al. (2014) analyzed replication profiles in three *D. melanogaster* cell cultures by high-throughput sequencing and concluded that the majority of early and late replicating domains, covering 60 % of the genome, demonstrated similar replication timing patterns in the cell lines. The highest density of ORC-binding sites was observed in domains replicating early in the cell lines. In dynamic domains with variable replication timing, the density of ORC binding sites was low regardless of the replication time (Lubelsky et al., 2014).

Work on cell cultures revealed general patterns of replication in *Drosophila*, but no criteria were suggested for borders of replication domains. The understanding of the subdivision of *Drosophila* genome into replication domains came from analysis of replication in polytene chromosomes.

Cytological studies of replication timing in polytene chromosomes

Polytene chromosomes are interphase chromosomes comprising multiple DNA copies stacked together. Their giant size makes it possible to visualize molecular processes, such as gene expression or replication, at the cytological level at a very high resolution. Characteristic polytene chromosome patterns of alternating dark bands and light interbands have been used for many years for fine mapping of various cyto-

genetic markers. There are two types of bands in *Drosophila* polytene chromosomes. The most pronounced black bands contain DNA in a very compact state, whereas gray bands are less condensed. The DNA of interbands is significantly decompacted (Spierer A., Spierer P., 1984; Kozlova et al., 1994; Vatolina et al., 2011; Zhimulev et al., 2014).

Since the 1960s, replication in polytene chromosomes of *Drosophila* and other Diptera has been extensively studied by cytological methods. A detailed cytological analysis of replication in polytene chromosomes of *D. melanogaster*, *Rhynchosciara angelae* (=*Rhynchosciara americana*), *Chironomus thummi*, *Anopheles stephensi* and some other species showed similar patterns of ^{3}H thymidine incorporation in different polyploid tissues (for references see (Zhimulev, 1999)). A stage of continuous labeling, when entire chromosomes were covered with ^{3}H thymidine was followed by a stage of discrete labeling, when only the compact bands were labeled. In addition, a stage of inverse discrete labeling was detected, when some interbands, puffs, and decondensed bands were labeled (for review see (Zhimulev, 1999)). This stage corresponds to the very early S-phase. To determine the order of replication for different regions, various elegant methods were employed, such as double radioactive labeling, determination of the DNA amount in replicating bands, and analysis of patterns after natural or artificial synchronization of endocycles (Keyl, Pelling, 1963; Plaut et al., 1966; Danieli, Rodinò, 1967; Mulder et al., 1968; Amabis, 1974; Stocker, Pavan, 1974; Achary et al., 1981; Redfern, 1981). In different species, the replication time of a band correlates with the amount of DNA in it (Keyl, 1965; Mulder et al., 1968; Hägele, 1976; Bedo, 1982). Big bands replicate late and complete replication in a similar temporal order in different polytene tissues of *Drosophila* (Sinha et al., 1987; Koryakov, Zhimulev, 2015) and mosquito *Anopheles stephensi* (Redfern, 1981).

The use of fluorescence detection increased the resolution of the method. Back in 1985, the replication of larval salivary gland polytene chromosomes of *Chironomus tummi* was investigated by BrdU incorporation followed by fluorescence detection (Allison et al., 1985). D. Koryakov and I. Zhimulev (2015) used BrdU to study replication in polytene chromosomes from nurse cells of *D. melanogaster otu* mutants. In recent years, replication detection with PCNA antibodies has been extensively used (Gibert, Karch, 2011; Kolesnikova et al., 2013, 2018; Andreyeva et al., 2017). The high resolution of the method revealed that at the continuous labeling stage, when replication was seen almost everywhere along the chromosome arms, large black bands had not even started replication, whereas some regions of loosely packed chromatin had already completed it (Kolesnikova et al., 2013; Koryakov, Zhimulev, 2015).

Similarity of replication timing in polytene chromosomes and chromosomes of diploid cells

Despite a great amount of papers devoted to cytological studies of replication in polytene chromosomes, until recently, it was difficult to compare these results with data obtained on diploid cells because morphological structures of polytene chromosome were mapped to the *Drosophila* genome assembly at insufficient resolution. Invoking cytological and molecular data, T.D. Kolesnikova et al. (2018) mapped borders of all 159 prominent black bands on chromosome 2R to the *Drosophila* genome assembly with a high accuracy. These 159 bands are characterized by the presence of silent chromatin *ruby* (Zhimulev et al., 2014) and hence were named rb-bands (Kolesnikova et al., 2018). The rb-bands demonstrate low gene density, low gene expression level in individual tissues, and enrichment in transcriptionally inactive types of chromatin. The INTervals between rb-bands, or INTs, comprised from gray bands and interbands, are characterized by high gene density, high proportion of actively transcribed genes, and open chromatin in different cell types. The rb-bands contain predominantly tissue-specific genes, the INTs are enriched in genes expressed at high levels in many tissues. The rb-bands correspond to 60 % of the euchromatic part of chromosome 2R. Their sizes vary within 15–500 kb, the median size being ~50 kb. INTs are generally smaller, the median size ~30 kb. Border regions between rb-bands and INTs often coincide with borders of topological domains. The borders are characterized by a dramatic difference in the presence of multiple epigenetics markers such as H3K27me3, Lamin, SUUR, and histone H1 (Kolesnikova et al., 2018). INTs and rb-bands split chromosome 2R into two types of domains with contrasting properties, which are conservative in different tissues (Kolesnikova et al., 2018). Analysis of rb-bands from another chromosome arm shows similar features (Kolesnikova, unpublished results).

Analysis of replication in polytene chromosomes using anti-PCNA immunostaining demonstrated that INTs replicate before rb-bands (Kolesnikova et al., 2018). Completion of replication in rb-bands correlates with the size of the regions. Comparison of publicly available replication data for D*rosophila* cell cultures (Schwaiger et al., 2009) revealed that regions corresponding to the rb-bands showed a similar replication timing in diploid Kc and Cl8 cells; that is, replication starts and complete later in these regions, and replication time correlates with the size of the region. INTs replicate early in the salivary glands and cell cultures (Kolesnikova et al., 2018).

Analysis of ORC distribution data for salivary glands and cell cultures (Sher et al., 2013) revealed that INTs are enriched with ORC binding sites in polytene chromosomes and a cell culture and serve as initiation zones of early replication in different cell types. The rb-bands are practically devoid of ORC binding sites and presumably replicated by forks entering from early replication zones. A negative correlation –0.6 between the size of a region and time of replication completion in the cell culture (a minimal replication score for a region) supports this assumption. M. Schwaiger et al. (2009) found that in *Drosophila* cell cultures average length of replicons originated from early and late origins are 80 and 30 kb, respectively and suggested that replication forks originated from early replication origins enter late replicating regions. Indeed, the size of early replicons (80 kb) exceeds that of early replicating INT regions (30 kb).

Finding early and late replication origins: challenges and obstacles

While early replication origins are well defined in the *Drosophila* and human genomes, the locations and the very nature of the late replication origins remain somewhat obscure.

Replication profiles are produced by analysis of a cell population. Comparison of replication initiation sites identified by different techniques in mammals shows that origins are often present in clusters, named replication initiation zones (Borowiec, Schildkraut, 2011; Cayrou et al., 2011; Mesner et al., 2011). Every replication initiation zone contains multiple replication origins. It is postulated that origins have low efficiency (probability of activation in a cell cycle). Replication initiates stochastically at any origin within the zone. Initiation of replication inactivates neighboring origins by 'interference' (Lebofsky et al., 2006; Petryk et al., 2016; Prioleau, MacAlpine, 2016). In humans, the median size of early replication zones is 150 kb (Petryk et al., 2016). Multiple ORC binding sites identified within these zones serve as potential replication origins. These origins are confined to particular regions and easily detected by different experimental techniques (Petryk et al., 2016). In *Drosophila*, INTs presumably act as early replication initiation zones and contain multiple ORC binding sites. Both in *Drosophila* and humans, early replication initiation zones are interwoven with late replicating regions characterized by U-shaped replication profiles. Replication origins in late replicating regions are less confined to specific sites, and they do not necessarily correspond to ORC binding sites; hence, they can escape detection by traditional approaches (Petryk et al., 2016). A 'cascade' model for sequential activation of internal origins by moving replication forks was proposed for late replicating regions (Chagin et al., 2010; Guilbaud et al., 2011; Petryk et al., 2016).

M. Schwaiger et al. (2009) identified two peaks of replication initiation events in *Drosophila* cell cultures. Analysis of data from different sources on replicon size, size of candidate replication initiation zones, and domains depleted of ORC binding sites suggests that a major proportion of the euchromatic part of the *Drosophila* genome is replicated by forks originating in early S-phase in INTs (Kolesnikova et al., 2018). Size of the rb-bands located between INTs rarely exceeds hundreds of kbs, and these regions often complete replication by forks entering from adjacent INTs before activation of the late origins. Only the longest rb-bands represented by regions of intercalary heterochromatin and pericentric heterochromatin regions initiate replication in the late S-phase in diploid cells. It is hypothesized that late origins never fire in polytene chromosomes (Lilly, Duronio, 2005; Lee et al., 2009).

Activation of origins occurs throughout all S-phase. However, origins can be divided in two classes according to their response to the system of DNA damage control, intra-S-phase checkpoint. Treatment of cells with hydroxyurea (HU) activates the intra-S-phase checkpoint. Origins capable of activation in the presence of HU are considered as early origins (Shirahige et al., 1998; Willis, Rhind, 2009). MacAlpine et al. (2010) mapped origins active in Kc167 cells after HU treatment. All identified HU-insensitive origins coincide with ORC binding sites. However, only 30 % of the ORC binding sites overlap HU-insensitive origins. It is unclear whether the remaining ORC binding sites correspond to late replication initiation sites or represent weak origins, which may be activated in a stochastic manner in a limited number of cells. In addition, any manipulations with the cell cycle and the inter-S-phase checkpoint can have a profound effect on the pattern of active origins (Kolesnikova et al., 2013).

Distinct properties of replication in *Drosophila* polytene chromosomes

Polytene chromosomes emerge in a modified cell cycle called endocycle. An endocycle lacks all steps of mitosis. Different tissues attain different ploidies. In *D. melanogaster* salivary gland cells typically achieve ~1,300C, whereas fat body cells ~256C and midgut enterocytes 32C (Shu et al., 2018). In salivary glands, the exit from the S-phase happens before the completion of the entire genome replication and approximately 30 % of the genome are underreplicated in polytene chromosomes (Lilly, Spradling, 1996; Royzman, Orr-Weaver, 1998; Doronkin et al., 2003; Lee et al., 2009; Zielke et al., 2013; Edgar et al., 2014; Shu et al., 2018). The underreplication points to a change in the system of DNA damage control. It is plausible that the intra S-phase checkpoint and activation of late origins are absent in polytene nuclei (Lilly, Duronio, 2005; Lee et al., 2009). The apoptosis mechanism is turned off in these cells (Hassel et al., 2014). Genes associated with DNA replication are expressed at relatively low levels in endocycling cells (Maqbool et al., 2010).

The *Drosophila* endocycle is driven by oscillation of Cyclin E/Cdk2 activity. Different levels of Cyclin E/Cdk2 activity control initiation and termination of endoreplication, as well as the origins reset (Lilly, Spradling, 1996; Edgar, Orr-Weaver, 2001; Lilly, Duronio, 2005; Edgar et al., 2014; Shu et al., 2018). Ectopic expression of *Cyclin E* under control of the *hsp70* promoter induces S-phase in salivary gland cells (Duronio, O'Farrell, 1995; Su, O'Farrell, 1998). The oscillation of Cyclin E-Cdk2 activity is controlled at many levels, including the transcriptional induction of the *Cyclin E* gene by E2F1 (Duronio, O'Farrell, 1995), the destruction of Cyclin E protein by SCFAgo E3 ubiquitin ligase (Moberg et al., 2001; Shcherbata et al., 2004; Zielke et al., 2011), and the oscillation of cyclin-dependent kinase inhibitor Dacapo (Hong et al., 2007; Swanson et al., 2015). The oscillation of Cyclin E-Cdk2 activity and, hence, endocycle regulation are sensitive to external factors, such as nutrient levels (Britton, Edgar, 1998). External factors act via a network of signaling and transcription factors. Well-studied upstream regulators of the endocycle include growth factors, in particular, insulin-like peptides, and epidermal growth factor (EGF) signaling. The downstream effectors in these pathways are PI3K, AKT, target of rapamycin (TOR), forkhead box O (FOXO), RAS, and MAPK (Britton et al., 2002; Saucedo et al., 2003; Pierce et al., 2004; Grewal et al., 2005; Demontis, Perrimon, 2009; Edgar et al., 2014). The nonuniform distribution of external factors can asynchronously induce endoreplication in cells within an organ. Nuclei in salivary gland are not synchronized, but the distribution of endocycle stages is not random (Rudkin, 1973; Smith, Orr-Weaver, 1991; Kolesnikova et al., 2013). For example, nuclei in proximal and distal ends of the *D. melanogaster* salivary gland differ by one or two rounds of endoreplication. Some of these differences are mediated by transcription factor Sunspot (Ssp). It promotes the expression of E2F1 and PCNA in *D. melanogaster* salivary glands. In the proximal region of the gland, the *sunspot* gene activity is inhibited by wingless (Wnt) signaling, resulting in reduced E2F1 expression, a longer endocycle, and decreased ploidy (Taniue et al., 2010). Ssp is negatively regulated by Arm (Taniue et al., 2010). The *arm*-GAL4 driver induces mosaic expression

of the reporter gene *LacZ* in salivary glands of *Drosophila* third instar larvae (Kolesnikova et al., 2005). Perhaps the Arm mosaic expression contributes to the asynchrony of endocycles in salivary gland. In *D. melanogaster* ovarian follicle cells, differentiation-associated endocycle exit is regulated through a combination of Notch signaling, Ecdysone receptor, and transcriptional repressor Tramtrack (Sun et al., 2008).

It has been shown in *Drosophila* that the duration of endocycles and, presumably, the S-phase increases with increase of ploidy (Rudkin, 1972, 1973). The speed of replication forks in salivary gland polytene chromosomes of *Drosophila* and *Rhynchosciara* is significantly lower than in diploid cells, especially during the late S-phase (Meneghini, Cordeiro, 1972; Cordeiro, Meneghini, 1973; Steinemann, 1981a, b; Lakhotia, Sinha, 1983). Studies of underreplication profiles in different *Drosophila* tissues have shown that underreplication zones can be tissue-specific, but they correspond to repressive chromatin areas lacking origins in other polytene tissues where they are completely replicated. The authors suggest that the difference in underreplication is caused by tissue-specific variation of replication rates in these regions (Hua et al., 2018).

The speed of replication forks in *Drosophila* polytene chromosomes depends on SUUR protein (Sher et al., 2012; Nordman et al., 2014). In $SuUR^{ES}$ polytene chromosomes the late replicating regions complete replication earlier (Zhimulev et al., 2003). Both in salivary glands and in diploid cells, all chromosome regions corresponding to rb-bands are enriched in SUUR protein (Kolesnikova et al., 2018). Targeting of SUUR protein to an early replication region on a polytene chromosome leads to the late completion of replication in this site (Pokholkova et al., 2015). It can be assumed that SUUR plays an important role in the replication delay associated with rb-bands at least in salivary gland chromosomes (Kolesnikova et al., 2018). Continuous *SuUR* overexpression in salivary glands starting from early embryogenesis leads to a miniature gland, indicative of a strong suppression of DNA replication in the cells (Volkova et al., 2003; Zhimulev et al., 2003). Overexpression of *SuUR* in the middle of the third larval instar under the salivary gland specific *Sgs3*-GAL4 driver leads to disappearance of nuclei at G-phase and accumulation of nuclei at early S-phase, when labelling is observed only in early replicating regions (Kolesnikova et al., 2011). The *SuUR* overexpression does not abolish incorporation of 5-ethynyl-2′-deoxyuridine (EdU) into the polytene chromosomes, hence DNA replication continues, albeit at a very low rate.

SuUR overexpression exerts a significantly less pronounced effect in diploid cells. There is no obvious effect on eye formation, but apoptosis was observed in the wing disc, and the wing shape was distorted. The effect is enhanced by mutations in the genes encoding E2F and PCNA proteins, suggesting that ectopic SUUR affects replication in diploid cells (Volkova et al., 2003). The cause of the more pronounced effect of SUUR on DNA replication in polytene chromosomes compared with diploid cells remains to be found out.

O.V. Posukh et al. (2015) suggests that the SUUR-dependent replication delay is important for a proper post-replication chromatin assembly. The *SuUR* mutation changes levels of H3K27me3 and H3K9me3 in late replicating regions of polytene chromosomes of salivary glands and pseudonurse cells of otu^{11} mutants (Koryakov et al., 2011; Sher et al., 2012; Posukh et al., 2017). According to O.V. Posukh et al. (2017), SUUR in *Drosophila* chromosomes is involved in the epigenetic inheritance of H3K27me3 in those regions where Polycomb complexes only establish, but do not maintain H3K27me3 silencing (Posukh et al., 2015, 2017).

The Rif1 protein, a candidate repressor of replication, is another important factor involved in the regulation of replication timing in *D. melanogaster*. Rif1 recruits phosphatase PP1 to multiple chromosomal sites, where the latter can dephosphorylate replicative helicase and block premature replication of heterochromatin sequences (Sreesankar et al., 2015; Seller, O'Farrell, 2018). During amplification of chorion genes in ovarian follicle cells, Rif1 is localized in active replication forks in a partially SUUR-dependent manner, where it directly regulates replication fork progression (Munden et al., 2018). It appears that in polytene chromosomes SUUR protein binds to chromatin domains where histone H1 is accumulated during the S-phase (Andreyeva et al., 2017). SUUR binds to the components of replication forks (Kolesnikova et al., 2013; Nordman et al., 2014) and attracts Rif1 protein (Munden et al., 2018). In turn, Rif1 protein attracts PP1 phosphatase, which dephosphorylates replicative helicase, thereby controlling the activation time of the origins and the replication fork rate (Sreesankar et al., 2015; Seller, O'Farrell, 2018).

Conclusion

In the last twenty years, the progress in understanding of replication was associated with advances in high-throughput techniques. Comparison of genomic replication profiles revealed significant similarity between *Drosophila* and other well-studied eukaryotic organisms, such as humans. Early replication is often confined to intensely transcribed gene-dense regions characterized by multiple replication initiation sites. Features of DNA replication in *Drosophila* may be explained by the compact genome (Petrov, 2002). The most important feature of replication in polytene chromosomes is their low replication rate and the dependence of S-phase duration on many factors: external and internal, local and global. In *D. melanogaster* polytene chromosomes, the speed of replication forks is affected by SUUR and Rif1 proteins. It is not known yet how universal the mechanisms associated with these factors are, but their study is very promising.

References

Achary P.M., Majumdar K., Duttagupta A., Mukherjee A.S. Replication of DNA in larval salivary glands of *Drosophila* after *in vivo* synchronization. Chromosoma. 1981;82(4):505-514.

Allison L., Arndt-Jovin D.J., Gratzner H., Ternynck T., Robert-Nicoud M. Mapping of the pattern of DNA replication in polytene chromosome from *Chironomus thummi* using monoclonal anti-bromodeoxyuridine antibodies. Cytometry. 1985;6(6):584-590.

Amabis J.M. Induction of DNA synthesis in *Rhynchosciara angelae* salivary gland. Cell Differ. 1974;3(4):199-207.

Andreyeva E.N., Bernardo T.J., Kolesnikova T.D., Lu X., Yarinich L.A., Bartholdy B.A., Guo X., Posukh O.V., Healton S., Willcockson M.A., Pindyurin A.V., Zhimulev I.F., Skoultchi A.I., Fyodorov D.V. Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in *Drosophila*. Genes Dev. 2017;31(6):603-616.

Bedo D.G. Patterns of polytene-chromosome replication in *Simulium ornatipes* (Diptera: Simuliidae). Genetica. 1982;59:9-21.

Belyakin S.N., Babenko V.N., Maksimov D.A., Shloma V.V., Kvon E.Z., Belyaeva E.S., Zhimulev I.F. Gene density profile re-

veals the marking of late replicated domains in the *Drosophila melanogaster* genome. Chromosoma. 2010;119(6):589-600.
Berezney R., Dubey D.D., Huberman J.A. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. Chromosoma. 2000;108:471-484.
Borowiec J.A., Schildkraut C.L. Open sesame: activating dormant replication origins in the mouse immunoglobulin heavy chain (Igh) locus. Curr. Opin. Cell Biol. 2011;23:284-292.
Boulos R.E., Drillon G., Argoul F., Arneodo A., Audit B. Structural organization of human replication timing domains. FEBS Lett. 2015; 589(20 Pt A):2944-2957.
Britton J.S., Edgar B.A. Environmental control of the cell cycle in *Drosophila*: nutrition activates mitotic and endoreplicative cells by distinct mechanisms. Development. 1998;125:2149-2158.
Britton J.S., Lockwood W.K., Li L., Cohen S.M., Edgar B.A. *Drosophila*'s insulin/PI3-kinase pathway coordinates cellular metabolism with nutritional conditions. Dev. Cell. 2002;2:239-249.
Cadoret J.C., Meisch F., Hassan-Zadeh V., Luyten I., Guillet C., Duret L., Quesneville H., Prioleau M.N. Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Natl. Acad. Sci. USA. 2008;105(41):15837-15842.
Cayrou C., Coulombe P., Vigneron A., Stanojcic S., Ganier O., Peiffer I., Rivals E., Puy A., Laurent-Chabalier S., Desprat R., Mechali M. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. 2011;21:1438-1449.
Chagin V.O., Stear J.H., Cardoso M.C. Organization of DNA replication. Cold Spring Harb. Perspect. Biol. 2010;2:a000737.
Cordeiro M., Meneghini R. The rate of DNA replication in the polytene chromosomes of *Rhynchosciara angelae*. J. Mol. Biol. 1973;78(2): 261-274.
Danieli G.A., Rodinò E. Larval moulting cycle and DNA synthesis in *Drosophila hydei* salivary glands. Nature. 1967; 213(5074):424-425.
Debatisse M., Le Tallec B., Letessier A., Dutrillaux B., Brison O. Common fragile sites: mechanisms of instability revisited. Trends Genet. 2012;28(1):22-32.
Demontis F., Perrimon N. Integration of Insulin receptor/Foxo signaling and dMyc activity during muscle growth regulates body size in *Drosophila*. Development. 2009;136:983-993.
Doronkin S., Djagaeva I., Beckendorf S.K. The COP9 signalosome promotes degradation of Cyclin E during early *Drosophila* oogenesis. Dev. Cell. 2003;4(5):699-710.
Durkin S.G., Glover T.W. Chromosome fragile sites. Annu. Rev. Genet. 2007;41:169-192.
Duronio R.J., O'Farrell P.H. Developmental control of the G1 to S transition in *Drosophila*: cyclin E is a limiting downstream target of E2F. Genes Dev. 1995;9(12):1456-1468.
Edgar B.A., Orr-Weaver T.L. Endoreplication cell cycles: more for less. Cell. 2001;105(3):297-306.
Edgar B.A., Zielke N., Gutierrez C. Endocycles: a recurrent evolutionary innovation for post-mitotic cell growth. Nat. Rev. Mol. Cell Biol. 2014;15(3):197-210.
Gibert J.M., Karch F. The Polycomb group protein CRAMPED is involved with TRF2 in the activation of the histone H1 gene. Chromosoma. 2011;120:297-307.
Gilbert D.M. Replication timing and transcriptional control: beyond cause and effect. Curr. Opin. Cell Biol. 2002;14(3):377-383.
Gilbert D.M., Cohen S.N. Bovine papilloma virus plasmids replicate randomly in mouse fibroblasts throughout S phase of the cell cycle. Cell. 1987;50(1):59-68.
Gilbert D.M., Takebayashi S.I., Ryba T., Lu J., Pope B.D., Wilson K.A., Hiratani I. Space and time in the nucleus: developmental control of replication timing and chromosome architecture. Cold Spring Harb. Symp. Quant. Biol. 2010;75:143-153.
Gillespie P.J., Blow J.J. Clusters, factories and domains: the complex structure of S-phase comes into focus. Cell Cycle. 2010;9:3218-3226.
Grewal S.S., Li L., Orian A., Eisenman R.N., Edgar B.A. Myc-dependent regulation of ribosomal RNA synthesis during *Drosophila* development. Nat. Cell Biol. 2005;7(3):295-302.
Guilbaud G., Rappailles A., Baker A., Chen C.L., Arneodo A., Goldar A., d'Aubenton-Carafa Y., Thermes C., Audit B., Hyrien O. Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. PLoS Comput. Biol. 2011;7(12):e1002322.
Hassel C., Zhang B., Dixon M., Calvi B.R. Induction of endocycles represses apoptosis independently of differentiation and predisposes cells to genome instability. Development. 2014;141(1):112-123.
Hägele K. Prolongation of replication time after doubling of the DNA content of polytene chromosome bands of *Chironomus*. Chromosoma. 1976;55(3):253-258.
Hiratani I., Gilbert D.M. Replication timing as an epigenetic mark. Epigenetics. 2009;4(2):93-97.
Hiratani I., Ryba T., Itoh M., Rathjen J., Kulik M., Papp B., Fussner E., Bazett-Jones D.P., Plath K., Dalton S., Rathjen P.D., Gilbert D.M. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. 2010;20:155-169.
Hiratani I., Ryba T., Itoh M., Yokochi T., Schwaiger M., Chang C.W., Lyou Y., Townes T.M., Schubeler D., Gilbert D.M. Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol. 2008;6:e245.
Hong A., Narbonne-Reveau K., Riesgo-Escovar J., Fu H., Aladjem M.I., Lilly M.A. The cyclin-dependent kinase inhibitor Dacapo promotes replication licensing during *Drosophila* endocycles. EMBO. J. 2007;26(8):2071-2082.
Hua B.L., Bell G.W., Kashevsky H., Von Stetina J.R., Orr-Weaver T.L. Dynamic changes in ORC localization and replication fork progression during tissue differentiation. BMC Genomics. 2018;19(1):623.
Jackson D.A., Pombo A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. J. Cell Biol. 1998;140(6):1285-1295.
Jackson L.P., Reed S.I., Haase S.B. Distinct mechanisms control the stability of the related S-phase cyclins Clb5 and Clb6. Mol. Cell. Biol. 2006;26:2456-2466.
Julienne H., Zoufir A., Audit B., Arneodo A. Human genome replication proceeds through four chromatin states. PLoS Comput. Biol. 2013;9(10):e1003233.
Keyl H.-G. Duplikationen von untereinheiten der chromosomalen DNS wahrend der Evolution von *Chironomus thummi*. Chromosoma. 1965;17:139-180.
Keyl H.-G., Pelling C. Differentielle DNS-Replikation in den Speicheldrusen-chromosopmen von *Chironomus thummi*. Chromosoma. 1963;14:347-359.
Kolesnikova T.D., Goncharov F.P., Zhimulev I.F. Similarity in replication timing between polytene and diploid cells is associated with the organization of the *Drosophila* genome. PLoS One. 2018;13(4): e0195207.
Kolesnikova T.D., Makunin I.V., Volkova E.I., Pirrotta V., Belyaeva E.S., Zhimulev I.F. Functional dissection of the Suppressor of UnderReplication protein of *Drosophila melanogaster*: identification of domains influencing chromosome binding and DNA replication. Genetica. 2005;124:187-200.
Kolesnikova T.D., Posukh O.V., Andreyeva E.N., Bebyakina D.S., Ivankin A.V., Zhimulev I.F. *Drosophila* SUUR protein associates with PCNA and binds chromatin in a cell cycle-dependent manner. Chromosoma. 2013;122(1-2):55-66.
Kolesnikova T.D., Semeshin V.F., Andreyeva E.N., Zykov I.A., Kokoza E.B., Kalashnikova D.A., Belyaeva E.S., Zhimulev I.F. Induced decondensation of heterochromatin in *Drosophila melanogaster* polytene chromosomes under condition of ectopic expression of the *Supressor of underreplication* gene. Fly(Austin). 2011;5(3):181-190.
Koryakov D.E., Walther M., Ebert A., Lein S., Zhimulev I.F., Reuter G. The SUUR protein is involved in binding of SU(VAR)3-9 and me-

thylation of H3K9 and H3K27 in chromosomes of *Drosophila melanogaster*. Chromosome Res. 2011;19(2):235-249.

Koryakov D.E., Zhimulev I.F. DNA replication in nurse cell polytene chromosomes of *Drosophila melanogaster otu* mutants. Chromosoma. 2015;124:95-106.

Kozlova T., Semeshin V.F., Tretyakova I.V., Kokoza E.B., Pirrotta V., Grafodatskaya V.E., Belyaeva E.S., Zhimulev I.F. Molecular and cytogenetical characterization of the 10A1-2 band and adjoining region in the *Drosophila melanogaster* polytene X chromosome. Genetics. 1994;136:1063-1073.

Lakhotia S.C., Sinha P. Replication in *Drosophila* chromosomes. X. Two kinds of active replicons in salivary gland polytene nuclei and their relation to chromosomal replication patterns. Chromosoma. 1983;88(4):265-276.

Lebofsky R., Heilig R., Sonnleitner M., Weissenbach J., Bensimon A. DNA replication origin interference increases the spacing between initiation events in human cells. Mol. Biol. Cell. 2006;17:5337-5345.

Lee H.O., Davidson J.M., Duronio R.J. Endoreplication: polyploidy with purpose. Genes Dev. 2009;23:2461-2477.

Lilly M.A., Duronio R.J. New insights into cell cycle control from the *Drosophila* endocycle. Oncogene. 2005;24(17):2765-2775.

Lilly M.A., Spradling A.C. The *Drosophila* endocycle is controlled by Cyclin E and lacks a checkpoint ensuring S-phase completion. Genes Dev. 1996;10:2514-2526.

Lubelsky Y., Prinz J.A., DeNapoli L., Li Y., Belsky J.A., MacAlpine D.M. DNA replication and transcription programs respond to the same chromatin cues. Genome Res. 2014;24:1102-1114.

Ma H., Samarabandu J., Devdhar R.S., Acharya R., Cheng P.C., Meng C., Berezney R. Spatial and temporal dynamics of DNA replication sites in mammalian cells. J. Cell Biol. 1998;143:1415-1425.

MacAlpine H.K., Gordân R., Powell S.K., Hartemink A.J., MacAlpine D.M. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res. 2010;20:201-211.

MacAlpine D.M., Rodriguez H.K., Bell S.P. Coordination of replication and transcription along a *Drosophila* chromosome. Genes Dev. 2004;18(24):3094-3105.

Maqbool S.B., Mehrotra S., Kolpakas A., Durden C., Zhang B., Zhong H., Calvi B.R. Dampened activity of E2F1-DP and Myb-MuvB transcription factors in *Drosophila* endocycling cells. J. Cell Sci. 2010;123(Pt 23):4095-4106.

Meneghini R., Cordeiro M. DNA replication in polytene chromosomes of *Rhynchosciara angelae*. Cell Differ. 1972;1(3):167-177.

Mesner L.D., Valsakumar V., Karnani N., Dutta A., Hamlin J.L., Bekiranov S. Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res. 2011;21:377-389.

Moberg K.H., Bell D.W., Wahrer D.C., Haber D.A., Hariharan I.K. Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. Nature. 2001;413(6853):311-316.

Mulder M.P., van Duijn P., Gloor H.J. The replicative organization of DNA in polytene chromosomes of *Drosophila hydei*. Genetica. 1968;39(3):385-428.

Munden A., Rong Z., Sun A., Gangula R., Mallal S., Nordman J.T. Rif1 inhibits replication fork progression and controls DNA copy number in *Drosophila*. Elife. 2018;7:e39140.

Nordman J.T., Kozhevnikova E.N., Verrijzer C.P., Pindyurin A.V., Andreyeva E.N., Shloma V.V., Zhimulev I.F., Orr-Weaver T.L. DNA copy-number control through inhibition of replication fork progression. Cell Rep. 2014;9(3):841-849.

Norio P., Kosiyatrakul S., Yang Q., Guan Z., Brown N.M., Thomas S., Riblet R., Schildkraut C.L. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. Mol. Cell. 2005;20:575-587.

Peric-Hupkes D., Meuleman W., Pagie L., Bruggeman S.W., Solovei I., Brugman W., Gräf S., Flicek P., Kerkhoven R.M., van Lohuizen M., Reinders M., Wessels L., van Steensel B. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell. 2010;38(4):603-613.

Petrov D.A. DNA loss and evolution of genome size in *Drosophila*. Genetica. 2002;115(1):81-91.

Petryk N., Kahli M., d'Aubenton-Carafa Y., Jaszczyszyn Y., Shen Y., Silvain M., Thermes C., Chen C.L., Hyrien O. Replication landscape of the human genome. Nat. Commun. 2016;7:10208.

Pickersgill H., Kalverda B., de Wit E., Talhout W., Fornerod M., van Steensel B. Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. Nat. Genet. 2006;38(9):1005-1014.

Pierce S.B., Yost C., Britton J.S., Loo L.W., Flynn E.M., Edgar B.A., Eisenman R.N. dMyc is required for larval growth and endoreplication in *Drosophila*. Development. 2004;131(10):2317-2327.

Plaut W., Nash D., Fanning T. Ordered replication of DNA in polytene chromosomes of *Drosophila melanogaster*. J. Mol. Biol. 1966;16:85-93.

Pokholkova G.V., Koryakov D.E., Pindyurin A.V., Kozhevnikova E.N., Belyakin S.N., Andreyenkov O.V., Belyaeva E.S., Zhimulev I.F. Tethering of SUUR and HP1 proteins results in delayed replication of euchromatic regions in *Drosophila melanogaster* polytene chromosomes. Chromosoma. 2015;124(2):209-220.

Pope B.D., Hiratani I., Gilbert D.M. Domain-wide regulation of DNA replication timing during mammalian development. Chromosome Res. 2010;18(1):127-136.

Pope B.D., Ryba T., Dileep V., Yue F., Wu W., Denas O., Vera D.L., Wang Y., Hansen R.S., Canfield T.K., Thurman R.E., Cheng Y., Gülsoy G., Dennis J.H., Snyder M.P., Stamatoyannopoulos J.A., Taylor J., Hardison R.C., Kahveci T., Ren B., Gilbert D.M. Topologically associating domains are stable units of replication-timing regulation. Nature. 2014;515(7527):402-405.

Posukh O.V., Maksimov D.A., Laktionov P.P., Koryakov D.E., Belyakin S.N. Functional dissection of *Drosophila melanogaster* SUUR protein influence on H3K27me3 profile. Epigenetics Chromatin. 2017;10(1):56.

Posukh O.V., Maksimov D.A., Skvortsova K.N., Koryakov D.E., Belyakin S.N. The effects of SUUR protein suggest its role in repressive chromatin renewal during replication in *Drosophila*. Nucleus. 2015;6(4):249-253.

Prioleau M.N., MacAlpine D.M. DNA replication origins-where do we begin? Genes Dev. 2016;30:1683-1697.

Redfern C.P.F. DNA replication in polytene chromosomes: Similarity of termination patterns in somatic and germ-line derived polytene chromosomes of *Anopheles stephensi* liston (Diptera: Culicidae). Chromosoma. 1981;84:33-47.

Rhind N., Gilbert D.M. DNA replication timing. Cold Spring Harb. Perspect. Biol. 2013;5(8):a010132.

Royzman I., Orr-Weaver T.L. S phase and differential DNA replication during *Drosophila* oogenesis. Genes Cells. 1998;3:767-776.

Rudkin G.T. Replication in polytene chromosomes. Ed. W. Beermann. Results and Problems in Cell Differentiation. Berlin; Heidelberg; New York: Springer-Verlag, 1972;4:59-85.

Rudkin G.T. Cyclic synthesis of DNA in polytene chromosomes of Diptera. Eds. M. Balls, F. Billet. Cell Cycle in Development and Differentiation. London: Cambridge Univ. Press, 1973;279-292.

Saucedo L.J., Gao X., Chiarelli D.A., Li L., Pan D., Edgar B.A. Rheb promotes cell growth as a component of the insulin/TOR signalling network. Nat. Cell Biol. 2003;5(6):566-571.

Schübeler D., Scalzo D., Kooperberg C., van Steensel B., Delrow J., Groudine M. Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. Nat. Genet. 2002;32(3):438-442.

Schwaiger M., Stadler M.B., Bell O., Kohler H., Oakeley E.J., Schubeler D. Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. Genes Dev. 2009;23:589-601.

Seller C.A., O'Farrell P.H. Rif1 prolongs the embryonic S phase at the *Drosophila* mid-blastula transition. PLoS Biol. 2018;16(5):e2005687.

Shcherbata H.R., Althauser C., Findley S.D., Ruohola-Baker H. The mitotic-to-endocycle switch in *Drosophila* follicle cells is executed by Notch-dependent regulation of G1/S, G2/M and M/G1 cell-cycle transitions. Development. 2004;131(13):3169-3181.

Sher N., Bell G.W., Li S., Nordman J., Eng T., Eaton M.L., Macalpine D.M., Orr-Weaver T.L. Developmental control of gene copy number by repression of replication initiation and fork progression. Genome Res. 2012;22(1):64-75.

Sher N., von Stetina J.R., Bell G.W., Matsuura S., Ravid K., Orr-Weaver T.L. Fundamental differences in endoreplication in mammals and *Drosophila* revealed by analysis of endocycling and endomitotic cells. Proc. Natl. Acad. Sci. USA. 2013;110(23):9368-9373.

Shirahige K., Hori Y., Shiraishi K., Yamashita M., Takahashi K., Obuse C., Tsurimoto T., Yoshikawa H. Regulation of DNA-replication origins during cell-cycle progression. Nature. 1998;395(6702): 618-621.

Shu Z., Row S., Deng W.M. Endoreplication: The Good, the Bad, and the Ugly. Trends Cell. Biol. 2018;28(6):465-474.

Sinha P., Mishra A., Lakhotia S.C. Chromosomal organization of *Drosophila* tumours. I. Polytene chromosome organization and DNA synthesis in ovarian pseudonurse cells in *otu* mutants of *D. melanogaster*. Chromosoma. 1987;95:108-116.

Smith A.V., Orr-Weaver T.L. The regulation of the cell cycle during *Drosophila* embryogenesis: the transition to polyteny. Development. 1991;112:997-1008.

Spierer A., Spierer P. Similar level of polyteny in bands and interbands of *Drosophila* giant chromosomes. Nature. 1984;307:176-178.

Sreesankar E., Bharathi V., Mishra R.K., Mishra K. Drosophila *Rif1* is an essential gene and controls late developmental events by direct interaction with PP1-87B. Sci. Rep. 2015;5:10679.

Steinemann M. Chromosomal replication of *Drosophila virilis*. II. Organization of active origins in diploid brain cells. Chromosoma. 1981a;82(2):267-288.

Steinemann M. Chromosomal replication in *Drosophila virilis*. III. Organization of active origins in the highly polytene salivary gland cells. Chromosoma. 1981b;82(2):289-307.

Stocker A.J., Pavan C. The influence of ecdysterone on gene amplification, DNA synthesis, and puff formation in the salivary gland chromosomes of *Rhynchosciara hollaenderi*. Chromosoma. 1974; 45(3):295-319.

Su T.T., O'Farrell P.H. Chromosome association of minichromosome maintenance proteins in *Drosophila* endoreplication cycles. J. Cell. Biol. 1998;140(3):451-460.

Sun J., Smith L., Armento A., Deng W.M. Regulation of the endocycle/gene amplification switch by Notch and ecdysone signaling. J. Cell Biol. 2008;182(5):885-896.

Swanson C.I., Meserve J.H., McCarter P.C., Thieme A., Mathew T., Elston T.C., Duronio R.J. Expression of an S phase-stabilized version of the CDK inhibitor Dacapo can alter endoreplication. Development. 2015;142(24):4288-4298.

Taniue K., Nishida A., Hamada F., Sugie A., Oda T., Ui-Tei K., Tabata T., Akiyama T. Sunspot, a link between Wingless signaling and endoreplication in Drosophila. Development. 2010;137(10):1755-1764.

Vatolina T.Y., Boldyreva L.V., Demakova O.V., Demakov S.A., Kokoza E.B., Semeshin V.F., Babenko V.N., Goncharov F.P., Belyaeva E.S., Zhimulev I.F. Identical functional organization of nonpolytene and polytene chromosomes in *Drosophila melanogaster*. PLoS One. 2011;6:e25960.

Volkova E.I., Yurlova A.A., Kolesnikova T.D., Makunin I.V., Zhimulev I.F. Ectopic expression of the *Suppressor of Underreplication* gene inhibits endocycles but not the mitotic cell cycle in *Drosophila melanogaster*. Mol. Genet. Genomics. 2003;270:387-393.

White E.J., Emanuelsson O., Scalzo D., Royce T., Kosak S., Oakeley E.J., Weissman S., Gerstein M., Groudine M., Snyder M., Schübeler D. DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. Proc. Natl. Acad. Sci. USA. 2004;101(51):17771-17776.

Willis N., Rhind N. Regulation of DNA replication by the S-phase DNA damage checkpoint. Cell Div. 2009;4:13.

Woodfine K., Beare D.M., Ichimura K., Debernardi S., Mungall A.J., Fiegler H., Collins V.P., Carter N.P., Dunham I. Replication timing of human chromosome 6. Cell Cycle. 2005;4:172-176.

Zhimulev I.F. Genetic organization of polytene chromosomes. Adv. Genet. 1999;39:1-589.

Zhimulev I.F., Belyaeva E.S., Makunin I.V., Pirrotta V., Volkova E.I., Alekseyenko A.A., Andreyeva E.N., Makarevich G.F., Boldyreva L.V., Nanayev R.A., Demakova O.V. Influence of the *SuUR* gene on intercalary heterochromatin in *Drosophila melanogaster* polytene chromosomes. Chromosoma. 2003;111:377-398.

Zhimulev I.F., Zykova T.Y., Goncharov F.P., Khoroshko V.A., Demakova O.V., Semeshin V.F., Pokholkova G.V., Boldyreva L.V., Demidova D.S., Babenko V.N., Demakov S.A., Belyaeva E.S. Genetic organization of interphase chromosome bands and interbands in *Drosophila melanogaster*. PLoS One. 2014;9:e101631.

Zielke N., Edgar B.A., DePamphilis M.L. Endoreplication. Cold Spring Harb. Perspect. Biol. 2013;5(1):a012948.

Zielke N., Kim K.J., Tran V., Shibutani S.T., Bravo M.J., Nagarajan S., van Straaten M., Woods B., von Dassow G., Rottig C., Lehner C.F., Grewal S.S., Duronio R.J., Edgar B.A. Control of *Drosophila* endocycles by E2F and CRL4(CDT2). Nature. 2011;480(7375): 123-127.

Acknowledgements. This work was financially supported by the Russian Science Foundation (project No. 14_14_00934). The authors are grateful to reviewer's valuable comments, which improved the manuscript.

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised January 15, 2019. Accepted January 16, 2019.

Вавиловский журнал генетики и селекции. 2019;23(2):148-153
DOI 10.18699/VJ19.474

Polytene chromosomes reflect functional organization of the *Drosophila* genome

D.S. Sidorenko[1] ✉, T.Yu. Zykova[1], V.A. Khoroshko[1], G.V. Pokholkova[1], S.A. Demakov[1], J. Larsson[2], E.S. Belyaeva[1], I.F. Zhimulev[1, 3]

[1] Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
[2] Department of Molecular Biology, Umeå University, Umeå, Sweden
[3] Laboratory of Structural, Functional and Comparative Genomics of the Novosibirsk State University, Novosibirsk, Russia
✉ e-mail: demidova.daria@mcb.nsc.ru

Polytene chromosomes of *Drosophila melanogaster* are a convenient model for studying interphase chromosomes of eukaryotes. They are giant in size in comparison with diploid cell chromosomes and have a pattern of cross stripes resulting from the ordered chromatid arrangement. Each region of polytene chromosomes has a unique banding pattern. Using the model of four chromatin types that reveals domains of varying compaction degrees, we were able to correlate the physical and cytological maps of some polytene chromosome regions and to show the main properties of genetic and molecular organization of bands and interbands, that we describe in this review. On the molecular map of the genome, the interbands correspond to decompacted aquamarine chromatin and 5′ ends of ubiquitously active genes. Gray bands contain lazurite and malachite chromatin, intermediate in the level of compaction, and, mainly, coding parts of genes. Dense black transcriptionally inactive bands are enriched in ruby chromatin. Localization of several dozens of interbands on the genome molecular map allowed us to study in detail their architecture according to the data of whole genome projects. The distribution of proteins and regulatory elements of the genome in the promoter regions of genes localized in the interbands shows that these parts of interbands are probably responsible for the formation of open chromatin that is visualized in polytene chromosomes as interbands. Thus, the permanent genetic activity of interbands and gray bands and the inactivity of genes in black bands are the basis of the universal banding pattern in the chromosomes of all *Drosophila* tissues. The smallest fourth chromosome of *Drosophila* with an atypical protein composition of chromatin is a special case. Using the model of four chromatin states and fluorescent *in situ* hybridization, its cytological map was refined and the genomic coordinates of all bands and interbands were determined. It was shown that, in spite of the peculiarities of this chromosome, its band organization in general corresponds to the rest of the genome. Extremely long genes of different *Drosophila* chromosomes do not fit the common scheme, since they can occupy a series of alternating bands and interbands (up to nine chromosomal structures) formed by parts of these genes.
Key words: *Drosophila melanogaster*; polytene chromosomes; interphase chromosomes; four chromatin state model; fluorescent *in situ* hybridization; genetic organization; bands and interbands of chromosomes.

For citation: Sidorenko D.S., Zykova T.Yu., Khoroshko V.A., Pokholkova G.V., Demakov S.A., Larsson J., Belyaeva E.S., Zhimulev I.F. Polytene chromosomes reflect functional organization of the *Drosophila* genome. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):148-153. DOI 10.18699/VJ19.474

Политенные хромосомы отражают функциональную организацию генома *Drosophila*

Д.С. Сидоренко[1] ✉, Т.Ю. Зыкова[1], В.А. Хорошко[1], Г.В. Похолкова[1], С.А. Демаков[1], Я. Ларссон[2], Е.С. Беляева[1], И.Ф. Жимулёв[1, 3]

[1] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Факультет молекулярной биологии Университета Умео, Умео, Швеция
[3] Новосибирский национальный исследовательский государственный университет, лаборатория структурной, функциональной и сравнительной геномики, Новосибирск, Россия
✉ e-mail: demidova.daria@mcb.nsc.ru

Политенные хромосомы *Drosophila melanogaster* – удобная модель для изучения интерфазных хромосом эукариот. Им свойственны гигантские размеры в сравнении с хромосомами диплоидных клеток и поперечная исчерченность, возникающая в связи с упорядоченным расположением хроматид. Каждый район политенных хромосом обладает уникальным дисковым рисунком. С использованием модели четырех типов хроматина, которая выявляет домены различной степени компактизации, удалось соотнести физическую и цитологическую карты некоторых районов политенных хромосом и показать основные свойства генетической и молекулярной организации дисков и междисков, описанию которых посвящен данный обзор. Междискам на молекулярной карте генома соответствуют декомпактный aquamarine хроматин и 5′-концы повсеместно активных генов. Серые диски содержат промежуточный по уровню компактизации lazurite и malachite хроматин и в основном кодирующие части генов. Черные плотные транскрипционно неактив-

ные диски обогащены ruby хроматином. Локализация нескольких десятков междисков на молекулярной карте генома позволила по данным полногеномных проектов подробно исследовать их архитектуру. Распределение белков и регуляторных элементов генома в промоторных районах генов, локализованных в междисках, показывает, что эти части междисков, вероятно, отвечают за формирование хроматина открытого типа, который визуализируется в политенных хромосомах как междиски. Таким образом, постоянная генетическая активность междисков и серых дисков и неактивность генов в черных дисках лежат в основе универсального дискового рисунка в хромосомах всех тканей дрозофилы. Особый случай представляет самая маленькая – четвертая – хромосома дрозофилы с нетипичной белковой композицией хроматина. При помощи модели четырех состояний хроматина и флуоресцентной *in situ* гибридизации была уточнена ее цитологическая карта и определены геномные координаты всех дисков и междисков. Показано, что, несмотря на особенности этой хромосомы, ее дисковая организация в целом соответствует остальному геному. Выбиваются из общей схемы особо длинные гены разных хромосом дрозофилы, которые могут занимать целую серию чередующихся дисков и междисков (до девяти структур), образованных частями этих генов.
Ключевые слова: *Drosophila melanogaster*; политенные хромосомы; интерфазные хромосомы; модель состояний хроматина; флуоресцентная *in situ* гибридизация; генетическая организация; диски и междиски хромосом.

Introduction

The *Drosophila melanogaster* polytene chromosomes have been used as a model for studying interphase chromosomes for many years, since this is the only type of chromosomes that can be visualized along their entire length in the interphase nucleus. These giant chromosomes are formed in specialized cells (e. g. salivary gland cells and nurse cells) due to numerous replication cycles without subsequent segregation of daughter chromatids, which remain closely connected to each other. The chromomeric pattern of interphase chromosomes, therefore, becomes contrastingly expressed in the form of alternating dark and light transverse stripes. Densely packed chromatin in polytene chromosomes forms black bands, moderately condensed regions form loose gray bands, and light decondensed regions are interbands.

The model of four chromatin states

In recent years, a large amount of whole genome data on the distribution of various histone modifications, chromatin proteins, and regulatory elements of the *Drosophila* genome has been accumulated (modENCODE Consortium..., 2010). These data have a marking corresponding to the physical genomic coordinates. To analyze the interrelation of the functional domains formed by these characteristics, with the structural, morphological, and genetic organization of *Drosophila* chromosomes, it is necessary to know the correspondence of polytene chromosome cytological structures (bands and interbands) to their genomic coordinates. To achieve this goal, we developed a computer model of four chromatin states, which has a good correspondence to the cytological structures of polytene chromosomes (Zhimulev et al., 2014).

Earlier, about 30 interbands have been localized on the physical map of the *Drosophila* genome using different methods. Using modENCODE data, a set of proteins associated with DNA sequences, corresponding to these interbands in cell cultures, was determined (Demakov et al., 1993, 2011; Semeshin et al., 2008; Vatolina et al., 2011a, b; Zhimulev et al., 2014). Based on the distribution of these proteins, using a computer algorithm, our model of four chromatin states, conventionally named as colors, was developed (Zhimulev et al., 2014). Further, in order to avoid intersections with other works, the colors of chromatin states were renamed according to the names of the minerals (Boldyreva et al., 2017).

Aquamarine chromatin (former cyan) contains maximum amount of open chromatin proteins, that are the basis of the model, and occupies about 13 % of the genome. This chromatin state corresponds to 5′ UTRs of ubiquitously active genes (Zhimulev et al., 2014; Zykova et al., 2018). All the interbands with determined genomic coordinates contain this chromatin type (Zhimulev et al., 2014). Ruby chromatin (former magenta) occupies about 50 % of the genome and it lacks proteins characteristic of open chromatin (Zhimulev et al., 2014). It is localized in black transcriptionally inactive bands, mainly containing tissue- and stage-specific genes, including the largest bands of intercalary heterochromatin (Khoroshko et al., 2016; Kolesnikova et al., 2018). Lazurite (former blue) and malachite (former green) chromatin states have an intermediate degree of compaction and mainly correspond to gray bands and transition zones between the interbands and black bands (Zhimulev et al., 2014; Boldyreva et al., 2017; Khoroshko et al., 2018). Lazurite chromatin is enriched with coding parts of genes, and malachite – with introns and intergenic spacers (Zykova et al., 2018). Thus, the model of four chromatin types confirms the identity of the organization of polytene chromosomes and chromosomes of diploid cells. This review shows examples of successful application of this model in determining the localization of bands and interbands on the physical map of the *Drosophila* genome.

The interbands are decompacted structures containing promoters of ubiquitously active genes

As mentioned above, aquamarine chromatin mainly corresponds to the interbands of polytene chromosomes containing promoters of ubiquitously active genes (Zhimulev et al., 2014). In the recent study by T.Yu. Zykova and co-authors (2019), the molecular and genetic organization of 33 interband sequences located on the physical map of the *Drosophila* genome was studied in detail. It turned out that the interbands have diverse genetic structure. The majority of the interbands studied contain one gene with one transcription start site, and the remaining interbands contain one gene with several alternative promoters, two or more unidirectional genes and genes with "head-to-head" orientation. In addition, there are complexly organized interbands with three or more genes having unidirectional or bidirectional orientation (Zykova et al., 2019).

Interbands containing unidirectional and bidirectional genes were studied in detail. The distribution of various factors that may influence the formation of open chromatin in the interbands is of particular interest. In this work, using modENCODE data, it was shown that in the interband sequences containing a single gene, the peak of CHRIZ and BEAF-32 insulator protein localization (Vogelmann et al., 2014) is located at the transcription start site, and the peak of RNA polymerase II lies 200 bp upstream of this site. In the interbands containing promoters localized "head-to-head", these insulator proteins lie between the transcription initiation sites, and the peaks of RNA polymerase II correspond to these two gene promoters (Zykova et al., 2019). Short noncoding RNAs, the products of the paused RNA polymerase II (Nechaev et al., 2010), replication complex protein ORC2 (Eaton et al., 2011), DNAse I hypersensitive sites (Kharchenko et al., 2011), *P*-element insertions (FlyBase r.5.57), and "broad" promoters (Hoskins et al., 2011) in the interbands under study are located in the regions of transcription start sites. Thus, they have one peak in the interbands with one gene, and two peaks in the interbands containing two opposite directed genes (Zykova et al., 2019).

Despite the fact that the average interband size is about 3.5 kb, it was shown that various open chromatin elements, characteristic of interbands, do not occupy the entire sequence of aquamarine chromatin fragments corresponding to these structures. They have narrow localization in regulatory regions of genes beginning in the interbands (Zykova et al., 2019). In addition, the location of 5′ gene ends with ubiquitous expression in aquamarine chromatin and its high enrichment in ORC2 protein (Zhimulev et al., 2014; Zykova et al., 2018) confirms the fact that the interbands combine the processes of transcription initiation and replication. The authors conclude that regulatory region of an active gene is apparently critical for the formation of an open chromatin domain, which is visualized as a light interband in polytene chromosomes (Zykova et al., 2019).

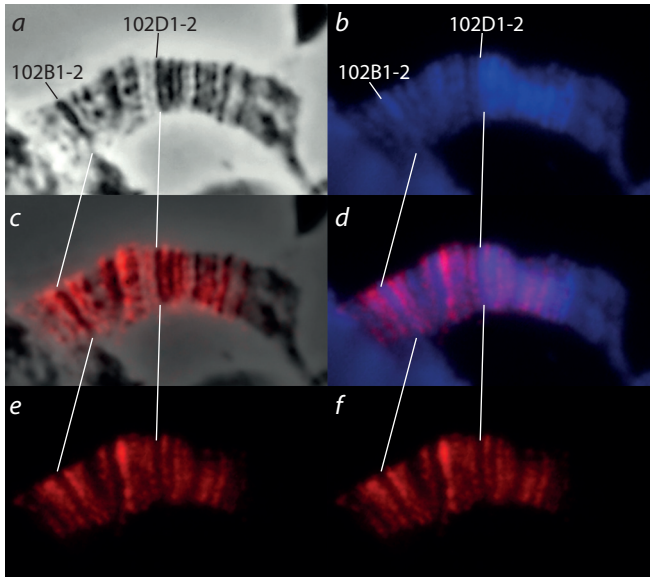
Atypical cases of gene localization in the structures of polytene chromosomes

To sum up, it has been shown that the banding pattern is a universal principle of both polytene chromosome and diploid cell chromosome organization, and housekeeping genes are embedded in two interphase chromosome structures: their promoters are localized in the interbands, and the structural parts lie in adjacent loose bands (Zhimulev et al., 2014). In addition, it has recently been discovered that long *D. melanogaster* genes can occupy several morphological structures of polytene chromosomes (Zhimulev et al., 2016; Khoroshko et al., 2019). For example, using fluorescent *in situ* hybridization (FISH), it was shown that long introns of the ubiquitous active genes *dlg1* (~18 kb) and *CG43867* (~106 kb), interrupted by short rare exons, form loose gray bands 10B8-9 and 1D1-2, respectively, and their 5′ and 3′ regulatory regions lie in the decompacted interbands of the X chromosome (Zhimulev et al., 2016). Thus, it has been found that in some cases polytene chromosome bands can be formed from the material of only one part of a gene, namely, long non-coding introns of actively transcribed genes, which allowed the authors to conclude that there is a differential pattern of chromatin condensation of gene parts during its activation (Zhimulev et al., 2016). A special case in studying gene localization in the morphological structures of polytene chromosomes is the *dunce* (*dnc*) gene, the size of which (167.3 kb) is 25-fold longer than the average length of a *D. melanogaster* gene (Khoroshko et al., 2019). The *dnc* gene has 17 transcripts responsible for various biological functions. About 94 % of the *dnc* transcript is occupied by introns, in which eight short genes are localized. The gene is mostly represented by inactive ruby and malachite chromatin states (93 %), which is consistent with the total size of introns. The majority of gene transcript 5′ ends corresponds to open aquamarine chromatin. To determine the localization of the *dnc* gene on polytene chromosome preparations, the authors performed FISH analysis using three probes from aquamarine chromatin corresponding to the 5′ and 3′ UTRs of the longest transcript of this gene and an alternative promoter of a group of transcripts. According to the hybridization results, the *dnc* gene is localized within nine cytological structures, namely, five interbands and four bands (3C7/C8-3D1 – 2/D3-4). The results do not correspond with the previously dominated ideas about the localization of genes in only one structure (Khoroshko et al., 2019). The material of introns, which constitute the largest part of the *dnc* gene length, forms not only loose gray and dense dark bands, but also interbands, in which chromatin is open for replication (Khoroshko et al., 2019).

The fourth chromosome is a specific domain of *Drosophila* chromatin

Mapping of the fourth chromosome, the smallest chromosome in the *D. melanogaster* genome, is of particular interest due to its unusual organization. The euchromatic arm of this chromosome, which looks like a dot on metaphase spreads, becomes visible after numerous endocycles in salivary gland cells. On the map by C.B. Bridges (1935), it corresponds to cytological sections 101 and 102, but there is no detailed cytological map. This *Drosophila* chromatin domain is unusual since it has high gene density and, at the same time, possesses heterochromatin characteristics (Riddle et al., 2009). Absence of recombination under normal conditions in the fourth chromosome of *D. melanogaster* leads to the accumulation of repeated sequences and mobile genome elements (Slawson et al., 2006). Chromatin in this domain shows high H3K9me enrichment (Riddle, Elgin, 2006), and this modification is mainly catalyzed by dSETDB1 histone methyltransferase specific for this chromosome (Seum et al., 2007; Figueiredo et al., 2012). The protein HP1a, which recognizes this mark and binds to typical heterochromatin regions, is also associated with the polytenised part of the fourth chromosome (James et al., 1989). A unique property of the *D. melanogaster* fourth chromosome is the protein POF (painting of fourth) that is localized mainly in the interbands (Larsson et al., 2001). Here we show the localization of POF on the fourth chromosome of the *SuUR Su(var)3-9* double mutant *D. melanogaster* larval salivary glands (Figure).

It was shown that POF and HP1a bind to the fourth chromosome interdependently and there is a mechanism balancing the action of POF and HP1 proteins, which provides fine-tuning of gene expression on this chromosome (Johansson et al., 2007a, b). It has been hypothesized that HP1a initially binds active gene promoters with high affinity regardless of



POF localization on the polytene fourth chromosome of *D. melanogaster*.
We used *SuUR*^{ES} *Su(var)3-9*^{06} double mutant flies because of suppressed underreplication for better spreading of the *Drosophila* fourth polytene chromosome on preparations (Andreyeva et al., 2007; Demakova et al., 2007). The preparations were made according to B. Czermin et al. (2002).
a – phase contrast image of the fourth chromosome; *b* – DAPI staining; *c* – merged phase contrast and POF staining; *d* – merged DAPI and POF staining; *e*, *f* – POF staining. Antibodies are the same as in J. Larsson et al. (2004). POF decorates predominantly the interbands and gray bands of the fourth chromosome.

H3K9me, and then it propagates through H3K9 methylation and loop contacts with chromatin containing this mark (Figueiredo et al., 2012). Recently, an optimal autonomous POF target has been identified. It consists of a gene and a block of X chromosome-specific 1.688 satellites downstream of the gene. The fourth chromosome seems to interact with POF due to the cooperative effect of a large number of suboptimal combinations of genes with repeats (Kim et al., 2018).

The study of the fourth chromosome at the cytological level is complicated by its small size, a large number of gray bands with an intermediate level of compaction, and frequent ectopic contacts of its distal end with the chromocenter. In our recent work, a fine mapping of all bands and interbands of this chromosome on the physical map of the *Drosophila* genome was carried out using the four chromatin state model (Zhimulev et al., 2014) and FISH (Sidorenko et al., 2018). In this study, to obtain preparations with good morphology of the fourth chromosome, we used a fly stock with suppressed underreplication, containing mutations in the *SuUR* and *Su(var)3-9* genes and linked X and Y chromosomes. Due to the matching of the fourth chromosome cytological map to the genomic coordinates, the distribution of various chromatin properties in bands and interbands was investigated using data from whole genome projects. It was shown that the interbands contain aquamarine chromatin and 5′ UTRs of ubiquitously active genes, gray bands contain lazurite chromatin and coding parts of these genes, and black bands contain ruby chromatin and are polygenic, or also contain coding parts of genes (Sidorenko et al., 2018). The distribution of various characteristics in the domains of active and repressed chromatin of the fourth chromosome was studied. POF (modENCODE data; Lundberg et al., 2013) and HP1a (modENCODE data) proteins, providing a special organization of this chromosome, are predominantly localized in aquamarine (interbands) and lazurite (gray bands) chromatin states (Sidorenko et al., 2018). Only ruby chromatin is enriched in the modified histone H3K27me3 (Sher et al., 2012), which corresponds to the data on other chromosomes (Sidorenko et al., 2018). DNase I hypersensitive sites, ORC2 protein and *P*-elements (modENCODE data) are predominantly located in open aquamarine chromatin, while the fourth chromosome-specific element *1360* (Kholodilov et al., 1988) occupies the band chromatin types (Sidorenko et al., 2018). Compared to the rest of the genome, the fourth chromosome contains half the size of compact ruby chromatin, which is associated with a large number of loose gray bands and the absence of intercalary heterochromatin bands. Notably, despite the peculiarities of the fourth chromosome, its band organization corresponds to the organization of the rest of the *D. melanogaster* genome (Sidorenko et al., 2018).

Conclusion

Thus, *D. melanogaster* polytene chromosomes together with the four chromatin state model (Zhimulev et al., 2014; Boldyreva et al., 2017; Zykova et al., 2018) provide a good model for studying the organization of interphase chromosomes. Using this system, we confirmed that the banding pattern is the general principle of interphase chromosome organization (Zhimulev et al., 2014). The model of four chromatin states in combination with FISH allows matching the morphological chromosome structures to the genomic coordinates to investigate them further using the data of genome-wide projects. The general regularity of molecular and genetic organization of interphase chromosome cytological structures is that the interbands contain mainly open aquamarine chromatin and 5′ UTRs of housekeeping genes, gray bands are enriched in lazurite and malachite chromatin states and coding parts of genes, black bands and intercalary heterochromatin bands contain compact ruby chromatin and tissue- and stage-specific genes (Zhimulev et al., 2014; Khoroshko et al., 2016; Zykova et al., 2018). This rule holds true for the organization of the fourth chromosome of *Drosophila* (Sidorenko et al., 2018). In addition, there are some atypical cases of band and interband genetic organization, with one long gene occupying up to nine cytological structures (Zhimulev et al., 2016; Khoroshko et al., 2019).

References

Andreyeva E.N., Kolesnikova T.D., Demakova O.V., Mendez-Lago M., Pokholkova G.V., Belyaeva E.S., Rossi F., Dimitri P., Villasante A., Zhimulev I.F. High-resolution analysis of *Drosophila* heterochromatin organization using *SuUR Su(var)3-9* double mutants. Proc. Natl. Acad. Sci USA. 2007;104(31):12819-12824.

Boldyreva L.V., Goncharov F.P., Demakova O.V., Zykova T.Y., Levitsky V.G., Kolesnikov N.N., Pindyurin A.V., Semeshin V.F., Zhimulev I.F. Protein and genetic composition of four chromatin types in *Drosophila melanogaster* cell lines. Curr. Genomics. 2017;18(2): 214-226.

Bridges C.B. Cytological data on chromosome four of *Drosophila melanogaster*. Trans. Dyn. Dev. 1935;10:463-473.

Czermin B., Melfi R., McCabe D., Seitz V., Imhof A., Pirrotta V. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 me-

thyltransferase activity that marks chromosomal Polycomb sites. Cell. 2002;111(2):185-196.

Demakov S.A., Semeshin V.F., Zhimulev I.F. Cloning and molecular genetic analysis of *Drosophila melanogaster* interband DNA. Mol. Gen. Genet. 1993;238(3):437-443.

Demakov S.A., Vatolina T.Y., Babenko V.N., Semeshin V.F., Belyaeva E.S., Zhimulev I.F. Protein composition of interband regions in polytene and cell line chromosomes of *Drosophila melanogaster*. BMC Genomics. 2011;12(1):566.

Demakova O.V., Pokholkova G.V., Kolesnikova T.D., Demakov S.A., Andreyeva E.N., Belyaeva E.S., Zhimulev I.F. The SU(VAR)3-9/HP1 complex differentially regulates the compaction state and degree of underreplication of X chromosome pericentric heterochromatin in *Drosophila melanogaster*. Genetics. 2007;175(2):609-620.

Eaton M.L., Prinz J.A., MacAlpine H.K., Tretyakov G., Kharchenko P.V., MacAlpine D.M. Chromatin signatures of the *Drosophila* replication program. Genome Res. 2011;21(2):164-174.

Figueiredo M.L., Philip P., Stenberg P., Larsson J. HP1a recruitment to promoters is independent of H3K9 methylation in *Drosophila melanogaster*. PLoS Genet. 2012;8(11):e1003061.

Hoskins R.A., Landolin J.M., Brown J.B., Sandler J.E., Takahashi H., Lassmann T., Yu C., Booth B.W., Zhang D., Wan K.H., Yang L., Boley N., Andrews J., Kaufman T.C., Graveley B.R., Bickel P.J., Carninci P., Carlson J.W., Celniker S.E. Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. Genome Res. 2011;21(2):182-192.

James T.C., Eissenberg J.C., Craig C., Dietrich V., Hobson A., Elgin S.C. Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of *Drosophila*. Eur. J. Cell Biol. 1989;50(1):170-180.

Johansson A.M., Stenberg P., Bernhardsson C., Larsson J. Painting of fourth and chromosome-wide regulation of the 4th chromosome in *Drosophila melanogaster*. EMBO J. 2007a;26(9):2307-2316.

Johansson A.M., Stenberg P., Pettersson F., Larsson J. POF and HP1 bind expressed exons, suggesting a balancing mechanism for gene regulation. PLoS Genet. 2007b;3(11):e209.

Kharchenko P.V., Alekseyenko A.A., Schwartz Y.B., Minoda A., Riddle N.C., Ernst J., Sabo P.J., Larschan E., Gorchakov A.A., Gu T., Linder-Basso D., Plachetka A., Shanower G., Tolstorukov M.Y., Luquette L.J., Xi R., Jung Y.L., Park R.W., Bishop E.P., Canfield T.K., Sandstrom R., Thurman R.E., MacAlpine D.M., Stamatoyannopoulos J.A., Kellis M., Elgin S.C., Kuroda M.I., Pirrotta V., Karpen G.H., Park P.J. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. Nature. 2011;471(7339):480-485.

Kholodilov N.G., Bolshakov V.N., Blinov V.M., Solovyov V.V., Zhimulev I.F. Intercalary heterochromatin in *Drosophila*. III. Homology between DNA sequences from the Y chromosome, bases of polytene chromosome limbs, and chromosome 4 of *D. melanogaster*. Chromosoma. 1988;97(3):247-253.

Khoroshko V.A., Levitsky V.G., Zykova T.Y., Antonenko O.V., Belyaeva E.S., Zhimulev I.F. Chromatin heterogeneity and distribution of regulatory elements in the late-replicating intercalary heterochromatin domains of *Drosophila melanogaster* chromosomes. PLoS One. 2016;11(6):e0157147.

Khoroshko V.A., Pokholkova G.V., Zykova T.Yu., Osadchiy I.S., Zhimulev I.F. Gene *dunce* localization in the polytene chromosome of *Drosophila melanogaster* long span batch of adjacent chromosomal structures. Dokl. Biochem. Biophys. 2019;484(4):1-4.

Khoroshko V.A., Zykova T.Yu., Popova O.O., Zhimulev I.F. Border structure of intercalary heterochromatin bands of *Drosophila melanogaster* polytene chromosomes. Dokl. Biochem. Biophys. 2018;479(1):114-117.

Kim M., Ekhteraei-Tousi S., Lewerentz J., Larsson J. The X-linked 1.688 satellite in *Drosophila melanogaster* promotes specific targeting by painting of fourth. Genetics. 2018;208(2):623-632.

Kolesnikova T.D., Goncharov F.P., Zhimulev I.F. Similarity in replication timing between polytene and diploid cells is associated with the organization of the *Drosophila* genome. PLoS One. 2018;13(4):e0195207.

Larsson J., Chen J.D., Rasheva V., Rasmuson-Lestander A., Pirrotta V. Painting of fourth, a chromosome-specific protein in *Drosophila*. Proc. Natl. Acad. Sci. USA. 2001;98(11):6273-6278.

Larsson J., Svensson M.J., Stenberg P., Mäkitalo M. Painting of fourth in genus *Drosophila* suggests autosome-specific gene regulation. Proc. Natl. Acad. Sci. USA. 2004;101(26):9728-9733.

Lundberg L.E., Kim M., Johansson A.M., Faucillion M.L., Josupeit R., Larsson J. Targeting of Painting of fourth to roX1 and roX2 proximal sites suggests evolutionary links between dosage compensation and the regulation of the fourth chromosome in *Drosophila melanogaster*. G3 (Bethesda). 2013;3(8):1325-1334.

modENCODE consortium, Roy S., Ernst J., Kharchenko P.V., Kheradpour P., Negre N., Eaton M.L., Landolin J.M., Bristow C.A., Ma L., Lin M.F., Washietl S., Arshinoff B.I., Ay F., Meyer P.E., Robine N., Washington N.L., Di Stefano L., Berezikov E., Brown C.D., Candeias R., Carlson J.W., Carr A., Jungreis I., Marbach D., Sealfon R., Tolstorukov M.Y., Will S., Alekseyenko A.A., Artieri C., Booth B.W., Brooks A.N., Dai Q., Davis C.A., Duff M.O., Feng X., Gorchakov A.A., Gu T., Henikoff J.G., Kapranov P., Li R., MacAlpine H.K., Malone J., Minoda A., Nordman J., Okamura K., Perry M., Powell S.K., Riddle N.C., Sakai A., Samsonova A., Sandler J.E., Schwartz Y.B., Sher N., Spokony R., Sturgill D., van Baren M., Wan K.H., Yang L., Yu C., Feingold E., Good P., Guyer M., Lowdon R., Ahmad K., Andrews J., Berger B., Brenner S.E., Brent M.R., Cherbas L., Elgin S.C., Gingeras T.R., Grossman R., Hoskins R.A., Kaufman T.C., Kent W., Kuroda M.I., Orr-Weaver T., Perrimon N., Pirrotta V., Posakony J.W., Ren B., Russell S., Cherbas P., Graveley B.R., Lewis S., Micklem G., Oliver B., Park P.J., Celniker S.E., Henikoff S., Karpen G.H., Lai E.C., MacAlpine D.M., Stein L.D., White K.P., Kellis M. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. Science. 2010;330(6012):1787-1797.

Nechaev S., Fargo D.C., dos Santos G., Liu L., Gao Y., Adelman K. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. Science. 2010;327(5963):335-338.

Riddle N.C., Elgin S.C. The dot chromosome of *Drosophila*: insights into chromatin states and their change over evolutionary time. Chromosome Res. 2006;14(4):405-416.

Riddle N.C., Shaffer C.D., Elgin S.C. A lot about a little dot – lessons learned from *Drosophila melanogaster* chromosome 4. Biochem. Cell. Biol. 2009;87(1):229-241.

Semeshin V.F., Demakov S.A., Shloma V.V., Vatolina T.Y., Gorchakov A.A., Zhimulev I.F. Interbands behave as decompacted autonomous units in *Drosophila melanogaster* polytene chromosomes. Genetica. 2008;132(3):267-279.

Seum C., Reo E., Peng H., Rauscher F.J., Spierer P., Bontron S. *Drosophila* SETDB1 is required for chromosome 4 silencing. PLoS Genet. 2007;3(5):e76.

Sher N., Bell G.W., Li S., Nordman J., Eng T., Eaton M.L., MacAlpine D.M., Orr-Weaver T.L. Developmental control of gene copy number by repression of replication initiation and fork progression. Genome Res. 2012;22(1):64-75.

Sidorenko D.S., Sidorenko I.A., Zykova T.Yu., Kolesnikova T.D., Goncharov F.P., Zhimulev I.F. Molecular and genetic organization of bands and interbands in the dot chromosome of *Drosophila melanogaster*. Chromosome 2018: International Conference. August 20–24, 2018, Novosibirsk, Russia. Abstracts. Novosibirsk: Institute of Molecular and Cellular Biology SB RAS; Novosibirsk State University, 2018;75.

Slawson E.E., Shaffer C.D., Malone C.D., Leung W., Kellmann E., Shevchek R.B., Craig C.A., Bloom S.M., Bogenpohl J., Dee J., Morimoto E.T., Myoung J., Nett A.S., Ozsolak F., Tittiger M.E., Zeug A., Pardue M.L., Buhler J., Mardis E.R., Elgin S.C. Comparison of dot chromosome sequences from *D. melanogaster* and

D. virilis reveals an enrichment of DNA transposon sequences in heterochromatic domains. Genome Biol. 2006;7(2):R15.

Vatolina T.Yu., Boldyreva L.V., Demakova O.V., Demakov S.A., Kokoza E.B., Semeshin V.F., Babenko V.N., Goncharov F.P., Belyaeva E.S., Zhimulev I.F. Identical functional organization of nonpolytene and polytene chromosomes in *Drosophila melanogaster*. PLoS One. 2011a;6(10):e25960.

Vatolina T.Iu., Demakov S.A., Semeshin V.F., Makunin I.V., Babenko V.N., Beliaeva E.S., Zhimulev I.F. Identification and molecular genetic characterization of the polytene chromosome interbands in *Drosophila melanogaster*. Russ. J. Genet. 2011b;47(5):521-532.

Vogelmann J., Le Gall A., Dejardin S., Allemand F., Gamot A., Labesse G., Cuvier O., Nègre N., Cohen-Gonsaud M., Margeat E., Nöllmann M. Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the *Drosophila* genome. PLoS Genet. 2014;10(8):e1004544.

Zhimulev I.F., Boldyreva L.V., Demakova O.V., Poholkova G.V., Khoroshko V.A., Zykova T.Y., Lavrov S.A., Belyaeva E.S. *Drosophila* polytene chromosome bands formed by gene introns. Dokl. Biochem. Biophys. 2016;466(1):57-60.

Zhimulev I.F., Zykova T.Y., Goncharov F.P., Khoroshko V.A., Demakova O.V., Semeshin V.F., Pokholkova G.V., Boldyreva L.V., Demidova D.S., Babenko V.N., Demakov S.A., Belyaeva E.S. Genetic organization of interphase chromosome bands and interbands in *Drosophila melanogaster*. PLoS One. 2014;9(7):e101631.

Zykova T.Yu., Levitsky V.G., Belyaeva E.S., Zhimulev I.F. Polytene chromosomes – a portrait of functional organization of the *Drosophila* genome. Curr. Genomics. 2018;19(3):179-191.

Zykova T.Yu., Levitsky V.G., Zhimulev I.F. Architecture of promoters of house-keeping genes in polytene chromosome interbands of *Drosophila melanogaster*. Dokl. Biochem. Biophys. 2019;485(1):1-6.

http://www.modecode.org/

Acknowledgements. The authors gratefully acknowledge the resources provided by the "Molecular and Cellular Biology" core facility of the IMCB SB RAS. Financial support: Russian Science Foundation grant No. 14-14-00934.

Conflict of interest. The authors declare no conflict of interest.

Received November 28, 2018. Revised December 14, 2018. Accepted December 18, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):154-159
DOI 10.18699/VJ19.475

ADF1 and BEAF-32 chromatin proteins affect nucleosome positioning and DNA decompaction in 61C7/C8 interband region of *Drosophila melanogaster* polytene chromosomes

M.B. Schwartz (Berkaeva), T.E. Pankova, S.A. Demakov ✉

Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
✉ e-mail: demakov@mcb.nsc.ru

The formation of interphase chromosomes is a multi-level process in which DNA is compacted several thousand-fold by association with histones and non-histone proteins. The first step of compaction includes the formation of nucleosomes – the basic repeating units of chromatin. Further packaging occurs due to DNA binding to histone H1 and non-histone proteins involved in enhancer-promoter and insulator interactions. Under these conditions, the genome retains its functionality due to the dynamic and uneven DNA compaction along the chromatin fiber. Since the DNA compaction level affects the transcription activity of a certain genomic region, it is important to understand the interplay between the factors acting at different levels of the packaging process. *Drosophila* polytene chromosomes are an excellent model system for studying the molecular mechanisms that determine DNA compaction degree. The unevenness of DNA packaging along the chromatin fiber is easily observed along these chromosomes due to their large size and specific banding pattern. The purpose of this study was to figure out the role of two non-histone regulatory proteins, ADF1 and BEAF-32, in the DNA packaging process from nucleosome positioning to the establishment of the final chromosome structure. We studied the impact of mutations that affect ADF1 and BEAF-32 binding sites on the formation of 61C7/C8 interband – one of the decompacted regions of *Drosophila* polytene chromosomes. We show that such mutations led to the collapse of an interband, which was accompanied with increased nucleosome stability. We also find that ADF1 and BEAF-32 binding sites are essential for the rescue of lethality caused by the null allele of *bantam* microRNA gene located in the region 61C7/C8.
Key words: *Drosophila*; nucleosome; polytene chromosome; interband; transcription.

For citation: Schwartz (Berkaeva) M.B., Pankova T.E., Demakov S.A. ADF1 and BEAF-32 chromatin proteins affect nucleosome positioning and DNA decompaction in 61C7/C8 interband region of *Drosophila melanogaster* polytene chromosomes. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):154-159. DOI 10.18699/VJ19.475

Хроматиновые белки ADF1 и BEAF-32 влияют на позиционирование нуклеосом и упаковку ДНК междиска 61C7/C8 политенных хромосом *Drosophila melanogaster*

М.Б. Шварц (Беркаева), Т.Е. Панкова, С.А. Демаков ✉

Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ e-mail: demakov@mcb.nsc.ru

Формирование интерфазной хромосомы представляет собой многоуровневый процесс, в результате которого ДНК упаковывается в тысячи раз. На первом этапе упаковки образуются нуклеосомы – базовые повторяющиеся единицы хроматина. Дальнейшая упаковка происходит за счет связывания ДНК с гистоном H1 и негистоновыми белками, участвующими в ближних и дальних энхансер-промоторных и инсуляторных взаимодействиях. При этом функциональность генома сохраняется за счет динамичной и неравномерной упаковки ДНК вдоль хромосомы, что проявляется уже на нуклеосомном уровне. Несмотря на долгую историю изучения процесса упаковки ДНК в интерфазном ядре, до сих пор до конца не ясно, от чего зависит степень упаковки разных участков ДНК и какое влияние оказывают друг на друга разные уровни упаковки. Превосходной модельной системой для изучения молекулярных механизмов, определяющих степень упаковки ДНК, являются политенные хромосомы слюнных желез личинок дрозофилы. За счет больших размеров и характерного диск/междискового рисунка они позволяют легко наблюдать неравномерность упаковки ДНК вдоль хромосом. В настоящей работе мы исследовали, какую роль играют негистоновые регуляторные белки ADF1 и BEAF-32 в позиционировании нуклеосом и формировании междиска 61C7/C8 – одного из декомпактных районов политенных хромосом. ADF1 – специфический транскрипционный фактор, а BEAF-32 – инсуляторный белок, ассоциированный с междисками. С использованием трансгенных линий мы показали, что

мутации сайтов связывания ADF1 или BEAF-32 приводят к тому, что трансген теряет способность формировать междиск в новом генетическом окружении. Кроме того, мутации нарушают нуклеосомную организацию трансгена, характеризующуюся повышением стабильности нуклеосом. Мы обнаружили, что сайты связывания ADF1 и BEAF-32 необходимы для спасения нуль-аллеля *bantam* – жизненно важного гена микроРНК, расположенного в районе 61C7/C8. Таким образом, мы можем проследить связь между степенью упаковки ДНК, нуклеосомной организацией и функцией конкретного участка интерфазной хромосомы.
Ключевые слова: дрозофила; нуклеосома; политенная хромосома; междиск; транскрипция.

Introduction

Eukaryotic DNA is organized in a compact nucleoprotein complex called chromatin. Nucleosomes is the basic unit of chromatin and consists of 147 bp of DNA wrapped around a histone octamere. Nucleosome arrays form a so-called "10-nm fibril" which is further folded into higher order chromatin structures by histone H1 and a large number of functionally different non-histone proteins. The resulting structure constitutes an eukaryotic chromosome. Despite the long history of chromosome formation, mutual influence between the different DNA compaction levels is still unclear. On the one hand, nucleosome positioning determines the availability of DNA for the binding of non-histone proteins that regulate transcription and replication. On the other hand, non-histone proteins affect nucleosome positioning: they can destroy a nucleosome by displacing histone octamer or recruit ATP-dependent remodeling complexes that move the nucleosomes along the chromosomes. Moreover, nucleosome arrays are thought to self-organize against the potential barrier introduced by DNA-bound non-histone protein complexes due to the steric exclusions between adjacent nucleosome particles (Chereji et al., 2016; Chereji, Clark, 2018).

Drosophila polytene chromosomes provide a convenient model of interphase chromatin and allow to study the interplay between DNA compaction, chromatin structure, and transcription activity. Polytene chromosomes are formed in salivary glands of third instar larvae in the process of endoreplication and display a characteristic repetitive pattern of densely compacted bands interleaving with decompacted regions, called interbands. The compaction level of a given DNA region is determined by its function and the corresponding chromatin state. Interbands are formed by open chromatin and usually contain gene promoters, enhancers and insulators. Furthermore, interbands coincide with the nucleosome-free regions or contain weakly positioned nucleosomes (Zhimulev et al., 2014). These features make interbands a unique tool for studying the role of non-histone regulatory proteins in nucleosome positioning and the formation of chromomeric pattern of interphase chromosomes.

Here, we study the role of two regulatory non-histone proteins – ADF1 and BEAF-32 – in the nucleosome positioning and interband formation in the region 61C7/C8 of polytene chromosomes. This region was chosen because the DNA sequence sufficient to establish and maintain the decompacted state of the 61C7/C8 interband was determined earlier (Semeshin et al., 2008). Interband 61C7/C8 encompasses a *bantam* miRNA gene and regulatory elements controlling its expression, including ADF1 and BEAF-32 binding sites (Brennecke et al., 2003). ADF1 is a transcription factor and BEAF-32 is an insulator protein that is also associated with the transcription activation. Both of the proteins are involved in the establishment and maintenance of local chromatin state (Jiang et al., 2009; Orsi et al., 2014). Previously, we showed that ADF1 and BEAF-32 localized to the 61C7/C8 interband in salivary gland polytene chromosomes (Berkaeva et al., 2009). BEAF-32 is known as an interband-specific protein (Zhimulev et al., 2014) and its binding to DNA is crucial for the polytene chromosome structure (Gilbert et al., 2006). Mapping of nucleosomes in *Drosophila* cell cultures suggests that binding of BEAF-32 to DNA causes the formation of potential barriers that determine the positioning of neighboring nucleosomes (Chereji et al., 2016).

We found that ADF1 and BEAF-32 binding sites were indispensable for 61C7/C8 interband formation. Although the mutations of both binding sites caused the interband disappearance, the associated changes in nucleosomal organization were distinct. Therefore, different molecular mechanisms underlying those changes were proposed. Additionally, we report here that ADF1 and BEAF-32 play an essential role in the regulation of *bantam* expression.

Materials and methods

Fly stocks. To obtain transgenic flies, we used a stock with an attP-site located in the 10A1-2 region of the X chromosome (Andreenkov et al., 2016). Transgenic stock "4.7" carried an insertion of the original unmutated "4.7" fragment, stock "ADF" contained the insertion of the "4.7" fragment with mutated ADF-1 binding site, and "BEAF" contained the insertion of the "4.7" fragment with mutated BEAF-32 binding site.

All transgenic flies had y^1 and $Df(1)w^{67c23}$ mutations in the X chromosome and a lethal deletion $ban^{\Delta 1}$ maintained against TM6B balancer.

Control "yw" flies had wild-type chromosomes except for the X which carried y^1 and $Df(1)w^{67c23}$ mutations.

Mutations in protein binding sites. The DNA fragment sufficient to form the 61C7/C8 interband was 4709 bp in length (3L: 637635-642343, release = r6.23) and was named as the "4.7" fragment. It encompassed ADF1 and BEAF-32 binding sites, *bantam* miRNA hairpin and two putative promoters of the *bantam* gene (Fig. 1).

ADF1 binding site in the "4.7" fragment consisted of three motifs which corresponded to the consensus (England et al., 1992). Two overlapping motifs were removed by *Sal*I-*Sph*I excision and the third one was disrupted by replacing 7 nucleotides with a stretch of As. Shown below is a portion of the "4.7" fragment that includes the ADF1 binding site:

…ATGcgacttcaaacatagttcggcatcgaaactttctagcacaccga
cacacatacgaacgcgatccagccgacacacacacacacacgcacgcagc
cacacacttaagcgactttcgaaaggtacaacttttacgaagtcgctgcct
cggccgctgtgcagccgacgcc<u>actgccgctgccgctgtcgctgcctc</u>tg
TCGACTTCGAATTCCAACGCCAAGATGAAAGATC
GGCGCAAAAGAAAAGAAATATTCATTCAGTA
AAATTTGA<u>TAGCTGC</u>**AAAAAAA**<u>AGCCGCATGG</u>…

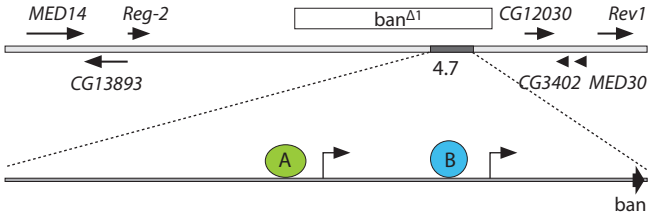


Fig. 1. Molecular and genetic organization of the 61C7/C8 region of 3L-chromosome.

Position of the "4.7" fragment on the genetic map is marked with a dark gray rectangle. Ovals "A" and "B" indicate ADF1 and BEAF-32 binding sites, respectively. Curved arrows denote putative promoters of the *bantam* gene (Brennecke et al., 2003; Qian et al., 2011). A thick black arrow indicates the position of bantam miRNA hairpin. White rectangle in the upper part of the figure indicates ban$^{\Delta 1}$ deletion.

The first nucleotide of the sequence above is 3L: 639461 (*Drosophila* genome release r6.23). The nucleotides included in the consensus motifs are underlined, deleted nucleotides are shown in lowercase, and inserted nucleotides are shown in bold.

BEAF-32 binding sites are characterized by the presence of CGATA motifs that are positioned in a certain way (Jiang et al., 2009). Two CGATA motifs on the forward and reverse DNA strands were deleted by *Cla*I excision:

GAATATCGATatgatatcgatGGGA

The first nucleotide of the sequence above is 3L: 641006 (release r6.23). Only forward DNA strand is presented. Forward and reverse CGATA motifs are underlined and deleted nucleotides are shown in lowercase.

Original "4.7" fragment and the mutated variants were cloned into pUni-mod vector (Andreenkov et al., 2016) using *Kpn*I and *Not*I restriction sites. The constructs obtained were used to create transgenic flies.

Cytology. Squashed preparations of the salivary gland polytene chromosomes were made by the standard technique and analyzed by phase-contrast microscopy, as described earlier (Zhimulev et al., 1982).

Nucleosome profiling. To study the nucleosome positioning, 100 pairs of salivary glands of third-instar larvae were extracted in PBS and then treated with 700 units of MNase (BioLabs; M0247S) at +25 °C for 3 min. Under those conditions, mono-, di- and tri-nucleosomes were obtained. Thus, the nucleosomes with different sensitivity to MNase hydrolysis (Chereji et al., 2016) were in our nucleosomal profile. Nucleosomal DNA was isolated by phenol extraction. Mononucleosomal DNA was eluted from 1.5 % agarose gel using a DNA gel extraction kit (BioSilica) and analyzed by Real-Time PCR with a set of primers spanning the entire "4.7" fragment. The concentration of DNA was determined using the method of calibration curves. To establish calibration curves, genomic DNA was isolated from the brain ganglia of the same larvae that were used for salivary gland isolation. The copy number of DNA along the entire chromosome is the same in diploid brain ganglia, while in salivary glands different parts of the same chromosome can be polytenized to a different extent. An amplicon from the *hsp26* locus corresponding to a precisely positioned nucleosome (Thomas, Elgin, 1988) was used as an internal control to take into account the pipetting errors. Nucleosome stability was assessed as the ratio of nucleosomal DNA to the genomic DNA isolated from salivary glands not treated with MNase.

Real-Time PCR data were quantified as follows. The concentrations of mononucleosomal ($M_{Amplicon}$) and untreated DNA ($U_{Amplicon}$) for each amplicon were normalized by the concentrations of the *hsp26* amplicon in mononucleosomal (M_{hsp26}) and untreated DNA (U_{hsp26}), respectively:

$$\frac{M_{Amplicon1}}{M_{hsp26}} = M_{\frac{A1}{hsp26}}; \frac{U_{Amplicon1}}{U_{hsp26}} = U_{\frac{A1}{hsp26}}.$$

The normalized concentration data for mono-nucleosomes were divided by the normalized concentration for untreated DNA to determine the representation of the amplicon in mononucleosome fraction:

$$\frac{M_{\frac{A1}{hsp26}}}{U_{\frac{A1}{hsp26}}} = M_1.$$

Some of the amplicons demonstrated nearly zero representation (for example, see Fig. 3, amplicons # 8, 9). This indicated that the underlying sequence was not fully protected from MNase hydrolysis. However, it cannot be excluded that the nucleosome is only slightly shifted relative to the amplicon, therefore one of the primers does not fit into the nucleosomal DNA.

All experiments were performed as two biological repeats, each consisted of three technical replicas. The representation data obtained in different transgenic flies were compared separately for each amplicon. The significance of differences was assessed using Student's criterion. The nucleosomal profiles of transgenes were studied in ban$^{\Delta 1}$ deletion background to eliminate the possible contribution of the native 61C7/C8 interband in the 3L chromosome (Brennecke et al., 2003).

Results and discussion

ADF1 and BEAF-32 binding sites are important for interband formation. The 4.7 kb DNA fragment from the 61C7/C8 region (hereinafter referred to as the "4.7" fragment) is able to form the interband when relocated into a new genomic position (Semeshin et al., 2008). This means that all the factors determining the interband formation are within this piece of DNA. We propose that the initial step of the interband formation involves DNA binding by some regulatory proteins which further recruit chromatin remodeling complexes capable of establishing and maintening the "open" chromatin structure. Putative binding sites for several proteins were identified within the "4.7" sequence, but only ADF1 and BEAF-32 were precisely mapped to the 61C7/C8 interband in salivary glands (Berkaeva et al., 2009). Therefore we asked whether these proteins may have a role in interband formation. We used transgenic assays to answer this question. The original "4.7" fragment and its modifications with disrupted ADF1 and BEAF-32 binding sites were inserted in the10A1-2 region which forms a thick band in the X polytene chromosome (Fig. 2, upper panel). All the transgenes were inserted exactly in the same position using phiC31-mediated attP-attB recombination system. In doing so, we were able distinguish the effects of mutations from the influence of genomic environment.

Cytological analysis of polytene chromosome squashed preparations showed that insertion of the original "4.7" fragment caused the formation of an ectopic interband that split the 10A1-2 band (see Fig. 2). It is worth noting that the insertion of the transgene was located on the very edge of the thick 10A1-2 band. Thus, the observed splitting of a thin gray band from the distal part of the original 10A1-2 band indicated that the ectopic interband was formed by the transgene material. It should be noted that no 10A1-2 band splitting was observed in the polytene chromosomes of original flies with the attP site in 10A1-2 region (see Fig. 2). Similarly, the mutated variants of the "4.7" fragment with disrupted ADF1 or BEAF-32 binding sites were unable to split the 10A1-2 band (see Fig. 2).

The absence of ectopic interband in 10A1-2 region of the polytene chromosomes of "ADF" and "BEAF" flies could be caused by a disturbance of the nucleosome positioning in the transgenes. It was shown previously that binding of ADF1 to the target promoters contributes to the subsequent binding of the transcription factor GAGA (GAF) (Talamillo et al., 2004; Orsi et al., 2014), which recruits NURF remodeling complex (Tsukiyama et al., 1994). It was also shown that depletion of NURF subunits leads to a large-scale redistribution of nucleosomes and disruption of the chromosome organization in *Drosophila* S2 cells (Moshkin et al., 2012). It was shown that BEAF-32 physically interacts with the insulator protein CP190 (Vogelmann et al., 2014), which functionally interacts with NURF301, a subunit of the NURF complex (Kwon et al., 2016). It should be noted that GAF, CP190, NURF, and BEAF-32 have been shown to bind the native 61C7/C8 region in *Drosophila* S2 cells (www.modencode.org).

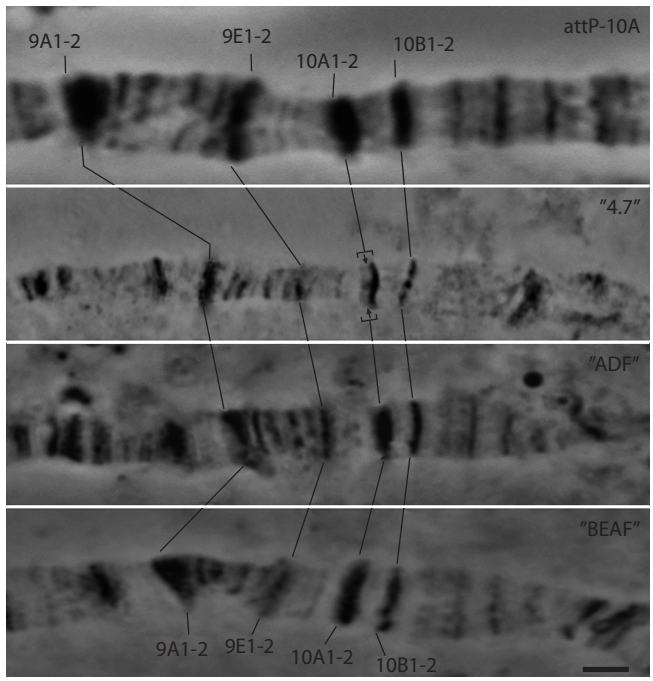


Fig. 2. Morphology of the X chromosome 10A1-2 region in transgenic larvae.

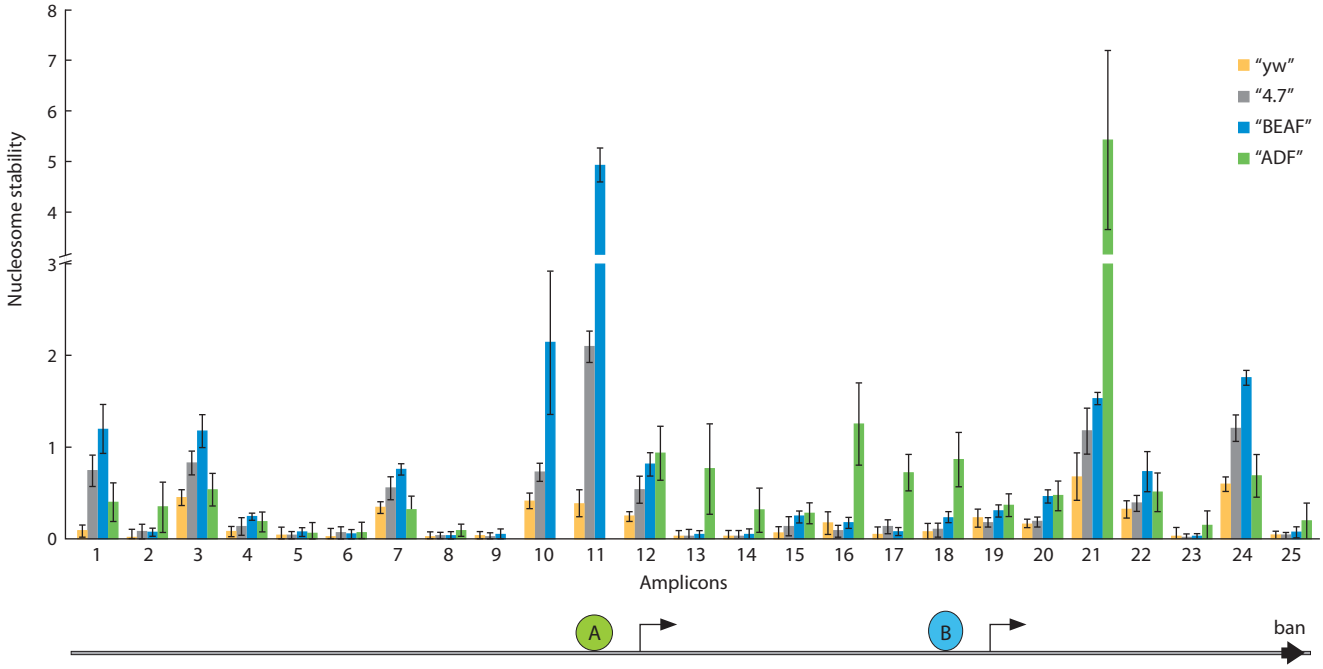
Stock names are indicated at the right upper corner of each panel. The top panel shows the chromosome of the original stock with the attP site in 10A1-2 region. Arrows on the "4.7" panel mark the ectopic interband. Scale bar corresponds to 5 μm.

Mutations in the ADF1 and BEAF-32 binding sites lead to disturbance of nucleosomal organization. To understand the role of nucleosomal organization in the formation of ectopic interband, we studied nucleosome profiles of all transgenes in larval salivary glands. First of all, we asked whether the transposition of "4.7" fragment into the new genomic locus was accompanied with perturbation in nucleosomal organization. We compared nucleosome profile of the "4.7" transgene with the profile of native 61C7/C8 region in the 3L chromosome of "yw" larvae. We found that the stability of nucleosomes increased in the "4.7" transgene compared to the native 61C7/C8 interband (Fig. 3). That was unexpected since the "4.7" transgene formed an interband visually indistinguishable from the native one (see Fig. 2). Genetic environment likely affects the nucleosome positioning in the inserted "4.7" fragment, although this impact is not visually manifested at the level of DNA packaging.

Next, we compared the nucleosome profile in the "ADF" transgene with the profile obtained for the "4.7" fragment. Deletion in the "ADF" transgene affects the area overlapped by the amplicons 9, 10, 11, so the data for them were unavailable. We found that the stability of nucleosomes between the amplicons 12 and 21 increased while the stability of nucleosomes at the edges of "ADF" transgene decreased in comparison with the "4.7" (see Fig. 3). This is consistent with our idea that NURF binding is disrupted in the "ADF" transgene. It was previously shown that depletion of NURF subunits in the cultured *Drosophila* cells led to increased nucleosome stability at the center of the areas occupied by the complex and to decreased nucleosome stability at the edges of those areas. The nucleosomes from the periphery seemed to gather at the place freed from the NURF complex (Moshkin et al., 2012). This is exactly was what we observed for the "ADF" transgene.

Nucleosome stability also increased in the "BEAF" transgene compared with "4.7" (see Fig. 3), however there was no shift of the nucleosomes from the periphery to the center, like in "ADF". Apparently the increase in nucleosome stability in this case was caused by the other factors than altered NURF binding. One of the explanation is based on the fact that BEAF-32 interacts with Chromator protein (Vogelmann et al., 2014) which attracts JIL-1 kinase that phosphorylates the S10 residue of histone H3. Depletion of JIL-1 or Chromator leads to the specific condensation of polytene chromosome interbands (Rath et al., 2006). We proposed that mutation of BEAF-32 binding site in "BEAF" transgene indirectly caused the displacement of JIL-1 and the corresponding decrease in H3S10 phosphorylation which might cause perturbation of nucleosomal organization. It is important to note that nucleosome stability around the mutated BEAF-32 binding site did not decrease. Apparently, at least in this case BEAF-32 does not set a potential barrier that delimits the phasing of surrounding nucleosomes.

ADF1 and BEAF-32 binding sites in *bantam* regulatory region are important for the viability of flies. The nucleosomal profile of transgenes was studied at $ban^{\Delta 1}$ background. We noticed that adult flies homozygous for the $ban^{\Delta 1}$ deletion were only recovered in the "4.7" stock. "ADF" and "BEAF" flies homozygous for the deletion died at the late pupal stage similarly to the flies containing the deletion without any rescue



Amplicon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
"yw"	0.006	0.204	0.014	0.177	0.491	0.273	0.033	0.405	0.438	0.012	0.001	0.022	0.436	0.457	0.145	0.114	0.110	0.368	0.217	0.229	0.042	0.171	0.402	0.006	0.472
"BEAF"	0.036	0.449	0.030	0.063	0.143	0.390	0.035	0.462	0.222	0.024	0.001	0.036	0.390	0.273	0.092	0.061	0.140	0.040	0.034	0.008	0.040	0.034	0.269	0.007	0.178
"ADF"	0.046	0.078	0.041	0.253	0.369	0.487	0.048	0.097	NA	NA	NA	0.037	0.035	0.055	0.078	0.012	0.010	0.013	0.038	0.027	0.014	0.158	0.105	0.021	0.116

Fig. 3. Comparison of the nucleosome profile of 61C7/C8 interband in the control "yw" and transgenic flies.

Bars indicate the representation of the corresponding amplicon in the nucleosomal DNA. Yellow bars show the nucleosomal profile of the native 61C7/C8 interband of "yw" flies; gray bars – in the "4.7" transgene, blue bars – in the "BEAF" transgene, and green bars – in "ADF" transgene. The table below the diagram contains the probability of a random difference in the representation of amplicons in the indicated genotypes compared to the "4.7" stock. Red color highlights the probabilities < 0.05; NA – data not available.

transgenes. It was shown that despite the large size of $ban^{\Delta 1}$ deletion (20 kb), the only essential gene within this deletion is *bantam*. Expression of *bantam* hairpin under the ubiquitous driver control was enough to rescue the lethality (Brennecke et al., 2003). The rest of the sequence deleted by $ban^{\Delta 1}$ seems to contain regulatory elements important for spatial and tissue-specific control of bantam expression (Brennecke et al., 2003; Martin et al., 2004; Peng et al., 2009; Oh, Irvine, 2011). The late pupal lethality of "ADF" and "BEAF" flies suggested that mutations of ADF1 and BEAF-2 binding sites led to *bantam* inactivation in the tissues important for adult fly development.

We would like to dwell in more detail on the mutation of the ADF1-binding site. The thing is that the mutation is not a point replacement but quite a big deletion which could remove more than only ADF-binding site. So we can not exclude the possibility that disruption of some other regulatory elements could cause *bantam* inactivation. However it is important to mention that our mutation does not destroy the core elements of the neighboring DPE-containing promotor (Berkaeva et al., 2009). Further experiments with point replacement in consensus sequences are needed to assert with the confidence that the ADF1-binding disruption is responsible for the effects we observe in "ADF" transgenic flies.

Interestingly, *bantam* was shown to be transcriptionally inactive in salivary glands (Kwon et al., 2011) – the tissue where the disappearance of the interband was observed. Previous studies (Zhimulev et al., 2014) suggest that polytene chromosome organization reflects the organization of interphase chromosomes in various cell types. So, we hypothesize that *bantam* inactivation caused by mutations in the ADF1 and BEAF-32 binding sites is accompanied by altered nucleosome positioning and chromatin structure not only in salivary glands but also in diploid tissues that are essential for the fly viability.

Conclusions

In this study, we investigated the role of non-histone regulatory proteins ADF1 and BEAF-32 in the positioning of nucleosomes and formation of 61C7/C8 interband – one of the decompacted regions of *Drosophila* polytene chromosomes. ADF1 is a specific transcription factor and BEAF-32 is an insulator protein associated with interbands. Using transgenic flies, we showed that ADF1 and BEAF-32 binding sites were indispensable for DNA decompaction in 61C7/C8 interband region. In addition, mutations of ADF1 and BEAF-32 binding sites disrupted nucleosome positioning in the transgenes, characterized by an increase of nucleosome stability. We also found that ADF1 and BEAF-32 binding sites were required for the rescue of null-allele *bantam* – an essential miRNA gene located in 61C7/C8 region. We hope that our findings will be useful for further understanding the mechanisms of interconnection between the degree of DNA packing, nucleosome organization and genomic functions of the particular regions of interphase chromosome.

References

Andreenkov O.V., Andreenkova N.G., Volkova E.I., Georgiev P.G., Goncharova A.A., Pokholkova G.V., Demakov S.A. Ectopic tethering of the chromator protein in UASDBD(GAL4) system as approach for studying of the insulator proteins in *Drosophila melanogaster* polytene chromosomes. Tsitologiia. 2016;58(6):493-497. (in Russian)

Berkaeva M., Demakov S., Schwartz Y.B., Zhimulev I. Functional analysis of *Drosophila* polytene chromosomes decompacted unit: the interband. Chromosome Res. 2009;17(6):745-754. DOI 10.1007/s10577-009-9065-7.

Brennecke J., Hipfner D.R., Stark A., Russell R.B., Cohen S.M. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. Cell. 2003;113(1):25-36.

Chereji R.V., Clark D.J. Major determinants of nucleosome positioning. Biophys. J. 2018;114(10):2279-2289. DOI 10.1016/j.bpj.2018.03.015.

Chereji R.V., Kan T.-W., Grudniewska M.K., Romashchenko A.V., Berezikov E., Zhimulev I.F., Guryev V., Morozov A.V., Moshkin Y.M. Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in *Drosophila melanogaster*. Nucleic Acids Res. 2016;44(3):1036-1051. DOI 10.1093/nar/gkv978.

England B.P., Admon A., Tjian R. Cloning of *Drosophila* transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc. Natl. Acad. Sci. USA. 1992;89(2):683-687.

Gilbert M.K., Tan Y.Y., Hart C.M. The *Drosophila* boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. Genetics. 2006;173(3):1365-1375. DOI 10.1534/genetics.106.056002.

Jiang N., Emberly E., Cuvier O., Hart C.M. Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in *Drosophila melanogaster* links BEAF to transcription. Mol. Cell. Biol. 2009;29(13):3556-3568. DOI 10.1128/MCB.01748-08.

Kvon E.Z., Demakov S.A., Zhimulev I.F. Chromatin decompaction in the interbands of *Drosophila* polytene chromosomes does not correlate with high transcription level. Russ. J. Genet. 2011;47(6):674-681. DOI 10.1134/S1022795411060135. (in Russian)

Kwon S.Y., Grisan V., Jang B., Herbert J., Badenhorst P. Genome-wide mapping targets of the metazoan chromatin remodeling factor NURF reveals nucleosome remodeling at enhancers, core promoters and gene insulators. PLoS Genet. 2016;12(4):1-26.

Martin F.A., Pérez-Garijo A., Moreno E., Morata G. The brinker gradient controls wing growth in *Drosophila.* Development. 2004;131(20):4921-4930. DOI 10.1242/dev.01385.

Moshkin Y.M., Chalkley G.E., Kan T.W., Reddy B.A., Ozgur Z., van Ijcken W.F.J., Dekkers D.H.W., Demmers J.A., Travers A.A., Verrijzer C.P. Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. Mol. Cell. Biol. 2012;32(3):675-688. DOI 10.1128/MCB.06365-11.

Oh H., Irvine K.D. Cooperative regulation of growth by Yorkie and Mad through bantam. Dev. Cell. 2011;20(1):109-122. DOI 10.1016/j.devcel.2010.12.002.

Orsi G.A., Kasinathan S., Hughes K.T., Saminadin-Peter S., Henikoff S., Ahmad K. High-resolution mapping defines the cooperative architecture of Polycomb response elements. Genome Res. 2014;24(5):809-820. DOI 10.1101/gr.163642.113.

Peng H.W., Slattery M., Mann R.S. Transcription factor choice in the Hippo signaling pathway: homothorax and yorkie regulation of the microRNA bantam in the progenitor domain of the *Drosophila* eye imaginal disc. Genes Dev. 2009;23(19):2307-2319. DOI 10.1101/gad.1820009.

Rath U., Ding Y., Deng H., Qi H., Bao X., Zhang W., Girton J., Johansen J., Johansen K.M. The chromodomain protein, Chromator, interacts with JIL-1 kinase and regulates the structure of *Drosophila* polytene chromosomes. J. Cell Sci. 2006;119(Pt 11):2332-2341.

Semeshin V.F., Demakov S.A., Shloma V.V., Vatolina T.Y., Gorchakov A.A., Zhimulev I.F. Interbands behave as decompacted autonomous units in *Drosophila melanogaster* polytene chromosomes. Genetica. 2008;132(3):267-279.

Talamillo A., Fernández-Moreno M.A., Martínez-Azorín F., Bornstein B., Ochoa P., Garesse R. Expression of the *Drosophila melanogaster* ATP synthase α subunit gene is regulated by a transcriptional element containing GAF and Adf-1 binding sites. Eur. J. Biochem. 2004;271(20):4003-4013. DOI 10.1111/j.1432-1033.2004.04336.x.

Thomas G.H., Elgin S.C. Protein/DNA architecture of the DNase I hypersensitive region of the *Drosophila* hsp26 promoter. EMBO J. 1988;7(7):2191-2201.

Tsukiyama T., Becker P.B., Wu C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature. 1994;367(6463):525-532. DOI 10.1038/367525a0.

Vogelmann J., Le Gall A., Dejardin S., Allemand F., Gamot A., Labesse G., Cuvier O., Nègre N., Cohen-Gonsaud M., Margeat E., Nöllmann M. Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the *Drosophila* genome. PLoS Genet. 2014;10(8):e1004544. DOI 10.1371/journal.pgen.1004544.

Zhimulev I.F., Semeshin V.F., Kulichkov V.A., Belyaeva E.S. Intercalary heterochromatin in *Drosophila*. I. Localization and general characteristics. Chromosoma. 1982;87(2):197-228. DOI 10.1007/BF00338489.

Zhimulev I.F., Zykova T.Y., Goncharov F.P., Khoroshko V.A., Demakova O.V., Semeshin V.F., Pokholkova G.V., Boldyreva L.V., Demidova D.S., Babenko V.N., Demakov S.A., Belyaeva E.S. Genetic organization of interphase chromosome bands and interbands in *Drosophila melanogaster*. PLoS One. 2014;9(7):e101631. DOI 10.1371/journal.pone.0101631.

Acknowledgements. The work was supported by IMCB SB RAS budget project (No. 0310-2018-0010).

Conflict of interest. The authors declare no conflict of interest.

Received December 10, 2018. Revised December 18, 2018. Accepted January 5, 2019.

Вавиловский журнал генетики и селекции. 2019;23(2):160-167
DOI 10.18699/VJ19.476

The role of ATP-dependent chromatin remodeling factors in chromatin assembly *in vivo*

Iu.A. Il'ina, A.Y. Konev ✉

Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina, Russia
✉ e-mail: konev_ay@pnpi.nrcki.ru

Chromatin assembly is a fundamental process essential for chromosome duplication subsequent to DNA replication. In addition, histone removal and incorporation take place constantly throughout the cell cycle in the course of DNA-utilizing processes, such as transcription, damage repair or recombination. *In vitro* studies have revealed that nucleosome assembly relies on the combined action of core histone chaperones and ATP-utilizing molecular motor proteins such as ACF or CHD1. Despite extensive biochemical characterization of ATP-dependent chromatin assembly and remodeling factors, it has remained unclear to what extent nucleosome assembly is an ATP-dependent process *in vivo*. Our original and published data about the functions of ATP-dependent chromatin assembly and remodeling factors clearly demonstrated that these proteins are important for nucleosome assembly and histone exchange *in vivo*. During male pronucleus reorganization after fertilization CHD1 has a critical role in the genome-scale, replication-independent nucleosome assembly involving the histone variant H3.3. Thus, the molecular motor proteins, such as CHD1, function not only in the remodeling of existing nucleosomes but also in *de novo* nucleosome assembly from DNA and histones *in vivo*. ATP-dependent chromatin assembly and remodeling factors have been implicated in the process of histone exchange during transcription and DNA repair, in the maintenance of centromeric chromatin and in the loading and remodeling of nucleosomes behind a replication fork. Thus, chromatin remodeling factors are involved in the processes of both replication-dependent and replication-independent chromatin assembly. The role of these proteins is especially prominent in the processes of large-scale chromatin reorganization; for example, during male pronucleus formation or in DNA repair. Together, ATP-dependent chromatin assembly factors, histone chaperones and chromatin modifying enzymes form a "chromatin integrity network" to ensure proper maintenance and propagation of chromatin landscape.
Key words: chromatin; chromatin assembly; ATP-dependent remodeling factors.

For citation: Il'ina Iu.A., Konev A.Y. The role of ATP-dependent chromatin remodeling factors in chromatin assembly *in vivo*. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):160-167. DOI 10.18699/VJ19.476

Роль АТФ-зависимых хроматин-ремоделирующих факторов в процессе сборки хроматина *in vivo*

Ю.А. Ильина, А.Ю. Конев ✉

Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт», Гатчина, Россия
✉ e-mail: konev_ay@pnpi.nrcki.ru

Сборка хроматина – фундаментальный процесс, необходимый для дупликации хромосом в процессе репликации ДНК. Кроме того, удаление гистонов и их инкорпорирование постоянно происходят в течение клеточного цикла в ходе процессов метаболизма ДНК, таких как транскрипция, восстановление повреждений или рекомбинация. Исследования *in vitro* показали, что сборка хроматина требует совместного действия гистоновых шаперонов и использующих энергию АТФ хроматин-ремоделирующих факторов – ACF или CHD1. Несмотря то, что АТФ-зависимые факторы сборки и ремоделирования хроматина хорошо охарактеризованы биохимически, оставалось неясным, до какой степени сборка нуклеосом является АТФ-зависимым процессом *in vivo*. Наши собственные и опубликованные в литературе данные о функциях АТФ-зависимых хроматин-ремоделирующих факторов показывают, что эти белки существенны для сборки нуклеосом и обмена гистонов и *in vivo*. CHD1 – критически важный фактор при реорганизации мужского пронуклеуса после оплодотворения, в процессе которой происходит независимая от репликации сборка хроматина, содержащего вариантный гистон H3.3. Следовательно, молекулярные моторные белки, такие как CHD1, функционируют *in vivo* не только в ремоделировании существующих нуклеосом, но также и в сборке нуклеосом *de novo* из ДНК и гистонов. АТФ-зависимые факторы сборки и ремоделирования хроматина участвуют в процессе обмена гистонов во время транскрипции и репарации ДНК, в поддержании центромерного хроматина и образовании и ремоделировании нуклеосом позади прохождения репликационной вилки. Таким образом, хроматин-ремоделирующие факторы участвуют в процессах как зависимой, так и не зависимой от репликации сборки хроматина. Их роль особенно заметна в процессах крупномасштабной реорганизации хроматина,

например при реорганизации хроматина мужского пронуклеуса или при восстановлении повреждений ДНК. Гистоновые шапероны, модифицирующие хроматин ферменты и АТФ-зависимые факторы сборки хроматина совместно образуют сеть факторов, обеспечивающих поддержание целостности хроматина.
Ключевые слова: хроматин; сборка хроматина; АТФ-зависимые хроматин-ремоделирующие факторы.

Itroduction

Chromatin assembly is a fundamentally important process, which is essential for chromosome duplication subsequent to DNA replication (Krude, Keller, 2001; Vincent et al., 2008), but also take place constantly throughout the cell cycle in the course of DNA-utilizing processes, such as transcription, damage repair or recombination (Venkatesh, Workman, 2015; Hauer, Gasser, 2017; Serra-Cardona, Zhang, 2018). Replication-independent chromatin assembly generally involves histone variants (e.g. H3.3, H2A.Z), while the so-called canonical histones (H2A, H2B, H3.1, H3.2, H4) are incorporated in the course of replication (Henikoff, Ahmad, 2005).

In vitro studies suggest that nucleosome assembly relies on the combined action of core histone chaperones and ATP-utilizing molecular motor proteins (Tyler, 2002; Haushalter, Kadonaga, 2003). The histone chaperones mediate the stepwise processes of disassembly and reassembly of nucleosomes and shield histones from promiscuous interactions and aggregation to ensure their proper storage, transport, post-translational modification, nucleosome assembly and turnover (Hammond et al., 2017). Various chaperones have been identified that specifically interact with certain histone variants driving nucleosome assembly and disassembly during replication-independent and replication-dependent processes. For example, H3.1 is found in a complex containing histone chaperones Asf1 and CAF-1, whereas H3.3 forms a complex with HIRA and Asf1 or with the DAXX chaperone (Tagami et al., 2004; Elsasser et al., 2012).

Biochemical evidence, however, has revealed that nucleosome assembly is an ATP-dependent process and requires the action of ATP-hydrolyzing molecular motor proteins such as ACF or CHD1 (Ito et al., 1997; Haushalter, Kadonaga, 2003; Lusser et al., 2005). These evolutionarily conserved factors promote the efficiency of core histone deposition as well as mediate the spacing of nucleosome arrays. ATP-utilizing molecular machines modify chromatin structure by modulating histone-DNA contacts, resulting in the repositioning or disruption of the nucleosome, changing the conformation of nucleosomal DNA or changing the composition of the histone octamer (Morettini et al., 2008; Narlikar et al., 2013; Clapier et al., 2017). Chromatin remodeling factors are compositionally and functionally diverse multi-protein complexes, yet they share the presence of a motor subunit that belongs to the SNF2-family of ATPases. The DNA-translocating core of SNF2-type proteins consists of two paired RecA-like domains or lobes, capable of rearranging during the ATP hydrolysis cycle to create a ratchet-like motion along DNA in single base increments (Clapier et al., 2017). Recent studies provided mechanistic insights into nucleosome assembly by the combined action of histone chaperones and ATP-dependent molecular motors.

These studies revealed the rapid formation of a stable non-nucleosomal histone-DNA intermediate, a prenucleosome, due to histone deposition by chaperones; this intermediate is then converted into nucleosomes by a motor protein such as ACF (Torigoe et al., 2011). The prenucleosome is a stable conformational isomer of the nucleosome, which consists of a histone octamer associated with 80 base pairs of DNA located at a position that corresponds to the central 80 bp of a nucleosome core particle (Fei et al., 2015). Monomeric prenucleosomes with free flanking DNA do not spontaneously fold into nucleosomes, but can be converted into canonical nucleosomes by ATP-dependent chromatin remodelers such as ACF or Chd1 (Torigoe et al., 2013). Analyses of the remodeling-defective Chd1 motor protein revealed a functional distinction between ATP-dependent nucleosome assembly and chromatin remodeling, and proposed a model for chromatin assembly in which randomly distributed nucleosomes are formed by the nucleosome assembly function of Chd1, and then regularly spaced nucleosome arrays are generated by the chromatin remodeling activity (Torigoe et al., 2013).

To date, the ATP-dependent assembly of extended, periodic arrays of nucleosomes has been observed *in vitro* with the *Drosophila* ACF, CHRAC (Ito et al., 1999; Fyodorov, Kadonaga, 2003; Kukimoto et al., 2004) and ToRC (Emelyanov et al., 2012) complexes as well as with the human RSF complex (Loyola et al., 2003). The catalytic subunit of each of these complexes belongs to the ISWI subfamily of the SNF2-like family of ATPases. ISWI by itself in conjunction with histone chaperone dNLP has a chromatin assembly activity (Khuong et al., 2017). Activity in an ATP-dependent chromatin assembly reaction was demonstrated for chromodomain-containing proteins of the CHD family: the *Drosophila*, yeast and mouse CHD1 protein (Robinson, Schultz, 2003; Lusser et al., 2005; Piatti et al., 2015) and human CHD2 (Liu et al., 2015). Unlike ACF, CHD1 cannot assemble chromatin containing the linker histone H1 (Lusser et al., 2005). Finally, the mammalian ATRX-DAXX complex catalyzes the deposition and remodeling of H3.3-containing nucleosomes (Drane et al., 2010; Lewis et al., 2010). ISWI and CHD1 remodelers share some structural and functional similarities (Narlikar et al., 2013; Clapier et al., 2017). Both ISWI and CHD1 proteins reposition nucleosomes to locations close to the center of short DNA fragments. The directionality and rate of nucleosome mobilization are determined in ISWI and CHD1 proteins by the SANT-SLIDE DNA binding domain (DBD), structurally similar between two proteins. DBD binds DNA at the nucleosomal entry site and in the adjacent linker anchoring the remodeler to the nucleosome. Although the N-terminal regions of ISWI and CHD1 are not related at the sequence level, they carry out similar functions. *Drosophila* ISWI protein contains two conserved regulatory regions, termed AutoN and NegC, which negatively regulate ATP hydrolysis (AutoN) or the coupling of ATP hydrolysis to DNA translocation (NegC). In CHD1 protein, a similar negative regulatory function is performed by the double chromodomain unit of the CHD1 remodeler, which can pack against the DNA-binding surface of the ATPase motor. Double chromodomains of mammalian CHD1 protein recognize the trimethylated histone H3

lysine 4, but chromodomains of yeast and *Drosophila* CHD1 and of mammalian CHD2 do not have such specificity. In yeast chromodomain mutations result in dissociation of Chd1 from chromatin. By contrast, in *Drosophila melanogaster* the chromodomains do not play an important role in its localization to chromatin; however, mutations in *Drosophila* CHD1 chromodomains negatively affect the chromatin assembly activities of CHD1 *in vitro* (Morettini et al., 2011). Similar biochemical characteristics probably determine the role of these two subfamilies of SNF2-type remodelers in chromatin assembly reaction as well as partial redundancy of these factors *in vivo*. The translocase domains of ATRX proteins have diverged from those of ISWI, and CHD1, suggesting another mechanism for histone exchange (Narlikar et al., 2013). The ATRX protein associates with the DAXX chaperone to couple chromatin dissociation with the reassembly of nucleosomes enriched for the histone variant H3.3 (Lewis et al., 2010).

Despite extensive biochemical characterization of ATP-dependent nucleosome loading and remodeling, it has remained unclear to what extent nucleosome assembly is an ATP-dependent process *in vivo*. We provided the first clear confirmation of the requirement of an ATP-utilizing motor protein (CHD1) in chromatin assembly *in vivo* (Konev et al., 2007). We generated *Chd1* null alleles and found that *Chd1* mutants are subviable but sterile. When maternal CHD1 is eliminated, the paternal genome does not undergo the first zygotic mitosis, leading to the development of haploid embryos. In contrast, CHD1 does not appear to affect the organization of maternal chromatin. It has been shown in *Drosophila* and mice that H3.3 is specifically present in the male pronucleus (Loppin et al., 2005; Torres-Padilla et al., 2006). Histones are deposited in decondensing sperm DNA by replication-independent mechanisms involving the HIRA chaperone complex, which use the histone variant H3.3, but not canonical H3 (Loppin et al., 2005; Orsi et al., 2013). Using the FLAG-tagged H3.3 transgene we demonstrated that, in contrast to wild-type animals, in Chd1-null embryos the H3.3-FLAG signal in the male pronucleus did not co-localize with the DNA and remained constrained to the nuclear periphery (Konev et al., 2007). We have found that CHD1 physically interacts with the H3.3-specific histone chaperone HIRA, which delivers histones H3.3-H4 to the male pronucleus. CHD1 is dispensable for protamine removal. This work established that CHD1 is a major factor in replacement histone metabolism and revealed a critical role for CHD1 in the genome-scale, replication-independent nucleosome assembly involving the histone variant H3.3. Thus, the molecular motor proteins, such as CHD1, function not only in remodeling of existing nucleosomes but also in *de novo* nucleosome assembly from DNA and histones *in vivo*.

Nevertheless, the results of the further studies of the ISWI- and CHD1-type remodelers' role in chromatin assembly *in vivo* are somewhat controversial. In yeasts *Saccharomyces cerevisiae* Iswi complexes (yeasts have two ISWI proteins: Isw1p and Isw2p) and Chd1 have partially redundant roles in nucleosome assembly and positioning (Xella et al., 2006; Gkikopoulos et al., 2011; Smolle et al., 2012; Yadav, Whitehouse, 2016). While individual *isw1*, *isw2* and *chd1* mutants are viable, only the *isw1 isw2 chd1* triple mutant combination results in severe growth defects as well as in wide-spread disruption of nucleosome positioning throughout the yeast genome (Vary et al., 2003; Gkikopoulos et al., 2011). Very similar results were obtained for the fission yeast *Schizosaccharomyces pombe*, which lacks ISWI-type remodelers, but has two CHD1-type ATPases, Hrp1 and Hrp3 (Hennig et al., 2012; Pointner et al., 2012). It was shown that *de novo* nucleosome assembly after DNA replication is altered in yeast *S. cerevisiae* mutants lacking histone chaperones (CAF-1) or chromatin remodelers Isw1 and Chd1 (Yadav, Whitehouse, 2016). ATP-dependent chromatin-remodeling enzymes Isw1 and Chd1 collaborate with histone chaperones to assemble and remodel nucleosomes as they are loaded behind a replication fork, allowing for rapid organization of chromatin during S phase. Chd1 is known to associate with elongating RNAPII through interaction with the PAF complex and other RNAPII-associated proteins (Simic et al., 2003). Some evidence has suggested a role for Chd1 in replication-independent histone exchange or assembly in the process of active transcription (Radman-Livaja et al., 2012; Smolle et al., 2012). Genome-wide analysis of H3 histone exchange in *S. cerevisiae* Chd1-deficient strains has shown that promoter and 5′ histone turnover are slower in *chd1Δ* cells, while maximal H3 replacement is instead observed at the 3′ ends of genes (Radman-Livaja et al., 2012). Thus, Chd1 accelerates H3 replacement at the 5′ ends of genes, while protecting the 3′ ends of genes from excessive H3 turnover, and specifically stabilizes nucleosomes over the 3′ ends of longer genes. Studies of H3 histone turnover in an Isw1b and Chd1 double mutant revealed that these chromatin remodeling factors both prevent co-transcriptional trans-histone exchange at largely overlapping, yet somewhat distinct groups of genes (Smolle et al., 2012). It is proposed, that yeast Isw1 and Chd1 may prevent histone exchange and promote chromatin integrity at lowly and highly transcribed genes respectively: Isw1 in the form of the Isw1b complex by remodeling nucleosomes that were retained on the DNA in spite of transcription, and Chd1 by reassembling nucleosomes in cis in the wake of RNAPII (Smolle et al., 2012). Still, it is clear that nucleosomes can be assembled without both Isw1b and Chd1. Analyses of nucleosome organization in the wild-type H2BK123A strain, incapable to monoubiquitinate histone H2B on Lys 123, and the *chd1Δ* strain indicate that in the absence of H2B monoubiquitination or Chd1, genes that are positively regulated by histone H2B monoubiquitination or Chd1 fail to adequately assemble nucleosomes +2, +3, and +4 in coding regions (Lee et al., 2012).

We have shown that in *Drosophila* CHD1 functions in H3.3 deposition during later stages of embryonic development, possibly in a processes associated with transcription (Konev et al., 2007). Analyses of polytene chromosomes revealed that the null mutations in *Drosophila Chd1* cause a decondensation of the male X chromosome, similar to that observed for the *ISWI* mutations (Konev et al., 2016). An effect of *Chd1* null mutations is increased by deficiency of one of two genes encoding the histone variant H3.3, *His3.3B*, suggesting that the role of CHD1 in the male X chromosome organization can be mediated by CHD1 activity in H3.3 histone deposition and exchange.

Mammals have two orthologs of yeast and *Drosophila* CHD1, Chd1 and Chd2. It was shown that in muscle development, transcriptional activation of the myogenic transcription factor MyoD is mediated by replication-independent histone

deposition (Yang et al., 2011). In this process, HIRA and Asf1a, but not CHD1 or Asf1b, mediate H3.3 incorporation in the promoter and the critical upstream regulatory regions of the *MyoD* gene. However, MyoD interacts with Chd2 and recruits it to downstream myogenic gene promoters, but not to housekeeping or silent gene promoters both in myoblasts and in differentiated cells (Harada et al., 2012). Genome-wide analysis of endogenous H3.3 incorporation demonstrates that knockdown of Chd2 prevents H3.3 deposition at differentiation-dependent, but not housekeeping, genes and inhibits myogenic gene activation. Thus, MyoD determines cell fate and facilitates differentiation-dependent gene expression through Chd2-dependent deposition of H3.3 at myogenic loci prior to differentiation (Harada et al., 2012). Chd2 is also required to maintain the differentiation potential of mouse embryonic stem cells (ESC). Chd2-depleted ESCs showed suppressed expression of developmentally regulated genes upon differentiation and alterations in the nucleosome occupancy of the histone variant H3.3 for developmentally regulated genes (Semba et al., 2017). Chd1 is also essential for open chromatin and pluripotency of embryonic stem cells, and for somatic cell reprogramming; downregulation of Chd1 leads to accumulation of heterochromatin (Gaspar-Maia et al., 2009). It is not known yet, whether or not effects of Chd1 in ESC are associated with changes in histone deposition. Analyses of genome-wide nucleosome specificity and function of chromatin remodelers in ES cells revealed that, while Chd1 is present near the 5′ ends of genes, the Chd2-nucleosome enrichment pattern encompassed the entire transcription unit and shared a high correlation with H3K36me3 (de Dieuleveult et al., 2016). This is consistent with how yeast Chd1 works, and thus yeast Chd1 may be functionally equivalent rather to mammalian Chd2.

ISWI protein is a catalytic subunit of four complexes, capable to assemble extended, periodic arrays of nucleosomes *in vitro*: ACF, CHRAC, Torch and RSF. In *Drosophila*, *ISWI* is an essential gene, and loss of ISWI function results in global structural alteration of the chromosomes, specifically the male X-chromosome (Deuring et al., 2000). It was found that ISWI is involved in the incorporation of the linker histone H1 into chromatin *in vivo* (Corona et al., 2007; Siriaco et al., 2009). Our work had shown that ISWI is unable to substitute for CHD1 in the deposition of H3.3 (Konev et al., 2007). However, at fertilization, maternal ISWI targets the paternal genome and drives its repackaging into de-condensed nucleosomal chromatin (Doyen et al., 2015). In embryos from *ISWI*^{1} mothers, the male pronucleus remained abnormally condensed, but core histones were deposited efficiently onto the condensed paternal genome in the absence of ISWI. These observations suggest that ISWI is not required for the incorporation of histones into the male pronucleus but is crucial for the formation of an open chromatin structure. As in the case of *Chd1* mutants (Konev et al., 2007; Podhraski et al., 2010), failure of proper transition from sperm chromatin to nucleosomal chromatin in *iswi* mutant embryos is followed by mitotic defects, aneuploidy, and haploid embryonic divisions (Doyen et al., 2015). ACF (ATP-utilizing chromatin assembly and remodeling factor) (Ito et al., 1999; Fyodorov, Kadonaga, 2003) and CHRAC (chromatin-accessibility-complex) (Kukimoto et al., 2004) contain one common subunit, termed Acf1, which stimulates and modulates the intrinsic chromatin assembly and remodeling activities of ISWI (Ito et al., 1999; Eberharter et al., 2001; Fyodorov, Kadonaga, 2002). In addition, CHRAC comprises two small, developmentally regulated subunits (CHRAC14, CHRAC16) which facilitate nucleosome sliding (Corona et al., 2000; Hartlepp et al., 2005). Expression of the signature subunit ACF1 is restricted during embryonic development, but remains high in primordial germ cells. Therefore, we explored roles for ACF1 during *Drosophila* oogenesis (Borner et al., 2016). We have isolated the novel loss – of function allele *Acf1(7)* which leads to defective egg chambers and their elimination through apoptosis. Two-fold overexpression of CHRAC (ACF1 and CHRAC-16) leads to increased apoptosis and packaging defects in egg chambers. Thus, finely tuned CHRAC levels are required for proper oogenesis. The mutant *acf1* fly line displays defects in heterochromatic silencing and exhibits reduced ATP-dependent chromatin assembly activity in embryonic cell extracts (Fyodorov et al., 2004). ACF/CHRAC contribute to establishing the repressive ground state of chromatin through organization of regular nucleosome arrays (Scacchetti et al., 2018). Surprisingly, despite the fact that ACF is the first discovered ATP-dependent chromatin assembly complex, direct evidence of its role in chromatin assembly *in vivo* is lacking. The human two-subunit complex RSF (Remodeling and Spacing Factor) contains, in addition to the ISWI homolog hSNFH, the p325 (Rsf-1) polypeptide, which is similar but not homologous to *Drosophila* Acf1 (Loyola et al., 2003). In *Drosophila* RSF contributes to histone H2Av replacement in the pathway of silent chromatin formation (Hanai et al., 2008).

We described a novel, evolutionarily conserved, ISWI-containing complex, ToRC (Toutatis-containing chromatin remodeling complex), which consist of ISWI, Toutatis/TIP5 and the transcriptional corepressor CtBP (Emelyanov et al., 2012). Toutatis (Tou) is the *Drosophila* ortholog of TIP5, a large subunit of a nucleolar-specific, SNF2H-containing chromatin remodeling factor, NoRC (Strohner et al., 2001). Tou exhibits strong similarity to Acf1 and shares with it a number of domains that are important for the chromatin assembly activity of ACF. ToRC facilitates ATP-dependent nucleosome assembly *in vitro* and all three subunits are required for its maximal biochemical activity. By contrast with NoRC, the ToRC complex is excluded from the nucleolus. ToRC tethering to its nuclear targets is dependent on the transcriptional corepressor CtBP. ToRC is involved in CtBP-mediated regulation of transcription by RNA polymerase II *in vivo*. However, native C-terminally truncated Toutatis isoforms do not associate with CtBP and localize predominantly to the nucleolus. Thus, Toutatis forms two alternative complexes, ToRC and dNoRC, comprising of Tou and ISWI, that have differential distribution and can participate in distinct aspects of nuclear DNA metabolism. It is remained to be determined whether or not ToRC or NoRC is involved in histone incorporation *in vivo*.

ATRX is a chromatin remodeler that together with its chaperone DAXX in mammalian cells deposits the histone variant H3.3 in pericentric and telomeric regions (Drane et al., 2010; Goldberg et al., 2010; Lewis et al., 2010; Wong et al., 2010). We performed biochemical and genetic analyses of the *Drosophila melanogaster* ortholog of ATRX (Xnp) (Emelyanov et al., 2010). *Drosophila* ATRX is expressed

throughout development in two isoforms, p185 and p125, which form distinct multisubunit complexes in fly embryos. The ATRX185 complex comprises p185 and heterochromatin protein HP1a. Consistently, ATRX185 but not ATRX125 in polytene chromosomes is highly concentrated in the pericentric beta-heterochromatin of the X chromosome, where it is required for HP1a deposition. The loss-of-function allele of the *Drosophila* ATRX/Xnp gene is semilethal and alleles of the ATRX/XNP gene that do not express p185 are strong suppressors of position effect variegation. In another study it was shown that the ATRX/XNP focus near the heterochromatin of the X chromosome corresponds to an unusual decondensed satellite DNA block, and both active genes and the XNP focus are sites of ongoing H3 to H3.3 nucleosome replacement (Schneiderman et al., 2009). Using an inducible system that produces nucleosome-depleted chromatin at the Hsp70 genes in *Drosophila* it was found that the ATRX/XNP chromatin remodeler and the Hira histone chaperone independently bind nucleosome-depleted chromatin (Schneiderman et al., 2012). H3.3 deposition assays reveal that Xnp and Hira are required for efficient nucleosome replacement, and double-mutants are lethal. Thus, the function of ATRX in H3.3 deposition is conserved in evolution.

ATP-dependent chromatin assembly and remodeling factors have also been implicated in the maintenance of centromeric chromatin, which is epigenetically specified in most eukaryotes by the centromere-specific histone H3 variant, CENP-A. Human RSF1 localizes to centromeres and promotes the stable association of CENP-A with centromeres (Perpelescu et al., 2009). Tethering RSF1 to an ectopic non-centromeric alphoid DNA array is sufficient to stimulate histone exchange, and this reaction was coupled with new H3.3 or CENP-A assembly (Ohzeki et al., 2016). CHD1 was found to stimulate the deposition of CENP-A at centromeres in fission yeast and humans (Walfridsson et al., 2005; Okada et al., 2009), however, it is dispensable for CID (*Drosophila* CENP-A) deposition in flies (Podhraski et al., 2010).

Chromatin dynamics plays a very important role during DNA repair (Hauer, Gasser, 2017). ATRX-deficient cells exhibit a defect in repairing exogenously induced DNA double-strand breaks (DSBs) by homologous recombination (HR) (Juhasz et al., 2018). DAXX- and H3.3-depleted cells exhibit identical HR defects as ATRX-depleted cells, and both ATRX and DAXX function to deposit H3.3 during DNA repair synthesis. This suggests that ATRX facilitates the chromatin reconstitution required for extended DNA repair synthesis and sister chromatid exchange during HR (Elbakry et al., 2018; Juhasz et al., 2018). Human CHD2 plays a crucial role in regulating non-homologous end-joining (NHEJ), a dominant DSB repair pathway in human cells (Luijsterburg et al., 2016). This mechanism is initiated by DNA damage-associated poly(ADP-ribose) polymerase 1 (PARP1), which recruits CHD2 through a poly(ADP-ribose)-binding domain; CHD2 in turn triggers rapid deposition of histone variant H3.3 at sites of DNA damage. PARP1, CHD2, and H3.3 regulate the assembly of NHEJ complexes at broken chromosomes to promote efficient DNA repair (Luijsterburg et al., 2016). Targeted disruption of the CHD1 gene in human cells leads to a defect in DSB repair via homologous recombination, resulting in hypersensitivity to ionizing radiation and in increased error-prone NHEJ repair for DSB repair (Kari et al., 2016; Shenoy et al., 2017; Zhou et al., 2018). We have shown that null mutations of *Drosophila Chd1* cause hypersensitivity to ionizing radiation at the third instar larval stage, suggesting that *Drosophila* CHD1 is also involved in DSB repair. Considering that *Drosophila* has only a single ortholog of human CHD1 and CHD2, it will be interesting to determine which of the two major pathways of DSB repair is affected and whether or not this effect is associated with the role of fly CHD1 in H3.3 deposition.

Conclusion

The studies of *in vivo* functions of ATP-dependent chromatin assembly and remodeling factors clearly demonstrated that these factors do participate in nucleosome assembly and histone exchange *in vivo*. Most of the evidence concerning the role of ATP-dependent chromatin assembly and remodeling factors in the process of chromatin assembly *in vivo* was obtained for recombination independent assembly processes, involving histone variants such as H3.3. Yet it was shown that ATP-dependent chromatin-remodeling enzymes Isw1 and Chd1 in collaboration with histone chaperones assemble and remodel nucleosomes as they are loaded behind a replication fork (Yadav, Whitehouse, 2016), suggesting a common mechanism for both types for chromatin assembly. Their role is especially prominent in the processes of large-scale chromatin reorganization, for example during male pronucleus formation or DNA repair. Analysis of effects of individual factors is hampered by a high level of redundancy in the function of ATP-dependent chromatin assembly and remodeling factors. Interpretations of the role of chromatin remodelers in chromatin assembly *in vivo* is further complicated by the fact that mutations in the ATP-dependent chromatin assembly and remodeling factors, such as Chd1, may affect the distribution of histone modifications known to affect nucleosome stability (Lee J.S. et al., 2012; Radman-Livaja et al., 2012; Lee Y. et al., 2017). Together, ATP-dependent chromatin assembly and remodeling factors, histone chaperones and chromatin modifying enzymes form a "chromatin integrity network" to ensure proper maintenance and propagation of the chromatin landscape as well as regulation of genetic processes. All identified ATP-dependent chromatin assembly factors are linked to cancer. Thus, dissecting a chromatin integrity network should not only provide further insight into the fundamental organization of the eukaryotic genome but also contribute to human health.

References

Borner K., Jain D., Vazquez-Pianzola P., Vengadasalam S., Steffen N., Fyodorov D.V., Tomancak P., Konev A., Suter B., Becker P.B. A role for tuned levels of nucleosome remodeler subunit ACF1 during *Drosophila* oogenesis. Dev. Biol. 2016;411(2):217-230. DOI 10.1016/j.ydbio.2016.01.039.

Clapier C.R., Iwasa J., Cairns B.R., Peterson C.L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell. Biol. 2017;18(7):407-422. DOI 10.1038/nrm.2017.26.

Corona D.F., Eberharter A., Budde A., Deuring R., Ferrari S., Varga-Weisz P., Wilm M., Tamkun J., Becker P.B. Two histone fold proteins, CHRAC-14 and CHRAC-16, are developmentally regulated subunits of chromatin accessibility complex (CHRAC). EMBO J. 2000;19(12):3049-3059. DOI 10.1093/emboj/19.12.3049.

Corona D.F., Siriaco G., Armstrong J.A., Snarskaya N., McClymont S.A., Scott M.P., Tamkun J.W. ISWI regulates higher-order chromatin structure and histone H1 assembly *in vivo*. PLoS Biol. 2007;5(9):e232. DOI 10.1371/journal.pbio.0050232.

de Dieuleveult M., Yen K., Hmitou I., Depaux A., Boussouar F., Bou Dargham D., Jounier S., Humbertclaude H., Ribierre F., Baulard C., Farrell N.P., Park B., Keime C., Carriere L., Berlivet S., Gut M., Gut I., Werner M., Deleuze J.F., Olaso R., Aude J.C., Chantalat S., Pugh B.F., Gerard M. Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. Nature. 2016; 530(7588):113-116. DOI 10.1038/nature16505.

Deuring R., Fanti L., Armstrong J.A., Sarte M., Papoulas O., Prestel M., Daubresse G., Verardo M., Moseley S.L., Berloco M., Tsukiyama T., Wu C., Pimpinelli S., Tamkun J.W. The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure *in vivo*. Mol. Cell. 2000;5(2):355-365.

Doyen C.M., Chalkley G.E., Voets O., Bezstarosti K., Demmers J.A., Moshkin Y.M., Verrijzer C.P. A Testis-specific chaperone and the chromatin remodeler ISWI mediate repackaging of the paternal genome. Cell Rep. 2015;13(7):1310-1318. DOI 10.1016/j.celrep.2015.10.010.

Drane P., Ouararhni K., Depaux A., Shuaib M., Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev. 2010; 13(7):1310-1318. DOI 10.1101/gad.566910.

Eberharter A., Ferrari S., Langst G., Straub T., Imhof A., Varga-Weisz P., Wilm M., Becker P.B. Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. EMBO J. 2001; 20(14):3781-3788.

Elbakry A., Juhasz S., Mathes A., Lobrich M. DNA repair synthesis and histone deposition partner during homologous recombination. Mol. Cell. Oncol. 2018;5(5):e1511210. DOI 10.1080/23723556.2018.1511210.

Elsasser S.J., Huang H., Lewis P.W., Chin J.W., Allis C.D., Patel D.J. DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. Nature. 2012;491(7425):560-565. DOI 10.1038/nature11608.

Emelyanov A.V., Konev A.Y., Vershilova E., Fyodorov D.V. Protein complex of *Drosophila* ATRX/XNP and HP1a is required for the formation of pericentric beta-heterochromatin *in vivo*. J. Biol. Chem. 2010;285(20):15027-15037. DOI 10.1074/jbc.M109.064790.

Emelyanov A.V., Vershilova E., Ignatyeva M.A., Pokrovsky D.K., Lu X., Konev A.Y., Fyodorov D.V. Identification and characterization of ToRC, a novel ISWI-containing ATP-dependent chromatin assembly complex. Genes Dev. 2012;26(6):603-614. DOI 10.1101/gad.180604.111.

Fei J., Torigoe S.E., Brown C.R., Khuong M.T., Kassavetis G.A., Boeger H., Kadonaga J.T. The prenucleosome, a stable conformational isomer of the nucleosome. Genes Dev. 2015;29(24):2563-2575. DOI 10.1101/gad.272633.115.

Fyodorov D.V., Blower M.D., Karpen G.H., Kadonaga J.T. Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin *in vivo*. Genes Dev. 2004;18(2):170-183.

Fyodorov D.V., Kadonaga J.T. Binding of Acf1 to DNA involves a WAC motif and is important for ACF-mediated chromatin assembly. Mol. Cell. Biol. 2002;22(18):6344-6353.

Fyodorov D.V., Kadonaga J.T. Chromatin assembly *in vitro* with purified recombinant ACF and NAP-1. Methods Enzymol. 2003;371: 499-515.

Gaspar-Maia A., Alajem A., Polesso F., Sridharan R., Mason M.J., Heidersbach A., Ramalho-Santos J., McManus M.T., Plath K., Meshorer E., Ramalho-Santos M. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature. 2009;460(7257):863-868. DOI 10.1038/nature08212.

Gkikopoulos T., Schofield P., Singh V., Pinskaya M., Mellor J., Smolle M., Workman J.L., Barton G.J., Owen-Hughes T. A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science. 2011;333(6050):1758-1760. DOI 10.1126/science.1206097.

Goldberg A.D., Banaszynski L.A., Noh K.M., Lewis P.W., Elsaesser S.J., Stadler S., Dewell S., Law M., Guo X., Li X., Wen D., Chapgier A., DeKelver R.C., Miller J.C., Lee Y.L., Boydston E.A., Holmes M.C., Gregory P.D., Greally J.M., Rafii S., Yang C., Scambler P.J., Garrick D., Gibbons R.J., Higgs D.R., Cristea I.M., Urnov F.D., Zheng D., Allis C.D. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell. 2010; 140(5):678-691. DOI 10.1016/j.cell.2010.01.003.

Hammond C.M., Stromme C.B., Huang H., Patel D.J., Groth A. Histone chaperone networks shaping chromatin function. Nat. Rev. Mol. Cell. Biol. 2017;18(3):141-158. DOI 10.1038/nrm.2016.159.

Hanai K., Furuhashi H., Yamamoto T., Akasaka K., Hirose S. RSF governs silent chromatin formation via histone H2Av replacement. PLoS Genet. 2008;4(2):e1000011. DOI 10.1371/journal.pgen.1000011.

Harada A., Okada S., Konno D., Odawara J., Yoshimi T., Yoshimura S., Kumamaru H., Saiwai H., Tsubota T., Kurumizaka H., Akashi K., Tachibana T., Imbalzano A.N., Ohkawa Y. Chd2 interacts with H3.3 to determine myogenic cell fate. EMBO J. 2012;31(13):2994-3007. DOI 10.1038/emboj.2012.136.

Hartlepp K.F., Fernandez-Tornero C., Eberharter A., Grune T., Muller C.W., Becker P.B. The histone fold subunits of *Drosophila* CHRAC facilitate nucleosome sliding through dynamic DNA interactions. Mol. Cell. Biol. 2005;25(22):9886-9896.

Hauer M.H., Gasser S.M. Chromatin and nucleosome dynamics in DNA damage and repair. Genes Dev. 2017;31(22):2204-2221. DOI 10.1101/gad.307702.117.

Haushalter K.A., Kadonaga J.T. Chromatin assembly by DNA-translocating motors. Nat. Rev. Mol. Cell. Biol. 2003;4(8):613-620.

Henikoff S., Ahmad K. Assembly of variant histones into chromatin. Annu. Rev. Cell Dev. Biol. 2005;21:133-153.

Hennig B.P., Bendrin K., Zhou Y., Fischer T. Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. EMBO Rep. 2012;13(11):997-1003. DOI 10.1038/embor.2012.146.

Ito T., Bulger M., Pazin M.J., Kobayashi R., Kadonaga J.T. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell. 1997;90(1):145-155.

Ito T., Levenstein M.E., Fyodorov D.V., Kutach A.K., Kobayashi R., Kadonaga J.T. ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. Genes Dev. 1999;15;13(12):1529-1539.

Juhasz S., Elbakry A., Mathes A., Lobrich M. ATRX promotes DNA repair synthesis and sister chromatid exchange during homologous recombination. Mol. Cell. 2018;71(1):11-24 e7. DOI 10.1016/j.molcel.2018.05.014.

Kari V., Mansour W.Y., Raul S.K., Baumgart S.J., Mund A., Grade M., Sirma H., Simon R., Will H., Dobbelstein M., Dikomey E., Johnsen S.A. Loss of CHD1 causes DNA repair defects and enhances prostate cancer therapeutic responsiveness. EMBO Rep. 2016; 19(10):pii: e46783. DOI 10.15252/embr.201846783.

Khuong M.T., Fei J., Cruz-Becerra G., Kadonaga J.T. A simple and versatile system for the ATP-dependent assembly of chromatin. J. Biol. Chem. 2017;292(47):19478-19490. DOI 10.1074/jbc.M117.815365.

Konev A.Y., Tiutiunnik A.A., Baranovskaya I.L. The influence of the *Chd1* chromatin assembly and remodeling factor mutations on *Drosophila* polythene chromosome organization. Tsitologiia. 2016; 58(4):281-284. (in Russian)

Konev A.Y., Tribus M., Park S.Y., Podhraski V., Lim C.Y., Emelyanov A.V., Vershilova E., Pirrotta V., Kadonaga J.T., Lusser A., Fyodorov D.V. CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin *in vivo*. Science. 2007;317(5841): 1087-1090.

Krude T., Keller C. Chromatin assembly during S phase: contributions from histone deposition, DNA replication and the cell division cycle. Cell. Mol. Life Sci. 2001;58(5-6):665-672.

Kukimoto I., Elderkin S., Grimaldi M., Oelgeschlager T., Varga-Weisz P.D. The histone-fold protein complex CHRAC-15/17 en-

hances nucleosome sliding and assembly mediated by ACF. Mol. Cell. 2004;13(2):265-277.

Lee J.S., Garrett A.S., Yen K., Takahashi Y.H., Hu D., Jackson J., Seidel C., Pugh B.F., Shilatifard A. Codependency of H2B monoubiquitination and nucleosome reassembly on Chd1. Genes Dev. 2012; 26(9):914-919. DOI 10.1101/gad.186841.112.

Lee Y., Park D., Iyer V.R. The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes. Nucleic Acids Res. 2017;45(12):7180-7190. DOI 10.1093/nar/gkx321.

Lewis P.W., Elsaesser S.J., Noh K.M., Stadler S.C., Allis C.D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc. Natl. Acad. Sci. USA. 2010;107(32):14075-14080. DOI 10.1073/pnas.1008850107.

Liu J.C., Ferreira C.G., Yusufzai T. Human CHD2 is a chromatin assembly ATPase regulated by its chromo- and DNA-binding domains. J. Biol. Chem. 2015;290(1):25-34. DOI 10.1074/jbc.M114.609156.

Loppin B., Bonnefoy E., Anselme C., Laurencon A., Karr T.L., Couble P. The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. Nature. 2005;4437(7063):1386-1390.

Loyola A., Huang J.Y., LeRoy G., Hu S., Wang Y.H., Donnelly R.J., Lane W.S., Lee S.C., Reinberg D. Functional analysis of the subunits of the chromatin assembly factor RSF. Mol. Cell. Biol. 2003; 23(19):6759-6768.

Luijsterburg M.S., de Krijger I., Wiegant W.W., Shah R.G., Smeenk G., de Groot A.J., Pines A., Vertegaal A.C., Jacobs J.J., Shah G.M., van Attikum H. PARP1 links CHD2-mediated chromatin expansion and H3.3 deposition to DNA repair by non-homologous end-joining. Mol. Cell. 2016;61(4):547-562. DOI 10.1016/j.molcel.2016.01.019.

Lusser A., Urwin D.L., Kadonaga J.T. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. Nat. Struct. Mol. Biol. 2005;12(2):160-166.

Morettini S., Podhraski V., Lusser A. ATP-dependent chromatin remodeling enzymes and their various roles in cell cycle control. Front. Biosci. 2008;13:5522-5532.

Morettini S., Tribus M., Zeilner A., Sebald J., Campo-Fernandez B., Scheran G., Worle H., Podhraski V., Fyodorov D.V., Lusser A. The chromodomains of CHD1 are critical for enzymatic activity but less important for chromatin localization. Nucleic Acids Res. 2011; 39(8):3103-3115. DOI 10.1093/nar/gkq1298.

Narlikar G.J., Sundaramoorthy R., Owen-Hughes T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell. 2013;54(3):490-503. DOI 10.1016/j.cell.2013.07.011.

Ohzeki J., Shono N., Otake K., Martins N.M., Kugou K., Kimura H., Nagase T., Larionov V., Earnshaw W.C., Masumoto H. KAT7/HBO1/MYST2 regulates CENP-A chromatin assembly by antagonizing Suv39h1-mediated centromere inactivation. Dev. Cell. 2016; 37(5):413-427. DOI 10.1016/j.devcel.2016.05.006.

Okada M., Okawa K., Isobe T., Fukagawa T. CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. Mol. Biol. Cell. 2009;20(18):3986-3995. DOI 10.1091/mbc.E09-01-0065.

Orsi G.A., Algazeery A., Meyer R.E., Capri M., Sapey-Triomphe L.M., Horard B., Gruffat H., Couble P., Ait-Ahmed O., Loppin B. *Drosophila* Yemanuclein and HIRA cooperate for de novo assembly of H3.3-containing nucleosomes in the male pronucleus. PLoS Genet. 2013;9(2):e1003285. DOI 10.1371/journal.pgen.1003285.

Perpelescu M., Nozaki N., Obuse C., Yang H., Yoda K. Active establishment of centromeric CENP-A chromatin by RSF complex. J. Cell Biol. 2009;185(3):397-407. DOI 10.1083/jcb.200903088.

Piatti P., Lim C.Y., Nat R., Villunger A., Geley S., Shue Y.T., Soratroi C., Moser M., Lusser A. Embryonic stem cell differentiation requires full length Chd1. Sci. Rep. 2015;5:8007. DOI 10.1038/srep0800.

Podhraski V., Campo-Fernandez B., Worle H., Piatti P., Niederegger H., Bock G., Fyodorov D.V., Lusser A. CenH3/CID incorporation is not dependent on the chromatin assembly factor CHD1 in *Drosophila*. PLoS One. 2010;5(4):e10120. DOI 10.1371/journal.pone.0010120.

Pointner J., Persson J., Prasad P., Norman-Axelsson U., Stralfors A., Khorosjutina O., Krietenstein N., Svensson J.P., Ekwall K., Korber P. CHD1 remodelers regulate nucleosome spacing *in vitro* and align nucleosomal arrays over gene coding regions in *S. pombe*. EMBO J. 2012;31(23):4388-4403. DOI 10.1038/emboj.2012.289.

Radman-Livaja M., Quan T.K., Valenzuela L., Armstrong J.A., van Welsem T., Kim T., Lee L.J., Buratowski S., van Leeuwen F., Rando O.J., Hartzog G.A. A key role for Chd1 in histone H3 dynamics at the 3′ ends of long genes in yeast. PLoS Genet. 2012;8(7):e1002811. DOI 10.1371/journal.pgen.1002811.

Robinson K.M., Schultz M.C. Replication-independent assembly of nucleosome arrays in a novel yeast chromatin reconstitution system involves antisilencing factor Asf1p and chromodomain protein Chd1p. Mol. Cell. Biol. 2003;23(22):7937-7946.

Scacchetti A., Brueckner L., Jain D., Schauer T., Zhang X., Schnorrer F., van Steensel B., Straub T., Becker P.B. CHRAC/ACF contribute to the repressive ground state of chromatin. Life Sci. Alliance. 2018;1(1):e201800024. DOI 10.26508/lsa.201800024.

Schneiderman J.I., Orsi G.A., Hughes K.T., Loppin B., Ahmad K. Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. Proc. Natl. Acad. Sci. USA. 2012;109(48): 19721-19726. DOI 10.1073/pnas.1206629109.

Schneiderman J.I., Sakai A., Goldstein S., Ahmad K. The XNP remodeler targets dynamic chromatin in *Drosophila*. Proc. Natl. Acad. Sci. USA. 2009;106(34):14472-14477. DOI 10.1073/pnas. 0905816106.

Semba Y., Harada A., Maehara K., Oki S., Meno C., Ueda J., Yamagata K., Suzuki A., Onimaru M., Nogami J., Okada S., Akashi K., Ohkawa Y. Chd2 regulates chromatin for proper gene expression toward differentiation in mouse embryonic stem cells. Nucleic Acids Res. 2017;6;45(15):8758-8772. DOI 10.1093/nar/gkx475.

Serra-Cardona A., Zhang Z. Replication-coupled nucleosome assembly in the passage of epigenetic information and cell identity. Trends Biochem. Sci. 2018;43(2):136-148. DOI 10.1016/j.tibs.2017.12.003.

Shenoy T.R., Boysen G., Wang M.Y., Xu Q.Z., Guo W., Koh F.M., Wang C., Zhang L.Z., Wang Y., Gil V., Aziz S., Christova R., Rodrigues D.N., Crespo M., Rescigno P., Tunariu N., Riisnaes R., Zafeiriou Z., Flohr P., Yuan W., Knight E., Swain A., Ramalho-Santos M., Xu D.Y., de Bono J., Wu H. CHD1 loss sensitizes prostate cancer to DNA damaging therapy by promoting error-prone double-strand break repair. Ann. Oncol. 2017;28(7):1495-1507. DOI 10.1093/annonc/mdx165.

Simic R., Lindstrom D.L., Tran H.G., Roinick K.L., Costa P.J., Johnson A.D., Hartzog G.A., Arndt K.M. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J. 2003;22(8):1846-1856.

Siriaco G., Deuring R., Chioda M., Becker P.B., Tamkun J.W. *Drosophila* ISWI regulates the association of histone H1 with interphase chromosomes *in vivo*. Genetics. 2009;182(3):661-669. DOI 10.1534/genetics.109.10205.

Smolle M., Venkatesh S., Gogol M.M., Li H., Zhang Y., Florens L., Washburn M.P., Workman J.L. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat. Struct. Mol. Biol. 2012;19(9):884-892. DOI 10.1038/nsmb.2312.

Strohner R., Nemeth A., Jansa P., Hofmann-Rohrer U., Santoro R., Langst G., Grummt I. NoRC-a novel member of mammalian ISWI-containing chromatin remodeling machines. EMBO J. 2001;20(17): 4892-4900.

Tagami H., Ray-Gallet D., Almouzni G., Nakatani Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell. 2004;1116(1):51-61.

Torigoe S.E., Patel A., Khuong M.T., Bowman G.D., Kadonaga J.T. ATP-dependent chromatin assembly is functionally distinct from chromatin remodeling. Elife. 2013;2:e00863. DOI 10.7554/eLife.00863.

Torigoe S.E., Urwin D.L., Ishii H., Smith D.E., Kadonaga J.T. Identification of a rapidly formed nonnucleosomal histone-DNA inter-

mediate that is converted into chromatin by ACF. Mol. Cell. 2011; 43(4):638-648. DOI 10.1016/j.molcel.2011.07.017.

Torres-Padilla M.E., Bannister A.J., Hurd P.J., Kouzarides T., Zernicka-Goetz M. Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int. J. Dev. Biol. 2006;50(5):455-461.

Tyler J.K. Chromatin assembly. Cooperation between histone chaperones and ATP-dependent nucleosome remodeling machines. Eur. J. Biochem. 2002;269(9):2268-2274.

Vary J.C. Jr., Gangaraju V.K., Qin J., Landel C.C., Kooperberg C., Bartholomew B., Tsukiyama T. Yeast Isw1p forms two separable complexes *in vivo*. Mol. Cell. Biol. 2003;23(1):80-91.

Venkatesh S., Workman J.L. Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol. 2015; 16(3):178-189. DOI 10.1038/nrm3941.

Vincent J.A., Kwong T.J., Tsukiyama T. ATP-dependent chromatin remodeling shapes the DNA replication landscape. Nat. Struct. Mol. Biol. 2008;15(5):477-484. DOI 10.1038/nsmb.1419.

Walfridsson J., Bjerling P., Thalen M., Yoo E.J., Park S.D., Ekwall K. The CHD remodeling factor Hrp1 stimulates CENP-A loading to centromeres. Nucleic Acids Res. 2005;33(9):2868-2879.

Wong L.H., McGhie J.D., Sim M., Anderson M.A., Ahn S., Hannan R.D., George A.J., Morgan K.A., Mann J.R., Choo K.H. ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res. 2010;20(3):351-360. DOI 10.1101/gr.101477.109.

Xella B., Goding C., Agricola E., Di Mauro E., Caserta M. The ISWI and CHD1 chromatin remodelling activities influence ADH2 expression and chromatin organization. Mol. Microbiol. 2006;59(5): 1531-1541.

Yadav T., Whitehouse I. Replication-coupled nucleosome assembly and positioning by ATP-dependent chromatin-remodeling enzymes. Cell Rep. 2016;15(4):715-723. DOI 10.1016/j.celrep.2016.03.059.

Yang J.H., Song Y., Seol J.H., Park J.Y., Yang Y.J., Han J.W., Youn H.D., Cho E.J. Myogenic transcriptional activation of MyoD mediated by replication-independent histone deposition. Proc. Natl. Acad. Sci. USA. 2011;108(1):85-90. DOI 10.1073/pnas.1009830108.

Zhou J., Li J., Serafim R.B., Ketchum S., Ferreira C.G., Liu J.C., Coe K.A., Price B.D., Yusufzai T. Human CHD1 is required for early DNA-damage signaling and is uniquely regulated by its N terminus. Nucleic Acids Res. 2018;46(8):3891-3905. DOI 10.1093/nar/gky128.

Acknowledgements. This work was partially supported by RFBR, grant 15-04-99583.

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised December 25, 2018. Accepted December 25, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):168-173
DOI 10.18699/VJ19.477

Functional properties of the Su(Hw) complex are determined by its regulatory environment and multiple interactions on the Su(Hw) protein platform

L.S. Melnikova ✉, M.V. Kostyuchenko, V.V. Molodina, P.G. Georgiev, A.K. Golovnin

Institute of Gene Biology, RAS, Moscow, Russia
✉ e-mail: lsm73@mail.ru

The Su(Hw) protein was first identified as a DNA-binding component of an insulator complex in *Drosophila*. Insulators are regulatory elements that can block the enhancer-promoter communication and exhibit boundary activity. Some insulator complexes contribute to the higher-order organization of chromatin in topologically associated domains that are fundamental elements of the eukaryotic genomic structure. The Su(Hw)-dependent protein complex is a unique model for studying the insulator, since its basic structural components affecting its activity are already known. However, the mechanisms involving this complex in various regulatory processes and the precise interaction between the components of the Su(Hw) insulators remain poorly understood. Our recent studies reveal the fine mechanism of formation and function of the Su(Hw) insulator. Our results provide, for the first time, an example of a high complexity of interactions between the insulator proteins that are required to form the (Su(Hw)/Mod(mdg4)-67.2/CP190) complex. All interactions between the proteins are to a greater or lesser extent redundant, which increases the reliability of the complex formation. We conclude that both association with CP190 and Mod(mdg4)-67.2 partners and the proper organization of the DNA binding site are essential for the efficient recruitment of the Su(Hw) complex to chromatin insulators. In this review, we demonstrate the role of multiple interactions between the major components of the Su(Hw) insulator complex (Su(Hw)/Mod(mdg4)-67.2/CP190) in its activity. It was shown that Su(Hw) may regulate the enhancer–promoter communication via the newly described insulator neutralization mechanism. Moreover, Su(Hw) participates in direct regulation of activity of vicinity promoters. Finally, we demonstrate the mechanism of organization of "insulator bodies" and suggest a model describing their role in proper binding of the Su(Hw) complex to chromatin.
Key words: Su(Hw); Mod(mdg4)-67.2; CP190; transcription regulation; insulation; insulator bodies; protein-protein interactions.

For citation: Melnikova L.S., Kostyuchenko M.V., Molodina V.V., Georgiev P.G., Golovnin A.K. Functional properties of the Su(Hw) complex are determined by its regulatory environment and multiple interactions on the Su(Hw) protein platform. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):168-173. DOI 10.18699/VJ19.477

Функциональные свойства Su(Hw)-зависимого комплекса определяются его регуляторным окружением и множественными взаимодействиями на белковой платформе Su(Hw)

Л.С. Мельникова ✉, М.В. Костюченко, В.В. Молодина, П.Г. Георгиев, А.К. Головнин

Институт биологии гена Российской академии наук, Москва, Россия
✉ e-mail: lsm73@mail.ru

Белок Su(Hw) был впервые идентифицирован как ДНК-связывающий компонент инсуляторного комплекса. Инсуляторы представляют собой регуляторные элементы, которые могут блокировать энхансер-промоторные взаимодействия и работать как границы между активным и репрессивным хроматином. Su(Hw)-зависимый белковый комплекс – уникальная модель для изучения инсуляторов, поскольку основные структурные компоненты, влияющие на его активность, уже известны. Однако механизмы, вовлекающие этот комплекс в различные регуляторные процессы, и детали взаимодействий между компонентами Su(Hw) инсуляторов остаются недостаточно изученными. Наши недавние работы выявили детальный механизм формирования и функционирования инсулятора Su(Hw). В представленном обзоре мы демонстрируем, как множественные взаимодействия между основными компонентами Su(Hw)-зависимого комплекса (Su(Hw)/Mod(mdg4)-67.2/CP190) влияют на его активность. Показываем, что Su(Hw) может регулировать энхансер-промоторные взаимодействия через новый механизм нейтрализации инсулятора и, кроме того, участвует в прямой регуляции активности близлежащих промоторов. Наконец, мы описываем механизм формирования

«инсуляторных телец» и предлагаем модель, объясняющую их роль в рекрутировании Su(Hw)-зависимого комплекса на хроматин.
Ключевые слова: Su(Hw); Mod(mdg4)-67.2; CP190; регуляция транскрипции; инсуляция; инсуляторные тельца; белок-белковые взаимодействия.

Multiple interactions between the components of the Su(Hw) complex

The best-studied *Drosophila* insulator complex consists of two proteins, the Mod(mdg4)-67.2 and CP190, which are recruited to chromatin through interactions with the DNA-binding Su(Hw) protein. The Su(Hw) protein contains the N-terminal region, an array of 12 C_2H_2-type zinc finger domains (ZF), and the C-terminal region (aa 716–892) responsible for enhancer blocking activity (Harrison et al., 1993). Thus far, little is known about the precise mechanism of organization of the Su(Hw) complex. Mod(mdg4)-67.2 is one of the isoforms produced by the *mod(mdg4)* locus, which encodes a large set of proteins with common N- and different C-terminal parts (Dorn, Krauss, 2003). The common part contains the BTB domain that functions as a protein interaction domain facilitating oligomer formation (Golovnin et al., 2007; Bonchuk et al., 2011). In its specific part, each Mod(mdg4) isoform includes the degenerative FLYWCH domain presumably employed in protein-protein interaction. Mod(mdg4)-67.2 has an additional C-terminal acidic domain (SID), which interacts with the region of the Su(Hw) protein responsible for the enhancer-blocking effect (Golovnin et al., 2007). Another Su(Hw) binding protein CP190 also contains the N-terminal BTB/POZ domain, which forms stable homodimers that may be involved in protein-protein interactions (Bonchuk et al., 2011, 2015). The CP190 is a shuttle protein, since it localizes to centrioles via its M domain during the mitotic stage (Plevock et al., 2015), while being recruited to chromatin via its interaction with the Su(Hw) and many other insulator proteins during the interphase (Melnikova et al., 2018a). An array of four ZF domains is probably involved in protein-protein interactions, since no interaction between DNA and CP190 protein has been identified thus far despite the large number of studies conducted.The available data have been expanded and summarized in our recent studies. CP190 was shown to interact with the N-terminal part of the Su(Hw) protein located between aa 88 and 202 via its BTB domain (Melnikova et al., 2018a). Interestingly, some other known architectural/insulator proteins, such as dCTCF and Pita, also interact with CP190 via its BTB domain (Bonchuk et al., 2015; Maksimenko et al., 2015). Therefore, it is possible that specificity of CP190 interactions with DNA-binding proteins is determined by additional components. In the case of Su(Hw) insulator, this role can be played by Mod(mdg4)-67.2 as we have identified that the BTB domain of Mod(mdg4)-67.2 directly interacts with the M domain of CP190 (Golovnin et al., 2007). The results of yeast two-hybrid assay show that the FLYWCH domain improves the interaction between the BTB domain of Mod(mdg4)-67.2 and the M domain of CP190. Both CP190 and Mod(mdg4)-67.2 participate in recruitment of Su(Hw) to chromatin as the knock-down of each of them reduces Su(Hw) binding to the genomic sites (Melnikova et al., 2017a, 2018a). According to our results, it appears that the specificity of organization of the Su(Hw) complex and its recruitment to chromatin is achieved by complex interactions of the Mod(mdg4)-67.2 SID, FLYWCH, and BTB domains with CP190/Su(Hw) proteins. It seems likely that Mod(mdg4)-67.2 is recruited to the Su(Hw) sites in complex with CP190 (Fig. 1).

Identification of the first endogenous Su(Hw)-dependent insulators

The *Drosophila gypsy* insulator including a set of twelve degenerative binding sites for Su(Hw) protein was found in the 5′ region of the *gypsy* retrotransposon (Parkhurst et al., 1988). This insulator exhibiting enhancer blocking and boundary activities. For example, the *gypsy* insulator blocks enhancer-promoter interactions when placed between the enhancers and a promoter in the model system of the *yellow* gene.

However, many endogenous Su(Hw) binding sites observed in polytene chromosomes are not associated with *gypsy* retrotransposon. The functional Su(Hw) insulator between the *yellow* gene and AS-C, which was able to block the enhancer-promotor communication of the *scute* gene over significant distances was first described using *in vivo* and *in vitro* assays (Golovnin et al., 1999, 2003; Parnell et al., 2003). This insulator site (later named the 1A2 insulator), unlike the *gypsy* retrotransposone insulator, includes only two Su(Hw) binding sites. Interestingly, in transgenic lines, the small 126 bp fragment that includes only two Su(Hw) binding sites can only

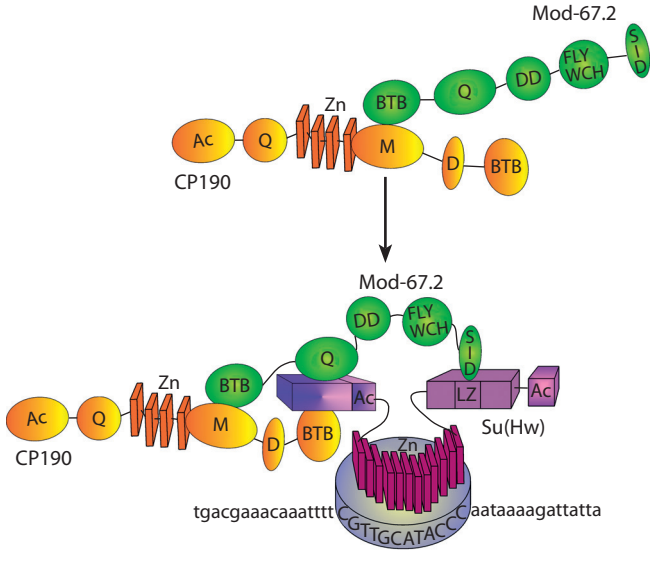


Fig. 1. Scheme of protein-protein interactions involved in formation of the (Su(Hw)/Mod(mdg4)-67.2/CP190) complex.

The CP190 domains are shown as yellow ovals and four zinc fingers, as yellow parallelepipeds; the Mod(mdg4)-67.2 domains are shown as green ovals; the Su(Hw) domains, as purple parallelepipeds. Bold capital letters indicate the Su(Hw) binding site. Domain abbreviations: Ac – acidic domains; Zn – zinc-finger domains; LZ – leucine zipper motif; BTB – BTB/POZ domain; Q – glutamine-rich (Q-rich) region; DD – dimerization domain; FLYWCH – FLYWCH-type zinc finger domain; SID – Su(Hw) interaction domain; D – aspartic acid-rich (D-rich) domain; M– centrosomal targeting domain.

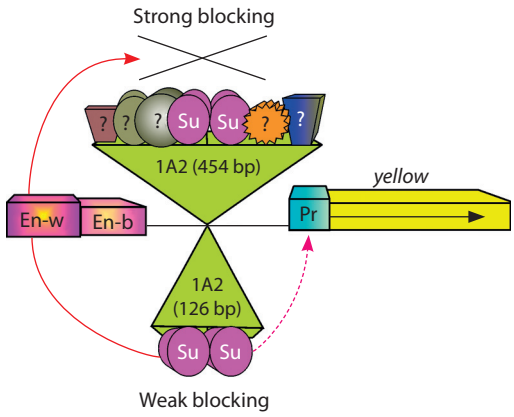


Fig. 2. The scheme of blocking the enhancer-promoter interactions with different fragments of 1A2 insulator in the model system of the *yellow* gene.

The boxes indicate the *yellow* promoter (Pr) and the wing (En-w) and body (En-b) enhancers. The 1A2 insertions are shown as triangles in which the purple ovals (Su) are the binding sites of the Su(Hw) protein and different figures with the question mark in them are unidentified proteins. The arrow indicates the direction of *yellow* gene transcription.

partially block the strong *yellow* enhancer, while the larger 454 bp fragment that includes the same Su(Hw) sites and neighboring sequences, completely blocks *yellow* enhancers. Thus, additional proteins binding to these sequences are required for strong insulator activity of the 1A2 insulator (Fig. 2).

It seems possible that in endogenous insulators, Su(Hw) cooperates with additional DNA-binding proteins to support insulator activity. Many endogenous Su(Hw)-dependent insulators were later identified and described (Parnell et al., 2006; Soshnev et al., 2011; Schwartz et al., 2012), suggesting that they play a significant role in gene regulation. Most of them have 2–4 binding sites for the Su(Hw) protein. They show different blocking activities, which do not directly correlate with the number of Su(Hw)-binding sites (Kuhn-Parnell et al., 2008). This fact is in line with the hypothesis that functionality of the Su(Hw) complex depends on the regulatory environment of Su(Hw) binding sites.

The mechanism of recruiting the Su(Hw) complexes to chromatin

As it follows from the limited experimental data, only some of the C_2H_2 domains of the Su(Hw) protein are usually involved in DNA binding, while the rest are involved in interactions with proteins or RNAs (Kim et al., 1996; Lei, Corces, 2006). Most of the genomic Su(Hw) binding regions have 2–4 consensus sites, while studies on transgenic lines have shown that only four reiterated binding sites for Su(Hw) can function effectively (Scott et al., 1999). Recent investigations suggest that the affinity of Su(Hw) for DNA and organization of the Su(Hw) complex may be dependent on the nature of the consensus site (Baxley et al., 2017). This model was described as the "Su(Hw) code", which assumed that Su(Hw) binds to a compound consensus sequence (approximately 26 nucleotide-long) consisting of three modules: the ZF6–ZF9 cluster binds to the main, central module; the ZF2–ZF4 cluster, to the downstream CG-rich module; and the ZF10–ZF12 cluster, to the upstream AT-rich module. The "Su(Hw) code" model predicts that the consensus site determines which of the Su(Hw) ZF domains are not involved in DNA binding and, hence, are free to interact with other proteins and/or complexes. In our recent study, we compared the influence of two mutations affecting the same tenth ZF (ZF10) of Su(Hw) protein on properties of *gypsy* insulator (Melnikova et al., 2018b). According to the "Su(Hw) code" model, *gypsy* insulator was mainly organized by the upstream and central modules (Baxley et al., 2017). We used the *su(Hw)^f* mutation involving a point substitution that affects the critical Zn-coordinating residue and thereby dramatically alters the structure of the C_2H_2 domain and leads to full deletion of ZF10 (Su(Hw)$^{\Delta 10}$). It was found that Su(Hw)^f becomes unable to interact with CP190 in the absence of DNA, while Su(Hw)$^{\Delta 10}$ restores the CP190–Su(Hw) interaction. These results suggest that the point mutation in ZF10 influencing the Su(Hw)^f conformation results in loss of interaction with CP190. The Su(Hw)$^{\Delta 10}$ mutant binds to the *gypsy* insulator better than Su(Hw)^f does and partially restores its enhancer-blocking activity but, in contrast to the wild-type "Su(Hw) protein", fails to interact with the *gypsy* insulator in the absence of Mod(mdg4)-67.2 protein *(mod(mdg4)^{u1}* mutant background) (Melnikova et al., 2018b). These results suggest that CP190 and Mod(mdg4)-67.2 are critical for binding of the mutant Su(Hw)$^{\Delta 10}$ to the *gypsy* insulator. Based on these findings we can draw a conclusion that both association with CP190 and Mod(mdg4)-67.2 partners and proper organization of DNA binding site are essential for efficient recruitment of the Su(Hw) complex.

Looping model describes the neutralization effect

However, in the transgenic experiments action of Su(Hw) insulator failed when two insulator sites were introduced at a distance from one another in the region between the enhancer and the promoter (Muravyova et al., 2001). The loss of insulator activity results from pairing between the two insulator complexes. "Looping out" of the sequences between the insulators brings the enhancer and promoter closer and may stimulate expression (Fig 3, *a*).

The insulator neutralization effects were observed only for the same type of insulators (Su(Hw)-Su(Hw)) but not for heterologous pairs (dCTCF-Su(Hw)) (Kyrchanova et al., 2008; Krivega et al., 2010). Such selectivity in insulator interaction implicates the existence of exclusive protein complexes organized on different insulator types. It is likely that these different complexes (despite the fact that all of them share the CP190 protein) are unable to interact with each other. Thus, restriction of interaction between the insulator elements may determinate the specificity of particular enhancer-promoter interaction.

The neutralization effect partially depends on tandem or opposite orientation of the tested insulator sites (Kyrchanova et al., 2008). Testing of the composite insulator fragment containing both dCTCF and Su(Hw) binding sites displayed strong dependence on orientation. Strong promoter stimulation was observed in the case of opposite orientation, while enhancer blocking was preserved in tandem orientation. These results demonstrate that loops having two different configurations and thus exhibiting different effects on enhancer activity can be formed depending on orientation of the insulator site (see Fig. 3, *b*). However, functional interaction between the

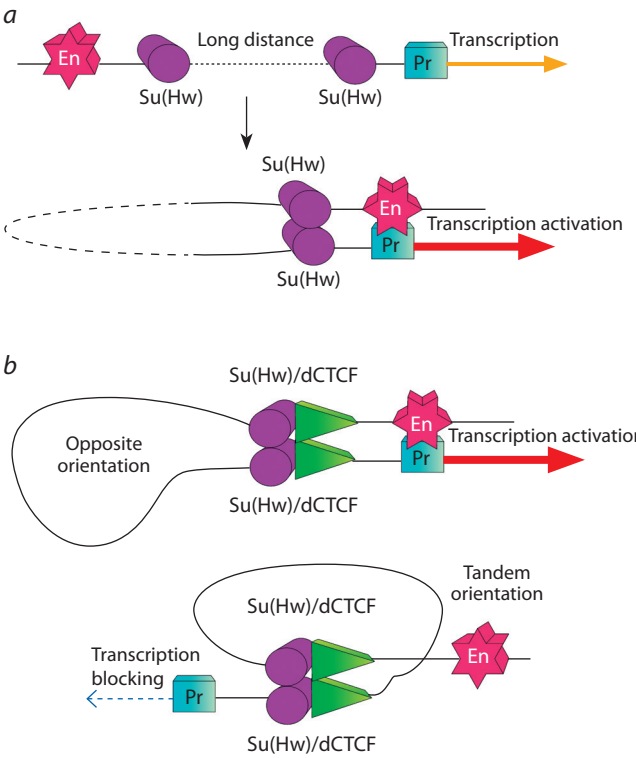


Fig. 3. The models of pairing between the insulators.

a – pairing of insulators (Su(Hw)) can provide the long-distance enhancer (En)–promoter (Pr) interactions; *b* – two models of pairing between the composed insulators (Su(Hw)/dCTCF) having either an opposite or the same orientation with respect to each other.

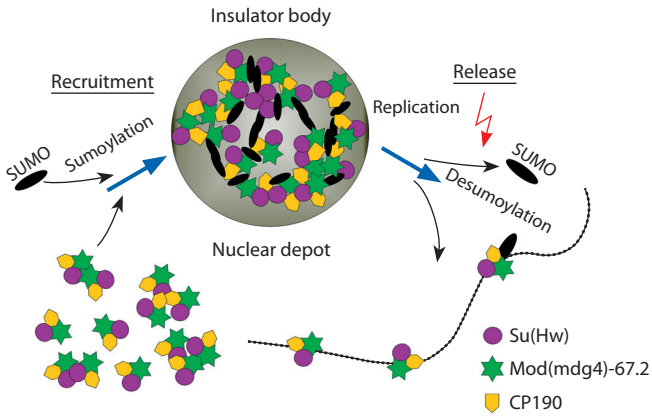


Fig. 4. The model describing the organization of insulator bodies (reproduced from (Golovnin et al., 2012)).

The circle, asterisk, and pentagon represent Mod(mdg4)-67.2, Su(Hw), and CP190 proteins, respectively; the black ovals are SUMO molecules; and the curve shows a chromatin fibril with insulators.

fragments containing only the Su(Hw) binding sites is less dependent on their relative orientation. It seems likely that the complex that can be formed on the Su(Hw) protein by Mod(mdg4)-67.2 and CP190 proteins is not sufficient to determine the long-range loop configuration. Thereby, in the context of endogenous Su(Hw) insulators, additional proteins located in the vicinity of the insulator site may determine the orientation-dependent interaction between insulators. Thus, the action of a particular insulator cannot be considered apart from its regulatory environment.

Direct regulation of gene expression by Su(Hw) insulators

Besides spatial organization of local enhancer-promoter communication, Su(Hw) insulator protein can behave as either a direct activator or repressor of transcription. It was shown that the *yellow* promoter weakened by deletion of the upstream regulatory region can be stimulated by the Su(Hw) protein in a distance-dependent manner. (Golovnin et al., 2005). Introduction of *su(Hw)*$^{-}$ background (*su(Hw)*^{v}/*su(Hw)*^{f}) completely abolishes the stimulation effect. This effect depends neither on the repression activity of neighboring regulatory elements nor on the boundary activity of Su(Hw) insulator as promoter stimulation was observed when the insulator was placed either upstream or downstream from the *yellow* promoter. This finding is in line with further studies that reported the stimulator effect of Su(Hw) on alcohol dehydrogenase promoter (Wei, Brennan, 2001) and weak *gypsy* promoter (Parkhurst, Corces, 1986; Smith, Corces, 1995). On the other hand, Su(Hw) may repress transcription of neural genes in ovaries in a tissue-specific manner (Soshnev et al., 2013). Hence, Su(Hw) protein may represent a platform that organizes the specific regulatory complex to provide particular activities.

"Maturation" of the insulator complex in the nucleus and it targeting to chromatin

The Su(Hw), Mod(mdg4)-67.2, and CP190 proteins in the interphase cell nucleus co-localize in discrete foci/speckles named "the insulator bodies" (Golovnin et al., 2007, 2008, 2012). It has been assumed that the "insulator bodies" arise via association of individual Su(Hw)-containing nucleoprotein complexes located at distant chromosomal sites, owing to interactions between the BTB domains of insulator proteins Mod(mdg4)-67.2 and CP190. Thus, it was supposed that the "insulator bodies" are responsible for insulator activity (Gerasimova et al., 1995). However, our results show that insulator bodies are rather aggregates of insulator proteins that comprise many unrelated proteins (Golovnin et al., 2008, 2012, 2015). Moreover, it was shown that the assembly of "insulator bodies" is determined by CP190 protein and sumolyation of Mod(mdg4)-67.2 (Golovnin et al., 2012). Sumolyation of Mod(mdg4)-67.2 is essential for incorporation of this protein and Su(Hw) into the "insulator bodies" but is dispensable for the stability of CP190-dependent speckles (Golovnin et al., 2012). The sumoylated Mod(mdg4)-67.2 and CP190 proteins interact with Su(Hw) and recruit it to the "insulator bodies". This protects the insulator complex from degradation. Moreover, the "insulator bodies" possibly facilitate the formation of complexes between Su(Hw)/Mod(mdg4)-67.2/CP190 and other transcription factors. "Mature" Su(Hw)-dependenent complexes may then transiently interact with chromatin fibril and be detached from the "insulator bodies" by means of desumoylation (Golovnin et al., 2008) (Fig. 4). Thus, the formation of "insulator bodies" does not directly correlate with insulator activity.

This model was confirmed by additional experiments with the EAST protein, which may serve as a structural basis for the

nuclear extrachromosomal compartment (Wasser, Chia, 2000). Altering EAST expression (overexpression or inactivation) affects the enhancer-blocking activity of the *gypsy* insulator, which, in its turn, directly correlates with the efficiency of Mod(mdg4)-67.2 and CP190 binding to the insulators (Melnikova et al., 2017b). There is no evidence that EAST directly binds to chromatin and to Su(Hw) insulators in particular under normal physiological conditions. However, by direct interacting with Mod(mdg4)-67.2 and CP190, EAST may also be directly involved in nucleation of "insulator bodies". The overexpression of EAST leads to segregation of the CP190 protein in independent speckles, thus inhibiting formation of "mature" insulator complexes (Golovnin et al., 2015). As a result, reduction of Mod(mdg4)-67.2 and CP190 proteins on the Su(Hw) binding sites and an increase in repression activity of Su(Hw) insulators were observed.

Conclusions

This review has summarized the results of the recent studies aimed to identify the mechanism of assembly and function of the Su(Hw)-dependent complex, one of the insulator/architectural complexes in *Drosophila*. We have demonstrated that activity of the Su(Hw) complex strongly depends regulatory elements in its vicinity and complex protein organization (tissues and stage of development in particular).

References

Baxley R.M., Bullard J.D., Klein M.W., Fell A.G., Morales-Rosado J.A., Duan T., Geyer P.K. Deciphering the DNA code for the function of the *Drosophila* polydactyl zinc finger protein Suppressor of Hairy-wing. Nucleic Acids Res. 2017;45(8):4463-4478. DOI 10.1093/nar/gkx040.

Bonchuk A., Denisov S., Georgiev P., Maksimenko O. *Drosophila* BTB/POZ domains of "ttk group" can form multimers and selectively interact with each other. J. Mol. Biol. 2011;412(3):423-436. DOI 10.1016/j.jmb.2011.07.052.

Bonchuk A., Maksimenko O., Kyrchanova O., Ivlieva T., Mogila V., Deshpande G., Wolle D., Schedl P., Georgiev P. Functional role of dimerization and CP190 interacting domains of CTCF protein in *Drosophila melanogaster*. BMC Biol. 2015;13:63. DOI 10.1186/s12915-015-0168-7.

Dorn R., Krauss V. The modifier of mdg4 locus in *Drosophila*: functional complexity is resolved by trans splicing. Genetica. 2003;117(2-3):165-177.

Gerasimova T.I., Gdula D.A., Gerasimov D.V., Simonova O., Corces V.G. A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. Cell. 1995;82(4):587-597.

Golovnin A., Biryukova I., Romanova O., Silicheva M., Parshikov A., Savitskaya E., Pirrotta V., Georgiev P. An endogenous Su(Hw) insulator separates the yellow gene from the Achaete-scute gene complex in *Drosophila*. Development. 2003;130(14):3249-3258.

Golovnin A., Gause M., Georgieva S., Gracheva E., Georgiev P. The su(Hw) insulator can disrupt enhancer-promoter interactions when located more than 20 kilobases away from the *Drosophila* achaete-scute complex. Mol Cell. Biol. 1999;19(5):3443-3456.

Golovnin A., Mazur A., Kopantseva M., Kurshakova M., Gulak P.V., Gilmore B., Whitfield W.G., Geyer P., Pirrotta V., Georgiev P. Integrity of the Mod(mdg4)-67.2 BTB domain is critical to insulator function in *Drosophila melanogaster*. Mol. Cell. Biol. 2007;27(3):963-974. DOI 10.1128/MCB.00795-06.

Golovnin A., Melnick E., Mazur A., Georgiev P. *Drosophila* Su(Hw) insulator can stimulate transcription of a weakened yellow promoter over a distance. Genetics. 2005;170(3):1133-1142. DOI 10.1534/genetics.104.034587.

Golovnin A., Melnikova L., Shapovalov I., Kostyuchenko M., Georgiev P. EAST organizes *Drosophila* insulator proteins in the interchromosomal nuclear compartment and modulates CP190 binding to chromatin. PLoS One. 2015;10(10):e0140991. DOI 10.1371/journal.pone.0140991.

Golovnin A., Melnikova L., Volkov I., Kostuchenko M., Galkin A.V., Georgiev P. 'Insulator bodies' are aggregates of proteins but not of insulators. EMBO Rep. 2008;9(5):440-445. DOI 10.1038/embor.2008.32.

Golovnin A., Volkov I., Georgiev P. SUMO conjugation is required for the assembly of *Drosophila* Su(Hw) and Mod(mdg4) into insulator bodies that facilitate insulator complex formation. J. Cell Sci. 2012;125:2064-2074. DOI 10.1242/jcs.100172.

Harrison D.A., Gdula D.A., Coyne R.S., Corces V.G. A leucine zipper domain of the suppressor of Hairy-wing protein mediates its repressive effect on enhancer function. Genes Dev. 1993;7(10):1966-1978.

Kim J., Shen B., Rosen C., Dorsett D. The DNA-binding and enhancer-blocking domains of the *Drosophila* suppressor of Hairy-wing protein. Mol. Cell. Biol. 1996;16(7):3381-3392.

Krivega M., Savitskaya E., Krivega I., Karakozova M., Parshikov A., Golovnin A., Georgiev P. Interaction between a pair of gypsy insulators or between heterologous gypsy and Wari insulators modulates Flp site-specific recombination in *Drosophila melanogaster*. Chromosoma. 2010;119(4):425-434. DOI 10.1007/s00412-010-0268-7.

Kuhn-Parnell E.J., Helou C., Marion D.J., Gilmore B.L., Parnell T.J., Wold M.S., Geyer P.K. Investigation of the properties of non-gypsy suppressor of hairy-wing-binding sites. Genetics. 2008;179(3):1263-1273. DOI 10.1534/genetics.108.087254.

Kyrchanova O., Chetverina D., Maksimenko O., Kullyev A., Georgiev P. Orientation-dependent interaction between *Drosophila* insulators is a property of this class of regulatory elements. Nucleic Acids Res. 2008;36(22):7019-7028. DOI 10.1093/nar/gkn781.

Lei E.P., Corces V.G. RNA interference machinery influences the nuclear organization of a chromatin insulator. Nat. Genet. 2006;38(8):936-941. DOI 10.1038/ng1850.

Maksimenko O., Bartkuhn M., Stakhov V., Herold M., Zolotarev N., Jox T., Buxa M.K., Kirsch R., Bonchuk A., Fedotova A., Kyrchanova O., Renkawitz R., Georgiev P. Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. Genome Res. 2015;25(1):89-99. DOI 10.1101/gr.174169.114.

Melnikova L., Kostyuchenko M., Molodina V., Parshikov A., Georgiev P., Golovnin A. Multiple interactions are involved in a highly specific association of the Mod(mdg4)-67.2 isoform with the Su(Hw) sites in *Drosophila*. Open Biol. 2017a;7(10):pii: 170150. DOI 10.1098/rsob.170150.

Melnikova L., Kostyuchenko M., Molodina V., Parshikov A., Georgiev P., Golovnin A. Interactions between BTB domain of CP190 and two adjacent regions in Su(Hw) are required for the insulator complex formation. Chromosoma. 2018a;127(1):59-71. DOI 10.1007/s00412-017-0645-6.

Melnikova L., Kostyuchenko M., Parshikov A., Georgiev P., Golovnin A. Role of Su(Hw) zinc finger 10 and interaction with CP190 and Mod(mdg4) proteins in recruiting the Su(Hw) complex to chromatin sites in *Drosophila*. PLoS One. 2018b;13(2):e0193497. DOI 10.1371/journal.pone.0193497.

Melnikova L., Shapovalov I., Kostyuchenko M., Georgiev P., Golovnin A. EAST affects the activity of Su(Hw) insulators by two different mechanisms in *Drosophila melanogaster*. Chromosoma. 2017b;126(2):299-311. DOI 10.1007/s00412-016-0596-3.

Muravyova E., Golovnin A., Gracheva E., Parshikov A., Belenkaya T., Pirrotta V., Georgiev P. Loss of insulator activity by paired Su(Hw) chromatin insulators. Science. 2001;291(5503):495-498. DOI 10.1126/science.291.5503.495.

Parkhurst S.M., Corces V.G. Mutations at the suppressor of forked locus increase the accumulation of gypsy-encoded transcripts in *Drosophila melanogaster*. Mol. Cell. Biol. 1986;6(6):2271-2274.

Parkhurst S.M., Harrison D.A., Remington M.P., Spana C., Kelley R.L., Coyne R.S., Corces V.G. The *Drosophila su(Hw)* gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. Genes Dev. 1988;2(10): 1205-1215.

Parnell T.J., Kuhn E.J., Gilmore B.L., Helou C., Wold M.S., Geyer P.K. Identification of genomic sites that bind the *Drosophila* suppressor of Hairy-wing insulator protein. Mol. Cell. Biol. 2006;26(16):5983-5993. DOI 10.1128/MCB.00698-06.

Parnell T.J., Viering M.M., Skjesol A., Helou C., Kuhn E.J., Geyer P.K. An endogenous suppressor of hairy-wing insulator separates regulatory domains in *Drosophila*. Proc. Natl. Acad. Sci. USA. 2003; 100(23):13436-13441. DOI 10.1073/pnas.2333111100.

Plevock K.M., Galletta B.J., Slep K.C., Rusan N.M. Newly characterized region of CP190 associates with microtubules and mediates proper spindle morphology in *Drosophila* stem cells. PLoS One. 2015;10(12):e0144174. DOI 10.1371/journal.pone.0144174.

Schwartz Y.B., Linder-Basso D., Kharchenko P.V., Tolstorukov M.Y., Kim M., Li H.B., Gorchakov A.A., Minoda A., Shanower G., Alekseyenko A.A., Riddle N.C., Jung Y.L., Gu T., Plachetka A., Elgin S.C.R., Kuroda M.I., Park P.J., Savitsky M., Karpen G.H., Pirrotta V. Nature and function of insulator protein binding sites in the *Drosophila* genome. Genome Res. 2012;22(11):2188-2198. DOI 10.1101/gr.138156.112.

Scott K.C., Taubman A.D., Geyer P.K. Enhancer blocking by the *Drosophila gypsy* insulator depends upon insulator anatomy and enhancer strength. Genetics. 1999;153(2):787-798.

Smith P.A., Corces V.G. The suppressor of Hairy-wing protein regulates the tissue-specific expression of the *Drosophila gypsy* retrotransposon. Genetics. 1995;139(1):215-228.

Soshnev A.A., Baxley R.M., Manak J.R., Tan K., Geyer P.K. The insulator protein Suppressor of Hairy-wing is an essential transcriptional repressor in the *Drosophila* ovary. Development. 2013; 140(17):3613-3623. DOI 10.1242/dev.094953.

Soshnev A.A., Ishimoto H., McAllister B.F., Li X., Wehling M.D., Kitamoto T., Geyer P.K. A conserved long noncoding RNA affects sleep behavior in *Drosophila*. Genetics. 2011;189(2):455-468. DOI 10.1534/genetics.111.131706.

Wasser M., Chia W. The EAST protein of *Drosophila* controls an expandable nuclear endoskeleton. Nat. Cell Biol. 2000;2(5):268-275. DOI 10.1038/35010535.

Wei W., Brennan M.D. The *gypsy* insulator can act as a promoter-specific transcriptional stimulator. Mol. Cell. Biol. 2001;21(22):7714-7720. DOI 10.1128/MCB.21.22.7714-7720.2001.

Acknowledgements. This study was supported by the Russian Science Foundation (project No. 18-14-00295).

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised December 17, 2018. Accepted December 17, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):174-179
DOI 10.18699/VJ19.478

The role of SAGA in the transcription and export of mRNA

E.N. Nabirochkina ✉, M.M. Kurshakova, S.G. Georgieva, D.V. Kopytova

Institute of Gene Biology, RAS, Moscow, Russia
✉ e-mail: elenan5@rambler.ru

SAGA/TFTC, which is a histone acetyltransferase complex, plays an important role in the regulation of transcription. We have identified that the metazoan TFTC/STAGA complexes had histone H2A and H2B deubiquitinase activity that is carried out by a DUBm (deubiquitination module). We studied the DUBm of SAGA in *Drosophila melanogaster* and identified *Drosophila* homologs of yeast DUBm components. Two subunits of DUBm (Sus1/ENY2 and Sgf11) were shown to have functions separate from DUBm function. Thus, Sus1/ENY2 was shown to be present in several different complexes. Sgf11 was found to be associated with the cap-binding complex (CBC) and recruited onto growing messenger ribonucleic acid (mRNA). Also, we have shown that Sgf11 interacted with the TREX-2/AMEX mRNA export complex and was essential for mRNA export from the nucleus. Immunostaining of the polytene chromosomes of *Drosophila* larvae revealed that Sgf11 is present at the sites of localization of snRNA genes. It was also found in immunostaining experiments that dPbp45, the subunit of the PBP complex, the key player in the snRNA transcription process, is associated not only with the snRNA gene localization sites, but with other sites of active transcription by PolII. We also revealed that Sgf11 was present at many active transcription sites in interbands and puffs on polytene chromosomes, Sgf11 was localized at all Brf1 (the component of the RNA polymerase III basal transcription complex) sites. We concluded that SAGA coactivated transcription of both the PolII and PolIII-dependent snRNA genes.
Key words: transcription; SAGA; DUB module; AMEX; Sgf11; ENY2; snRNA genes; PolII-dependent transcription; PolIII-dependent transcription.

For citation: Nabirochkina E.N., Kurshakova M.M., Georgieva S.G., Kopytova D.V. The role of SAGA in the transcription and export of mRNA. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2): 174-179. DOI 10.18699/VJ19.478

Роль SAGA комплекса в транскрипции и экспорте мРНК

Е.Н. Набирочкина ✉, М.М. Куршакова, С.Г. Георгиева, Д.В. Копытова

Институт биологии гена Российской академии наук, Москва, Россия
✉ e-mail: elenan5@rambler.ru

Гистон-ацетил трансферазный комплекс SAGA/TFTC играет важную роль в регуляции транскрипции. Нами было обнаружено, что TFTC/STAGA комплексы метазоа деубиквитинируют гистоны H2Aи H2B. Был определен модуль TFTC/STAGA комплекса, который обладает деубиквитиназной активностью (DUBm). Мы исследовали DUBm у *Drosophila melanogaster* и идентифицировали гомологи – компоненты DUBm дрожжей. Нами было показано, что белок Sgf11, один из компонентов DUBm у *Drosophila*, обладает другой, отличной от DUBm функцией. Белок Sgf11 ассоциирован с CAP-содержащим комплексом и рекрутируется на растущую матричную рибонуклеиновую кислоту (мРНК). Кроме того, мы обнаружили, что Sgf11 взаимодействует с TREX-2/AMEX комплексом экспорта мРНК, и этот белок необходим для экспорта мРНК из ядра. Другие две субъединицы DUBm *Drosophila* также обладают функциями, отличными от функции DUBm. Так, выявлено, что Sus1/ENY2 присутствует в нескольких различных комплексах. Эксперименты по иммуноокрашиванию политенных хромосом личинок *Drosophila* показали, что Sgf11 присутствует на всех сайтах локализации генов, кодирующих мяРНК, и что так же, как hSNAPC1, dPbp45, субъединица PBP комплекса, играющая ключевую роль в транскрипции мяРНК, присутствует не только в сайтах генов мяРНК, но и в других сайтах активной транскрипции, осуществляемой РНК-полимеразой II (PolII). Мы провели колокализацию на политенных хромосомах белков Sgf11 и Brf1 (компонента комплекса РНК-полимеразы III) и обнаружили, что Sgf11 находится во многих сайтах активной транскрипции и присутствует в тех же сайтах, что Brf1. Таким образом, мы показали, что SAGA коактивирует транскрипцию как РНК-полимеразы II-зависимых, так и РНК-полимеразы III-зависимых генов малых ядерных РНК.
Ключевые слова: транскрипция; SAGA; DUB модуль; AMEX; Sgf11; ENY2; гены мяРНК; PolII-зависимая транскрипция; PolIII-зависимая транскрипция.

Introduction

A large number of coactivator complexes are organized into transcription systems to provide accurate and precise functioning of the RNA polymerase II (RNAP II) machine. Coactivators are multisubunits complexes that are recruited to chromatin to promote transcription initiation by direct interaction with general transcription factors (GTFs) or RNAPII. They also may act indirectly through modification of chromatin structure (Li et al., 2007).

The yeast SAGA histone acetyltransferase (HAT) complex and its *Drosophila* and human homologs STAGA (the SPT3/TAF9/GCN5 acetyltransferase complex)/TFTC (the TBP-free

TAF complex) are multisubunits complexes that facilitate access of GTFs to DNA through histone acetylation mediated by the catalytic activity of the GCN5 subunit (Martinez, 2002). Human homologs of most SAGA subunits have been identified in TFTC/STAGA; biochemical and functional characterization demonstrated that the described human complexes are almost identical (hereafter called TFTC/STAGA) (Wu et al., 2004; Nagy, Tora, 2007).

The new components of the yeast SAGA complex were identified in biochemical studies bringing in additional functions to the complex (Powell et al., 2004; Rodriguez-Navarro et al., 2004). Ubp8 is a ubiquitin-specific protease, which has a histone H2B ubiquitin protease activity in this complex and is incorporated into SAGA through interaction with Sgf11 (Lee et al., 2005). H2B ubiquitination plays an important role in gene silencing and in activation of specific genes. Both ubiquitination and deubiquitination of H2B are required for optimal gene activation (Henry et al., 2003; Daniel et al., 2004). In higher eukaryotes, histone H2A is also monoubiquitinated (H2Aub1). It is interesting that Sus1, which is the component of both SAGA and the Sac3-Thp1 mRNA export complex in yeast (Rodriguez-Navarro et al., 2004), interacts with Sgf11 and Ubp8 and regulates the deubiquitination activity of the complex (Kohler et al., 2006).

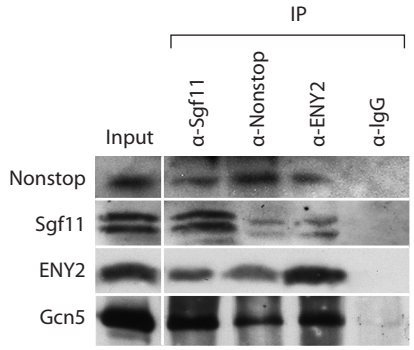


Fig. 1. Sgf11 interacts with DUBm subunits and the Gcn5 component of SAGA in nuclear extract from *Drosophila* embryos in co-immunoprecipitation experiments (IP).

Antibodies against Sgf11, Nonstop, ENY2 and Gcn5 or preimmune serum (PI) were used. Bands indicated with an asterisk correspond to antibodies (adapted from (Gurskiy et al., 2012)).

TFTC/STAGA complex has a modular organization that links the deubiquitination function to HAT and DUBm

We analyzed a highly purified human TFTC fraction by MS-MS mass spectrometry and identified three additional subunits of TFTC/STAGA (USP22, ATXN7L3 and ENY2). Human USP22 is the homolog of yeast Ubp8; human ATXN7L3 is the homolog of the yeast Sgf11 protein; and ENY2, a 101 amino acid protein, is homologous to yeast Sus1. We showed that the ubiquitin protease USP22 together with ATXN7L3 and ENY2 forms a deubiquitination module (DUBm). Also, we revealed that two different TFTC/STAGA subunits, TAF5L and ATXN7, interact with this module and may mediate its association with TFTC/STAGA (Zhao et al., 2008).

Then, we studied the DUBm of SAGA in *Drosophila melanogaster*. *Drosophila* homologs of yeast DUBm components have been identified. Nonstop, Sgf11 and ENY2 are homologous to yeast Ubp8, Sgf11 and Sus1, respectively, and were shown to be components of SAGA (Weake et al., 2008). *Drosophila* Nonstop and Sgf11 have a role in H2B deubiquitination (Weake et al., 2008). A putative *Drosophila* ortholog of yeast Sgf73 was also identified (Weake et al., 2009). However, the existence of an integrated DUBm in *Drosophila* has not been shown. We demonstrated that endogenous *Drosophila* ENY2, Sgf11 and Nonstop form an integrated DUBm associated with SAGA.

First, we raised antibodies against Sgf11 and Nonstop (subunits of DUBm) in rabbits, antibodies against the ENY2 component were described previously (Georgieva et al., 2001) and performed immunoprecipitation experiments. Antibodies against Sgf11 co-precipitated Nonstop from the nuclear extract of *Drosophila* embryos. Moreover, antibodies against either Sgf11 or Nonstop co-immunoprecipitated not only ENY2, but also the Gcn5 subunit of SAGA (Fig. 1). Thus, we confirmed the existence of a SAGA-associated DUBm in *Drosophila melanogaster*. It should be noted that a certain amount of Sgf11 remained in the extract after treatment with antibodies against Nonstop or ENY2. Hence Sgf11 also may perform their function in a complex other than SAGA DUBm.

Our previous data showed that the SAGA complex participates in the transcription of the *Drosophila* heat-shock protein 70 (*hsp70*) gene during heat shock (Lebedeva et al., 2005). So, we analyzed the function of Sgf11 in gene expression using the *hsp70* gene model. The antibodies against Sgf11 strongly stained *hsp70* puffs on *Drosophila* larval polytene chromosomes after heat shock, indicating that Sgf11 participates in the transcription of *hsp70*. Then we used a chromatin immunoprecipitation (ChIP) assay to study the occupancy of the *hsp70* promoter by Sgf11 and other DUBm subunits before and after gene activation. Sgf11 was detected on the *hsp70* promoter, its association with the promoter turned out to be RNA-dependent unlike that of ENY2 or Nonstop.

Earlier we demonstrated an interaction between E(y)2/ENY2 and the nuclear pore complex (NPC) and showed that SAGA/TFTC also contacts the NPC at the

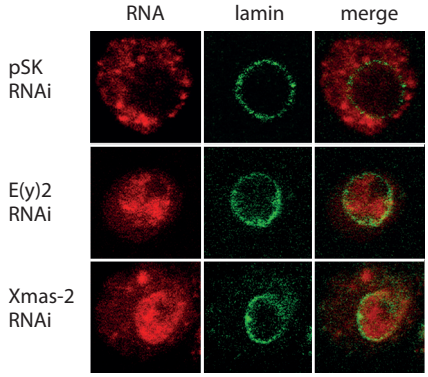


Fig. 2. E(y)2 and Xmas-2 are required for mRNA export from the nucleus. E(y)2 and Xmas-2 are required for poly(A)+ RNAexport from the nucleus.

RNAi was performed using either dsRNA corresponding to a fragment of pSK II vector as a control or the E(y)2 or the Xmas-2 cDNAs. RNA FISH was carried out using a Cy3-labelled oligo(dT) probe to identify poly(A)+RNA. Nuclear envelope is stained with lamin. Representative examples of cells are shown (magnification, ×1000) (adapted from (Kurshakova et al., 2007)).

nuclear periphery. E(y)2/ENY2 also forms a complex with the X-linked male sterile 2 (Xmas-2) protein to regulate mRNA export from nucleus to cytoplasm both in normal conditions and after heat shock. This complex was named AMEX/TREX2 (Fig. 2). Importantly, E(y)2/ENY2 and Xmas-2 knockdown decreased the contact between the *hsp70* gene loci and the nuclear envelope before and after activation and interfered with transcription.

Based on these data, we performed RNA-IP experiments and found that

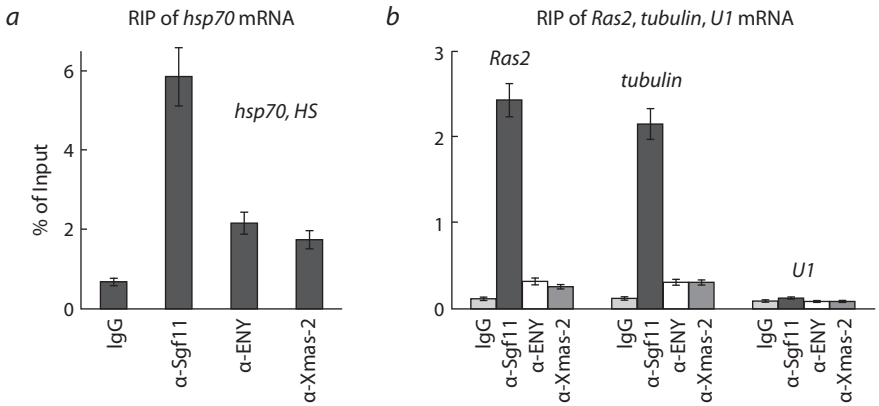


Fig. 3. Sgf11 is associated with mRNAs of several genes.

a – RIP experiments with *hsp70* mRNA after heat shock were performed using antibodies against Sgf11 or components of the mRNA-interacting AMEX complex (ENY2, Xmas-2); nonimmune IgG was used as control. The results are shown as a percentage of input; *b* – Sgf11 binds to mRNAs of *ras* and *tubulin* genes under normal conditions. The U1 snRNA was used as a control. Antibodies used in RIP experiments were the same as in Fig. 3, *a*. The results are shown as a percentage of input (adapted from (Gurskiy et al., 2012)).

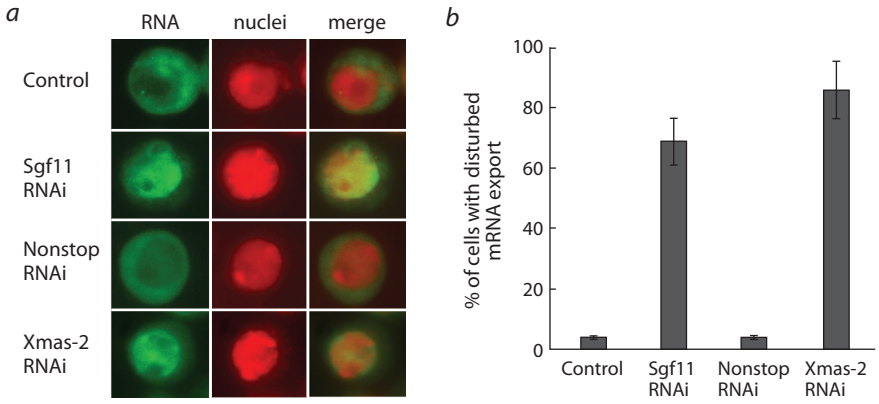


Fig. 4. RNAi knockdown of Sgf11 interferes with general mRNA export.

a – RNAi knockdown of Sgf11, but not Nonstop, interferes with general mRNA export. Cells were treated with GFP dsRNA (control) or dsRNA corresponding to Sgf11 and Nonstop. Xmas-2 RNAi knockdown was performed as a positive control. Representative examples of the distribution of mRNA (green staining) and cell nuclei (red staining) and the corresponding merged images are shown for control cells and cells after Sgf11 or Nonstop knockdown (magnification, ×1000). RNA FISH was carried out using a Cy3-labeled oligo(dT) probe to identify poly(A)+ RNA. The nuclei were stained blue with DAPI. The images were recolored in Photoshop for better visualization.

b – quantitative presentation of the results of experiments shown in Fig. 4, *a*. Bars show the percentage of cells with disturbed *hsp70* mRNA nuclear export (about 200 cells per RNAi experiment were examined) (adapted from (Gurskiy et al., 2012)).

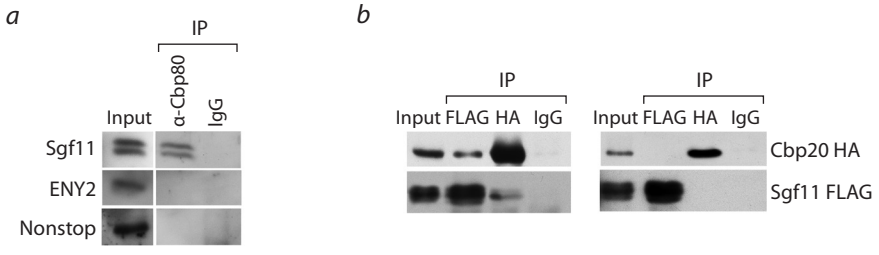


Fig. 5. Sgf11 is associated with Cbp80 in a separate complex.

a – co-immunoprecipitation experiments with nuclear extract of *Drosophila* embryos reveal no interactions between Cbp80 and ENY2 or Nonstop DUBm components.

b – recombinant Sgf11 interact with Cbp80 but not with Cbp20. FLAG-tagged Sgf11 was co-expressed with HA-tagged Cbp80 or HA-tagged Cbp20 in transiently transfected S2 cells. Immunoprecipitation was performed with anti-FLAG or anti-HA antibodies or with IgG. The Western blot was stained with anti-FLAG of anti-HA antibodies. About 10 % of the input and 50 % of the precipitate were loaded onto the gel (adapted from (Gurskiy et al., 2012)).

Sgf11 was associated with *hsp70* mRNA and with mRNAs of two other genes (*Ras2* and *tubulin*). Antibodies against Sgf11 immunoprecipitated mRNA even more efficiently than did antibodies against Xmas-2 or ENY2 (Fig. 3). We have also found that Sgf11 interacts with the AMEX/TREX2 complex and similarly to Xmas-2 and ENY2 co-localizes with NPC. We demonstrated that knockdown of Sgf11 disrupted mRNA export of both *hsp70* mRNA and the total mRNA (Fig. 4). Therefore, Sgf11 interaction with AMEX/TREX-2 may play an important role in general mRNA export (Gurskiy et al., 2012).

Sgf11 is present in several complexes in the *Drosophila* embryo nuclear extract and is associated with Cbp80 independently of the DUBm

Then, we purified Sgf11-containing complexes from the embryonic nuclear extract and revealed that Sgf11 was associated with the Cbp80 subunit of the Cap-Binding Complex (CBC). This result was confirmed in co-immunoprecipitation experiments (Fig. 5). Moreover, our data on co-expression of recombinant Sgf11 and Cbp80 in *Drosophila* S2 cell culture demonstrated that Sgf11 directly interacts with Cbp80, and Cbp80 was necessary for Sgf11 recruitment; this interaction was independent of the other subunit of the CBC complex, Cbp20. Thus, Sgf11 functions independently of the DUBm in mRNA export. In fact, the other two subunits of the DUBm were also shown to have functions separate from the DUBm function. Thus, Sus1/ENY2 was shown to be present in several different complexes (Kurshakova et al., 2007; Kopytova et al., 2010). Human ataxin 7, the homolog of the yeast Sgf73 subunit of the DUBm, was found in the cytoplasm and is involved in the regulation of cytoskeleton dynamics (Nakamura et al., 2012). Taken together, our data provide evidence that although *Drosophila* Sgf11 is an integral component of the SAGA DUBm, it also forms a complex with Cbp80 and associates with nascent mRNA. It interacts with the AMEX/TREX-2 mRNA export complex and is involved in competent mRNP translocation to the cytoplasm.

SAGA is present at snRNA genes and interacts with Pbp45

Mapping the binding sites for Sgf11 component of the SAGA complex on the polytene chromosomes of *Drosophila* larvae by immunostaining revealed that Sgf11 is present at the sites of localization of snRNA genes (Gurskiy et al., 2012). To verify this result, we performed double immunostaining of polytene chromosomes from the salivary glands of *Drosophila* using antibodies against Sgf11 and Pbp45, the subunit of the PBP complex, the key player in the snRNA transcription process. It was found that Pbp45 can be found not only at the snRNA loci but also at other actively transcribed sites (in puffs and interbands). Similar data had previously been obtained by the whole genome sequencing of hSNAPC1, the human Pbp45 homolog (Baillat et al., 2012). Our findings are the first indication that similar to hSNAPC1, dPbp45 is associated in addition to the snRNA gene localization sites, with other sites of active transcription by PolII.

Sgf11 and Pbp45 colocalized at many actively transcribed genes on polytene chromosomes, including those sites where *U1–U6* snRNA genes are located. To confirm that the SAGA complex is indeed present at the promoters of the snRNA genes, we performed ChIP with antibodies against the components of SAGA (Sgf11, ENY2 and Gcn5) and Pbp45. All tested proteins were detected at the promoter regions of snRNA genes: at promoters of the genes transcribed by RNA polymerase II (snRNA U1 and snRNA U2) and at promoters of the U6 snRNA genes which are transcribed by RNA polymerase III.

To determine whether the subunits in the SAGA complex physically interact with the snRNA gene transcription apparatus, we co-immunoprecipitated the components of SAGA modules: the HAT module (Gcn5 and Ada2b factors), the DUB module (Sgf11, ENY2, and Nonstop) and TRRAP, with the antibodies against the Pbp45 protein from the nuclear extract of the *Drosophila* S2 cells. Results obtained showed that protein components of both the HAT and DUB modules interact with Pbp45, and thus, with the PBP complex. Therefore, SAGA proteins are not only present at the promoter regions of snRNA genes, but they also interact with the snRNA transcription machinery. Moreover, SAGA participates in snRNA gene transcription guided by both PolII and PolIII.

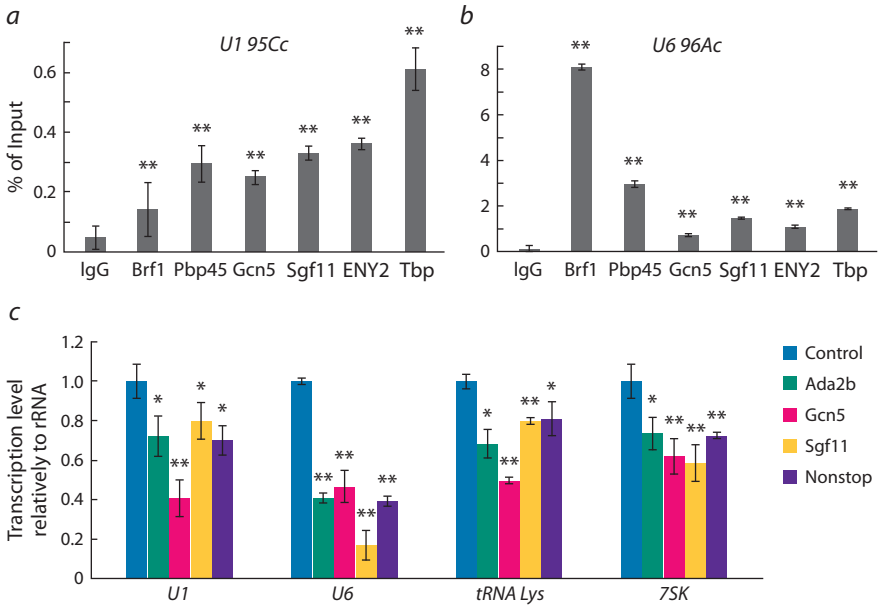


Fig 6. The participation of the SAGA complex in snRNA transcription.

(*a* and *b*) – occupation of the U1 and U6 promoters (*U1 95Cc*, *U6 96Ac*) by Brf1, Pbp45, Gcn5, Sgf11, ENY2, and TBP in *Drosophila* pupat. The protein level was measured by ChIP. The results of ChIP are provided as a percentage of input.

c – effects of decrease of Ada2b, Gcn5, Sgf11, and Nonstop transcription levels in mutant strains on the level of transcription of *U1*, *U6*, *tRNA Lys* and *7SK* genes. *TM6B*, *Tb[1]/+* strain was used as a control. The transcription levels were normalized on *28S rRNA*. *t* - tests have been performed to compare the means ($^{*}p < 0.05$; $^{**}p < 0.01$) (adapted from (Popova et al., 2018)).

To detect SAGA at the RNA polymerase III-transcribed genes, we analyzed colocalization of Sgf11 and Brf1 (the component of the RNA polymerase III basal transcription complex) on polytene chromosomes. In contrast to Sgf11, which is present at many active transcription sites in interbands and puffs, Brf1 was found in a relatively low number of sites of the PolIII-transcribed genes. However, the immunostaining experiments revealed that Sgf11 was present at all Brf1 sites. In particular, they colocalize at the loci corresponding to *U6* and tRNA genes, as well as other RNA polymerase III target genes.

Recently, a ChIP-seq analysis of *Drosophila* embryos has shown that Ada2b, Nonstop, and Sgf11 occupied promoter regions of snRNA genes (Li et al., 2017). Our results obtained for *Drosophila* support these data and we have demonstrated that SAGA participates in the transcription of snRNAs. We have also shown that the SAGA complex participates in the transcription of other PolIII target genes: the *U6* snRNA genes.

Chromatin immunoprecipitation using antibodies against the components of the SAGA complex and Brf1 confirmed that SAGA is present not only at the promoters of individual U6 snRNA genes, but also at the promoters of other genes (*RNase MRP* and *tRNA Lys*) transcribed by RNA polymerase III (Fig. 6, *a–c*). Additionally, we have demonstrated that the protein components of both the HAT module and the DUB module interact with Brf1. All the tested SAGA subunits interacted with Brf1 in the nuclear extract. Therefore, the SAGA complex is present at the promoters of the RNA polymerase III-transcribed genes and interacts with the RNA polymerase III transcription factors.

The effect of mutations in the genes encoding SAGA complex subunits on snRNA transcription

Next, we have checked whether mutations of subunits of the SAGA complex could change snRNA gene transcription in flies. The effect of mutations in genes encoding SAGA components on snRNA transcription in *Drosophila* showed that it is indeed the case. First, we showed that all studied factors were present on the promoters of snRNA genes in flies (chromatin was prepared from *Droso-*

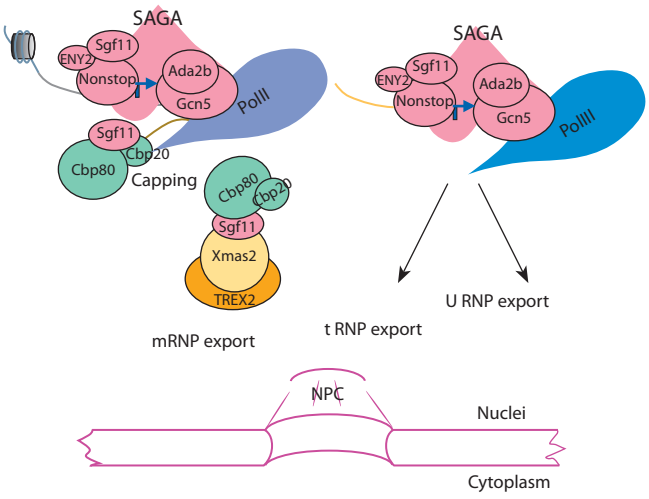


Fig. 7. SAGA proteins participate in the expression of both RNA PolII- and RNA PolIII-dependent genes.

The SAGA complex coactivates transcription from the promoters regulated by both RNA PolII and RNA PolIII. Following transcription activation of RNA PolII-dependent genes, Sgf11 subunit of SAGA interacts with Cbp80 protein of the CBC complex and associates with growing mRNA. Sgf11 associates with mRNP, interacts with TREX-2/AMEX and is essential for mRNA export through the NPC.

phila pupae). This is in line with our results obtained for snRNA genes in the S2 cells. So, SAGA also takes part in snRNA gene regulation at the level of the whole organism.

For our genetic crosses we used mutant strains which carried mutations in genes encoding subunits of the DUB module (Sgf11 and Nonstop) and the HAT module (Gcn5 and Ada2b). Each mutation was lethal in homozygotes because there was no test protein in the mutant line. Thus, we performed our experiments on heterozygous strains. In all strains, the level of investigated gene expression was decreased. The *TM6*;*TB*^{1}/+ heterozygotes were used as a control.

Using qRT-PCR we compared the snRNA (*U1* and *U6*) transcription level in the mutant flies with that in the control strain. We also measured the transcription levels of two other PolIII dependent genes (*tRNA Lys* and *7SK*). Our results showed that the mutations in all tested SAGA subunit genes caused significant decrease in the *U1* and *U6 snRNA* transcription levels and had a weaker effect on *U1* snRNA transcription. Thus, the results confirmed that SAGA is indeed the coactivator of the transcription of snRNAs, as well as of PolIII-dependent genes. Decreased transcription levels were also observed for the *tRNA Lys* and *7SK* genes.

It is important to indicate that mutations in SAGA subunits affected the attachment of Brf1 to promoters of *U6* genes. The high level of Brf1 detected on *U6* promoters was significantly decreased in the fly strains with mutations. Thus, the SAGA complex may also be participating in the recruitment of PolIII transcription factors.

We addressed the question as to whether the level of H2B monoubiquitylation could also be influenced by SAGA mutations. For our investigation, we choose H2B monoubiquitylation of *U6* 96Ac, since, as had been shown previously, this gene was transcribed more actively than the other studied genes and we could better estimate the effect of mutation on H2B monoubiquitylation. The antibodies against nonmodified H2B and against monoubiquitylated H2B (H2BK120Ub) were used. The data represent the ratio of H2BK120Ub to nonmodified H2B, verified by ChIP in the control and mutant strains. The results obtained demonstrated that H2B monoubiquitylation increased in all mutated flies comparatively to the control.

Conclusion

Recently, X. Li et al. (2017) investigating SAGA-independent properties of the DUB module, detected by a ChIP-seq assay an occupancy of snRNA promoter regions by Sgf11, Spt3, and Ada2b subunits. Our results are in agreement with the data that the DUB module has both SAGA-related and independent functions. In summary, it can be argued that the SAGA complex in metazoans is widely involved in the regulation of gene transcription (Popova et al., 2018).

Our model suggests that the SAGA complex is involved both in PolII and PolIII transcription, while some subunits of the complex participate in subsequent transcription steps (Fig. 7).

References

Baillat D., Gardini A., Cesaroni M., Shiekhattar R. Requirement for SNAPC1 in transcriptional responsiveness to diverse extracellular signals. Mol. Cell. Biol. 2012;32(22):4642-4650. DOI 10.1128/mcb.00906-12.

Daniel J.A., Torok M.S., Sun Z.W., Schieltz D., Allis C.D., Yates J.R., Grant P.A. Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. J. Biol. Chem. 2004;279(3):1867-1871. DOI 10.1074/jbc.C300494200.

Georgieva S., Nabirochkina E., Dilworth F.J., Eickhoff H., Becker P., Tora L., Georgiev P., Soldatov A. The novel transcription factor e(y)2 interacts with TAF(II)40 and potentiates transcription activation on chromatin templates. Mol. Cell. Biol. 2001;21(15):5223-5231. DOI 10.1128/mcb.21.15.5223-5231.2001.

Gurskiy D., Orlova A., Vorobyeva N., Nabirochkina E., Krasnov A., Shidlovskii Y., Georgieva S., Kopytova D. The DUBm subunit Sgf11 is required for mRNA export and interacts with Cbp80 in *Drosophila*. Nucleic Acids Res. 2012;40(21):10689-10700. DOI 10.1093/nar/gks857.

Henry K.W., Wyce A., Lo W.S., Duggan L.J., Emre N.C., Kao C.F., Pillus L., Shilatifard A., Osley M.A., Berger S.L. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev. 2003;17(21):2648-2663. DOI 10.1101/gad.1144003.

Kohler A., Pascual-Garcia P., Llopis A., Zapater M., Posas F., Hurt E., Rodriguez-Navarro S. The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. Mol. Biol. Cell. 2006;17(10):4228-4236. DOI 10.1091/mbc.e06-02-0098.

Kopytova D.V., Orlova A.V., Krasnov A.N., Gurskiy D.Y., Nikolenko J.V., Nabirochkina E.N., Shidlovskii Y.V., Georgieva S.G. Multifunctional factor ENY2 is associated with the THO complex and promotes its recruitment onto nascent mRNA. Genes Dev. 2010;24(1):86-96. DOI 10.1101/gad.550010.

Kurshakova M.M., Krasnov A.N., Kopytova D.V., Shidlovskii Y.V., Nikolenko J.V., Nabirochkina E.N., Spehner D., Schultz P., Tora L., Georgieva S.G. SAGA and a novel *Drosophila* export complex anchor efficient transcription and mRNA export to NPC. EMBO J. 2007;26(24):4956-4965. DOI 10.1038/sj.emboj.7601901.

Lebedeva L.A., Nabirochkina E.N., Kurshakova M.M., Robert F., Krasnov A.N., Evgen'ev M.B., Kadonaga J.T., Georgieva S.G., Tora L. Occupancy of the *Drosophila* hsp70 promoter by a subset

of basal transcription factors diminishes upon transcriptional activation. Proc. Natl. Acad. Sci. USA. 2005;102(50):18087-18092. DOI 10.1073/pnas.0509063102.

Lee K.K., Florens L., Swanson S.K., Washburn M.P., Workman J.L. The deubiquitylation activity of Ubp8 is dependent upon Sgf11 and its association with the SAGA complex. Mol. Cell. Biol. 2005; 25(3):1173-1182. DOI 10.1128/mcb.25.3.1173-1182.2005.

Li B., Carey M., Workman J.L. The role of chromatin during transcription. Cell. 2007;128(4):707-719. DOI 10.1016/j.cell.2007.01.015.

Li X., Seidel C.W., Szerszen L.T., Lange J.J., Workman J.L., Abmayr S.M. Enzymatic modules of the SAGA chromatin-modifying complex play distinct roles in *Drosophila* gene expression and development. Genes Dev. 2017;31(15):1588-1600. DOI 10.1101/gad.300988.117.

Martinez E. Multi-protein complexes in eukaryotic gene transcription. Plant Mol. Biol. 2002;50(6):925-947.

Nagy Z., Tora L. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene. 2007;26(37):5341-5357. DOI 10.1038/sj.onc.1210604.

Nakamura Y., Tagawa K., Oka T., Sasabe T., Ito H., Shiwaku H., La Spada A.R., Okazawa H. Ataxin-7 associates with microtubules and stabilizes the cytoskeletal network. Hum. Mol. Genet. 2012; 21(5):1099-1110. DOI 10.1093/hmg/ddr539.

Popova V.V., Orlova A.V., Kurshakova M.M., Nikolenko J.V., Nabirochkina E.N., Georgieva S.G., Kopytova D.V. The role of SAGA coactivator complex in snRNA transcription. Cell Cycle. 2018; 17(15):1859-1870. DOI 10.1080/15384101.2018.1489175.

Powell D.W., Weaver C.M., Jennings J.L., McAfee K.J., He Y., Weil P.A., Link A.J. Cluster analysis of mass spectrometry data reveals a novel component of SAGA. Mol. Cell. Biol. 2004;24(16):7249-7259. DOI 10.1128/mcb.24.16.7249-7259.2004.

Rodriguez-Navarro S., Fischer T., Luo M.J., Antunez O., Brettschneider S., Lechner J., Perez-Ortin J.E., Reed R., Hurt E. Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. Cell. 2004;116(1): 75-86.

Weake V.M., Lee K.K., Guelman S., Lin C.H., Seidel C., Abmayr S.M., Workman J.L. SAGA-mediated H2B deubiquitination controls the development of neuronal connectivity in the *Drosophila* visual system. EMBO J. 2008;27(2):394-405. DOI 10.1038/sj.emboj.7601966.

Weake V.M., Swanson S.K., Mushegian A., Florens L., Washburn M.P., Abmayr S.M., Workman J.L. A novel histone fold domain-containing protein that replaces TAF6 in *Drosophila* SAGA is required for SAGA-dependent gene expression. Genes Dev. 2009;23(24):2818-2823. DOI 10.1101/gad.1846409.

Wu P.Y., Ruhlmann C., Winston F., Schultz P. Molecular architecture of the *S. cerevisiae* SAGA complex. Mol. Cell. 2004;15(2):199-208. DOI 10.1016/j.molcel.2004.06.005.

Zhao Y., Lang G., Ito S., Bonnet J., Metzger E., Sawatsubashi S., Suzuki E., Le Guezennec X., Stunnenberg H.G., Krasnov A., Georgieva S.G., Schule R., Takeyama K., Kato S., Tora L., Devys D. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Mol. Cell. 2008;29(1):92-101. DOI 10.1016/j.molcel.2007.12.011.

ORCID ID

E.N. Nabirochkina orcid.org/0000-0002-7942-3924
M.M. Kurshakova orcid.org/0000-0001-6772-6955
S.G. Georgieva orcid.org/0000-0003-3665-0390
D.V. Kopytova orcid.org/0000-0003-1086-6329

Acknowledgements. This study was supported by the grant RFBR No. 18-04-00514 (Kopytova D.V., Nabirochkina E.N.).

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised December 13, 2018. Accepted December 13, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):180-183
DOI 10.18699/VJ19.479

Identification of proteins that can participate in the recruitment of Ttk69 to genomic sites of *Drosophila melanogaster*

I.S. Osadchiy, T.N. Fedorova, P.G. Georgiev, O.G. Maksimenko ✉

Institute of Gene Biology, RAS, Moscow, Russia
✉ e-mail: maksog@mail.ru

The proteins with the BTB domain play an important role in the processes of activation and repression of transcription. Interestingly, BTB-containing proteins are widely distributed only among higher eukaryotes. Many BTB-containing proteins are transcriptional factors involved in a wide range of developmental processes. One of the key regulators of early development is the BTB-containing protein Ttk (tramtrack), which is able to interact with the *Drosophila* nucleosome remodeling and histone deacetylation (dNuRD) complex. Ttk69 directly interacts with two protein components of the dNuRD complex, dMi-2 and MEP1. It can be assumed that Ttk69 represses some target genes by remodeling chromatin structure through the recruitment of the dNuRD complex. However, it is still unknown what provides for specific recruitment of Ttk to chromatin in the process of negative/positive regulation of a target gene expression. Although Ttk69 has DNA-binding activity, no extended specific motif has been identified. The purpose of this study was to find proteins that can participate in the recruitment of Ttk to regulatory elements. To identify Ttk partner proteins, screening in the yeast two-hybrid system was performed against a collection of proteins with clusters of C2H2 domains, which bind effectively and specifically to sites on chromatin. As a results, the CG10321 and CG1792 proteins were identified as potential DNA-binding partners of Ttk. We suppose that the CG10321 and CG1792 proteins provide specificity for the recruitment of Ttk and, as a result, of the NuRD-complex to the genome regulatory elements. We found that the Ttk protein is able to interact with the MEP1 and ZnF proteins at once.
Key words: *Drosophila*; protein-protein interaction; gene regulation; transcription factor; tramtrack.

For citation: Osadchiy I.S., Fedorova T.N., Georgiev P.G., Maksimenko O.G. Identification of proteins that can participate in the recruitment of Ttk69 to genomic sites of *Drosophila melanogaster*. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):180-183. DOI 10.18699/VJ19.479

Идентификация белков, участвующих в привлечении белка Ttk69 к геномным сайтам у *Drosophila melanogaster*

И.С. Осадчий, Т.Н. Федорова, П.Г. Георгиев, О.Г. Максименко ✉

Институт биологии гена Российской академии наук, Москва, Россия
✉ e-mail: maksog@mail.ru

Белки, имеющие в своем составе BTB-домены, играют важную роль в процессах активации и репрессии транскрипции. Белки, содержащие BTB-домены, широко распространены только среди высших эукариот. Многие из таких белков являются транскрипционными факторами, принимающими участие в процессах развития организма. Один из ключевых регуляторов процессов раннего развития – BTB-содержащий белок Ttk (tramtrack), способный взаимодействовать с комплексом ремоделирования нуклеосом и деацетилирования гистонов (dNuRD) дрозофилы. Белок Ttk69 способен напрямую взаимодействовать с двумя белками комплекса dNuRD complex, dMi-2 и MEP1. Можно предположить, что Ttk69 репрессирует мишеневые гены путем ремоделирования хроматина через привлечение комплекса dNuRD. Однако до сих пор неизвестно, что обеспечивает специфичность рекрутирования Ttk на хроматин в процессе негативной/позитивной регуляции генной экспрессии. Несмотря на то что Ttk69 обладает ДНК-связывающей активностью, протяженных специфичных мотивов связывания для него найдено не было. Целью данного исследования был поиск белков, которые могут участвовать в привлечении Ttk к геномным регуляторным элементам. Для поиска белков-партнеров Ttk был проведен скрининг в дрожжевой двугибридной системе против коллекции белков с кластерами доменов «цинковые пальцы» C2H2-типа, способных эффективно и специфично связываться с сайтами в хроматине. В результате два белка, CG10321 и CG1792, были описаны в качестве потенциальных ДНК-связывающих партнеров Ttk. Мы предполагаем, что CG10321 и CG1792 обеспечивают специфичность посадки Ttk на сайты и, как следствие, привлечение NuRD-комплекса к геномным регуляторным элементам. Мы обнаружили также, что белок Ttk может одновременно взаимодействовать с MEP1 и ZnF-белками.
Ключевые слова: дрозофила; белок-белковое взаимодействие; генная регуляция; транскрипционный фактор; tramtrack.

Introduction

The genome of higher eukaryotes is a complex, but at the same time flexible and tightly regulated system. The spatial and temporal specificity of gene activation or repression is realized in different ways. It is possible that proteins with the BTB domain play an important role in the processes of activation and repression of transcription. Interestingly, BTB-containing proteins are widely distributed only among higher

eukaryotes. Thus, 85 BTB proteins are known in *Drosophila melanogaster*, and 183 BTB proteins in *Homo sapiens*. The BTB/POZ (broad complex, Tramtrack, bric-a-brac/Pox virus and zinc finger) is an evolutionarily conserved protein-protein interaction domain. Most of the BTB domains are capable of forming multimers. Many BTB-containing proteins are transcriptional factors involved in a wide range of developmental processes. A special place among them is occupied by transcription factors that bind DNA using "zinc fingers" (BTB-ZF). Fifteen such proteins were found in *Drosophila*, and 43, in humans. All known BTB-ZF proteins are involved in the regulation of transcription, in particular in the action of enhancers, promoters and insulators (Stogios et al., 2005; Perez-Torrado et al., 2006). Typically, the BTB domain is located at the N-terminus of a protein, while closer to the C-terminus are zinc fingers of C2H2-type (ZF), which are involved in binding to DNA and interaction with other proteins. These proteins are often involved in the regulation of transcription and the organization of enhancer-promoter interactions (e. g., PLZF, Bcl6, GAF, Mod (mdg4), Ttk). The significance of the BTB domain in gene regulation has not been revealed. However, the BTB domain is involved in interactions with a number of proteins. In the case of Bcl6 and PLZF, the BTB domain directly interacts with the corepressors BcoR, N-CoR (SMRT), mSin3A (Melnick et al., 2002) to form a transcriptional repressor complex that includes the HDAC-1 histone deacetylase. In addition to interacting with corepressors, some BTB domains (GAF, bric-a-brac), in contrast, are associated with the TAFII155 transcription activator (Pointud et al., 2001; Chopra et al., 2008). The BTB domain of the Kaiso transcription factor is able to interact with the C-terminal domain of the human CTCF protein, which weakens the insulator activity of the latter (Defossez et al., 2005).

One of the key regulators of early development is the BTB-containing protein Ttk, which occurs in two isoforms, Ttk69 and Ttk88. These isoforms share the same BTB-domain but possess alternative sets of Zn fingers. Ttk69 is a repressor of *tailless* expression (Chen et al., 2002) and is a master repressor of enteroendocrine cell specification of *Drosophila* intestinal stem cell lineages (Wang et al., 2015). Also, Ttk69 regulates *Drosophila* oogenesis through selective and negative regulation of gene expression in follicle and germline cells (Boyle, Berg, 2009; Peters et al., 2013) and participates in photoreceptor development (Lai, Li, 1999). The BTB domain of Ttk69 genetically interacts with the transcriptional corepressor dCtBP (*Drosophila* homolog of human C-terminal-binding protein), which might be an interacting partner of Ttk69 for the control of cell fate decision and cellular differentiation (Wen et al., 2000). In addition, it was demonstrated that the Pits protein is able to interact with the BTB domain of Ttk69 and the Sin3A corepressor (Liaw, 2016). Moreover, Ttk69 directly interacts with two protein components of the *Drosophila* nucleosome remodeling and histone deacetylation (dNuRD) complex, dMi-2 (Murawsky et al., 2001) and MEP1 (Reddy et al., 2010). It can be assumed that Ttk69 represses some target genes by remodeling chromatin structure through the recruitment of the dNuRD complex.

Although Ttk has DNA-binding activity, no extended specific motif has been identified (Chen et al., 2002; Kulakovskiy, Makeev, 2009). Moreover, it was shown that Ttk69-mediated repression of GAGA-depended promoters occurs in the absence of Ttk69 binding to DNA (Pagans et al., 2004). These results, together with the detected interactions of Ttk69 with dCtBP and dMi2, suggest that depending on the promoter and the developmental stage, Ttk69 might repress transcription through different mechanisms (Pagans et al., 2004). However, it is still unknown what provides for specific recruitment of Ttk to chromatin in the process of negative/positive regulation of target gene expression.

The purpose of this study was to find proteins that can participate in the recruitment of Ttk69 to regulatory elements. To identify Ttk69 partner proteins, screening in the yeast two-hybrid system was performed against a collection of proteins with clusters of C2H2 domains, which bind effectively and specifically to sites on chromatin.

Materials and methods

Fusion proteins were cloned into vectors pGBT9 and pGAD424 from Clontech and verified by sequencing. Ttk69 (and deletion derivatives) were expressed in fusion with the Gal4 DNA binding domain (DBD) along with ZnF proteins (and deletion derivatives) fused to the Gal4 activation domain (AD) in a pairwise manner. The yeast two-hybrid assay was carried out using the yeast strain pJ69-4A (MATa trp1-901 leu2-3,112 ura3-52 his3-200 gal4Δ gal80Δ GAL2-ADE2 LYS2::GAL1-HIS3 met2::GAL7-lacZ), with plasmids and protocols from Clontech. For growth assays, plasmids were transformed into yeast strain pJ69-4A by the lithium acetate method, as described by the manufacturer with some modifications. All cells were grown at 30 °C in an orbital shaker at 250–300 rpm. Yeast colonies were transferred into a 15 mL culture tube with 6–7 mL of YPDA medium and grown for one day. The culture was 10-fold diluted in a 0.5 L culture flask with YPDA medium and cultivated for 3 h. Aliquots of 1.5 mL of cell suspension were pelleted by centrifugation at 4000 g for10 s, and the supernatant was removed. Pelleted cells were resuspended in 1 mL of 0.1 M LiAcO, incubated for 30 min and pelleted as above. To the pellet, 240 μL 50 % (v/v) PEG 3380, 36 μL of 1 M LiAcO, and 50 μL of a mixture of two plasmids (the amount of each plasmid in a mixture of 400–800 ng) was added sequentially and the cells were suspended to homogeneity. The tube was incubated at 30 °C for 30 min, then at 42 °C for 5 min and then placed on ice for 1–2 min. Cells were pelleted at 4000 g for 15–20 s and resuspended in 100 μL of sterile ddH2O. The resuspended cells were plated on selective medium lacking Leu and Trp ("medium-2"). The plates were incubated at 30 °C for 2–3 days. Afterwards, the colonies were streaked out on plates on selective medium lacking either Leu, Trp and His ("medium-3"), or lacking adenine in addition to the three amino acids ("medium-4"), or lacking the three amino acids but containing 5 mM 3-amino-1,2,4-triazole ("medium-3 + 5 mM 3AT"). The plates were incubated at 30 °C for 3–4 days and growth was assessed. Each assay was prepared as three independent biological replicates with three technical repeats.

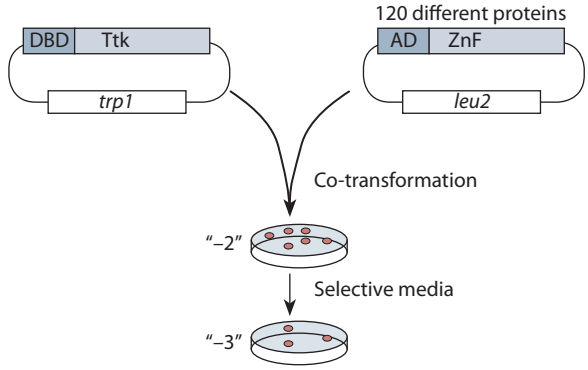
Results

To identify new DNA-binding factors that might directly interact with Ttk, we performed a yeast two-hybrid assay against a collection of ZnF proteins. Ttk was expressed in

fusion with a Gal4 DNA binding domain (DBD) along with ZnF proteins fused to the Gal4 activation domain (AD). Being co-expressed in yeast, the two associating proteins reconstitute the function of the transcription factor GAL4 that activates the *his3* gene in the yeast strain pJ69-4A, which is an auxotroph for histidine (Fig. 1, *a*). Yeasts with interacting DBD and AD fusion proteins can grow on plates lacking histidine, leucine and tryptophan. The assay revealed interaction of Ttk69 with CG10321 and CG1792. We also confirmed the previously described interaction of the Ttk69 protein with MEP1 (see Fig. 1, *b*).

Presumably, the CG10321 and CG1792 proteins provide specificity for the DNA-binding of Ttk69 and, as a result, the recruiting of the NuRD-complex to the genome regulatory elements. Therefore, the Ttk69 protein must be capable of a simultaneous interaction with the MEP1 and ZnF proteins. To test this assumption, we decided to localize the interacting domains of these proteins. Using appropriate deletion constructs, we found that the zinc fingers of Ttk69 are sufficient to interact with CG10321 and CG1792 (Fig. 2, *a*). While earlier, the Ttk region upstream of the zinc fingers was identified as interacting with MEP1 (Reddy et al., 2010). These results demonstrate that the Ttk69 protein is able to interact simultaneously with the NuRD complex and DNA-binding ZnF proteins. In addition, we localized interacting regions of CG10321 (see Fig. 2, *b*) and CG1792 (see Fig. 2, *c*) at the middle parts of the proteins, i. e., between the N-terminal ZAD domain and zinc fingers. So the interactions of these proteins with Ttk69 do not interfere with the DNA-binding activity of CG10321 and CG1792.

a



b

	AD	CG10321	CG1792	MEP1
DBD	–	–	–	–
DBD-Ttk	–	+	+	+

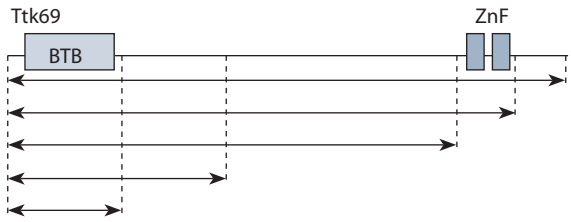
Fig. 1. Ttk is able to direct interaction with ZnF proteins.

a – principle of the yeast two-hybrid assay (Y2H); *b* – positive results of Y2H screening. The results are summarized in columns, with the '+' and '–' signs referring to the presence and absence of interaction, respectively.

Discussion

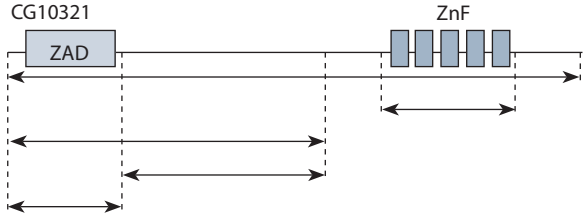
The structure of chromatin is related with the processes of activation or repression of gene transcription. Chromatin remodeling complexes are multi-component systems that can modulate the position of nucleosomes and insert or remove certain histone marks. One of the main complexes of this type in eukaryotes is Mi-2/NuRD. Studies have shown that Mi-2/NuRD complexes not only move nucleosomes through ATP-

a



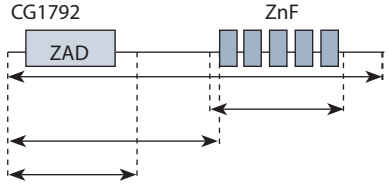
	CG10321	CG1792	AD
[1-643]	+	+	–
[1-570]	+	+	–
[1-500]	–	–	–
[1-265]	–	–	–
[1-120]	–	–	–

b



	Ttk	DBD
[1-855]	+	–
[605-709]	–	–
[1-504]	+	–
[122-504]	+	–
[1-122]	–	–

c



	Ttk	DBD
[1-372]	+	–
[180-340]	–	–
[1-196]	+	–
[1-107]	–	–

Fig. 2. Localization of interacting regions in Ttk, CG10321 and CG1792.

a – localization of the Ttk region interacting with CG10321 and CG1792 in the yeast two-hybrid assay. Full-sized or fragments of Ttk were fused to the GAL4 DNA-binding domain (DBD) and tested for interaction with CG10321 and CG1792 fused to the GAL4 activating domain (AD). Ttk fusions were tested for the absence of interaction with the GAL4 activating domain alone; b, c – localization of CG10321 and CG1792 regions interacting with CLAMP in the yeast two-hybrid assay. Different fragments of these proteins were fused to the GAL4 activating domain and tested for interaction with Ttk fused to the GAL4 DNA-binding domain.

dependent chromatin remodeling, but also contain subunits that deacetylate histones and bind methylated DNA.

In *Drosophila*, one of the variants of the Mi-2 complex, dMec, was purified. A unique component of this complex is the conserved protein MEP1, which is responsible for the binding profile of the entire complex at chromatin sites. It was shown that this type of the complex is critically necessary for somatic differentiation at the stage of transition from early to late embryogenesis. Presumably, the MEP1 protein is the structure-forming component of this complex. Previously it was shown that, in particular, it binds to the transcription repressor protein Ttk69. It is possible that this repressing activity is associated with the recruitment of the NuRD complex to genomic targets. However, no specific binding motifs are known for either MEP1 or Ttk69. According to our hypothesis, MEP1 and Ttk69 form many contacts with DNA-binding proteins which ensure the specific positioning of the NuRD complex on chromatin. As part of testing such a hypothesis, we searched for protein partners of Ttk69 by screening among DNA-binding ZnF proteins. As a result, we found the CG10321 and CG1792 proteins, which can act as potential recruiters of the Ttk-containing complex to chromatin. We localized interaction regions of these proteins and showed that the unstructured regions of the CG10321 and CG1792 proteins interact with Ttk69's zinc fingers, while zinc fingers of CG10321 and CG1792 remain free for DNA binding. At the same time, MEP1 binds to the area adjacent to the zinc fingers in Ttk69. This result indicates the possibility of simultaneous interaction of the Ttk69 protein with MEP1 and DNA-binding ZnF proteins.

Conclusions

At the present time, several dozen complexes participating in epigenetic regulation of gene expression are described. Nevetheless, mechanisms of their specific recruitment to target loci are still unclear. One of the current models implies that a specific ensemble of DNA-binding proteins on a regulatory element may serve as a platform for the targeted complex assembly. Here, we report on two new earlier undescribed proteins with DNA-binding domains, which potentially can recruit the NuRD complex to its genomic loci. Further research on the aspects of its interactions with the NuRD complex can contrtibute to our better understanding of epigenetic complexes recruitment.

References

Boyle M.J., Berg C.A. Control in time and space: Tramtrack69 cooperates with Notch and Ecdysone to repress ectopic fate and shape changes during *Drosophila* egg chamber maturation. Development. 2009;136(24):4187-4197. DOI 10.1242/dev.042770.

Chen Y.-J., Chiang C.-S., Weng L.-C., Lengyel J.A., Liaw G.-J. Tramtrack69 is required for the early repression of tailless expression. Mech. Dev. 2002;116:75-83.

Chopra V.S., Srinivasan A., Kumar R.P., Mishra K., Basquin D., Docquier M., Seum C., Pauli D., Mishra R.K. Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. Dev. Biol. 2008;317(2):660-670.

Defossez P.A., Kelly K.F., Filion G.J., Pérez-Torrado R., Magdinier F., Menoni H., Nordgaard C.L., Daniel J.M., Gilson E. The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. J. Biol. Chem. 2005;280(52):43017-43023.

Kulakovskiy I.V., Makeev V.J. Discovery of DNA motifs recognized by transcription factors through integration of different experimental sources. Biophysics. 2009;54:667-674.

Lai Z.C., Li Y. Tramtrack69 is positively and autonomously required for *Drosophila* photoreceptor development. Genetics. 1999;152(1):299-305.

Liaw G.J. Pits, a protein interacting with Ttk69 and Sin3A, has links to histone deacetylation. Sci. Rep. 2016;6:33388. DOI 10.1038/srep33388.

Melnick A., Carlile G., Ahmad K.F., Kiang C.L., Corcoran C., Bardwell V., Prive G.G., Licht J.D. Critical residues within the BTB domain of PLZF and Bcl-6 modulate interaction with corepressors. Mol. Cell. Biol. 2002;22(6):1804-1818.

Murawsky C.M., Brehm A., Badenhorst P., Lowe N., Becker P.B., Travers A.A. Tramtrack69 interacts with the dMi-2 subunit of the *Drosophila* NuRD chromatin remodelling complex. EMBO Rep. 2001;2(12):1089-1094.

Pagans S., Piñeyro D., Kosoy A., Bernués J., Azorín F. Repression by TTK69 of GAGA-mediated activation occurs in the absence of TTK69 binding to DNA and solely requires the contribution of the POZ/BTB domain of TTK69. J. Biol. Chem. 2004;279(11):9725-9732.

Perez-Torrado R., Yamada D., Defossez P.A. Born to bind: the BTB protein-protein interaction domain. BioEssays. 2006;28(12):1194-1202.

Peters N.C., Thayer N.H., Kerr S.A., Tompa M., Berg C.A. Following the 'tracks': Tramtrack69 regulates epithelial tube expansion in the *Drosophila* ovary through Paxillin, Dynamin, and the homeobox protein Mirror. Dev. Biol. 2013;378(2):154-169. DOI 10.1016/j.ydbio.2013.03.017.

Pointud J.C., Larsson J., Dastugue B., Couderc J.L. The BTB/POZ domain of the regulatory proteins Bric à brac 1 (BAB1) and Bric à brac 2 (BAB2) interacts with the novel *Drosophila* TAF(II) factor BIP2/dTAF(II)155. Dev. Biol. 2001;237(2):368-380.

Reddy B.A., Bajpe P.K., Bassett A., Moshkin Y.M., Kozhevnikova E., Bezstarosti K., Demmers J.A., Travers A.A., Verrijzer C.P. *Drosophila* transcription factor Tramtrack69 binds MEP1 to recruit the chromatin remodeler NuRD. Mol. Cell. Biol. 2010;30(21):5234-5244. DOI 10.1128/MCB.00266-10.

Stogios P.J., Downs G.S., Jauhal J.J., Nandra S.K., Privé G.G. Sequence and structural analysis of BTB domain proteins. Genome Biol. 2005;6(10):R82.

Wang C., Guo X., Dou K., Chen H., Xi R. Ttk69 acts as a master repressor of enteroendocrine cell specification in *Drosophila* intestinal stem cell lineages. Development. 2015;142(19):3321-3331. DOI 10.1242/dev.123208.

Wen Y., Nguyen D., Li Y., Lai Z.C. The N-terminal BTB/POZ domain and C-terminal sequences are essential for Tramtrack69 to specify cell fate in the developing *Drosophila* eye. Genetics. 2000;156(1):195-203.

Acknowledgements. This study was supported by the Russian Foundation for Basic Research (project No. 18-04-01272). The work was performed using the equipment of the IGB RAS facilities supported by the Ministry of Science and Education of the Russian Federation.

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised December 11, 2018. Accepted December 11, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):184-189
DOI 10.18699/VJ19.480

Different functions of PHF10 isoforms – subunits of the PBAF chromatin remodeling complex

A.A. Sheynov[1], V.V. Tatarskiy[1], A.M. Azieva[1, 2], S.G. Georgieva[1], N.V. Soshnikova[1]

[1] Institute of Gene Biology, RAS, Department of Eukariotic Transcription Factors, Moscow, Russia
[2] National Research Center "Kurchatov Institute", Moscow, Russia
e-mail: so2615nat@gmail.com

Chromatin remodelling multiprotein complexes play an important role in regulation of gene expression in embryogenesis and in the adult organism. Mutations in the subunits of the complexes are often lethal or lead to developmental defects. Complexes consist of core subunits and a specific module. The core consists of ATPase and structure subunits, specific subunits of the module are necessary for chromatin binding. PHF10 (PHD finger protein 10) is a subunit of the PBAF (polybromo-associated BAF) chromatin remodelling complex subfamily. Conserved and highly regulated PHF10 is ubiquitously expressed in mammals as four different isoforms. The isoforms of PHF10 differ by domain structures and posttranslational modifications. All isoforms are highly regulated and included in the PBAF complex in a mutually exclusive manner. Two of the PHF10 isoforms (PHF10-P) are expressed at a high level in neuronal and myeloid progenitors and are necessary for cell proliferation. These isoforms contain PHD (plant homeodomain) fingers for nucleosome binding and recruit RNA polymerase II on the promoters of cell cycle genes. Two other isoforms (PHF10-S) instead of PHD have PDSM (phosphorylation-dependent sumoylation motif), the motif for SUMO1 conjugation. PHF10 is the most unstable subunit of the PBAF complex. Stability can alter the turnover rate of the subunits of the PBAF complex. All PHF10 isoforms are degraded by β-TrCP ubiquitin ligase but PHF10-S isoforms contain a cluster of serins (X-cluster) for multiple phosphorylation by casein kinase I. This phosphorylation protects the β-TrCP degron from β-TrCP recognition and subsequently stabilizes the PHF10-S isoforms. Thus, the incorporation of PHF10 isoforms with different phosphorylation patterns and different stability into the PBAF complexes alters the functions of the entire PBAF complex and determines the range of genes undergoing remodelling.
Key words: chromatin remodeling; PBAF complex; PHF10 isoforms.

For citation: Sheynov A.A., Tatarskiy V.V., Azieva A.M., Georgieva S.G., Soshnikova N.V. Different functions of PHF10 isoforms – subunits of the PBAF chromatin remodeling complex. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):184-189. DOI 10.18699/VJ19.480

Функции изоформ PHF10 – субъединицы PBAF комплекса, ремоделирующего хроматин

А.А. Шейнов[1], В.В. Татарский[1], А.М. Азиева[1, 2], С.Г. Георгиева[1], Н.В. Сошникова[1]

[1] Институт биологии гена Российской академии наук, Москва, Россия
[2] Национальный исследовательский центр «Курчатовский институт», Москва, Россия
e-mail: so2615nat@gmail.com

Комплексы, ремоделирующие хроматин, играют важную роль в экспрессии генов при эмбриональном развитии и во взрослом организме. Мутации субъединиц этого комплекса часто летальны или приводят к дефектам развития. Один из основных комплексов эукариот, изменяющих структуру хроматина, – комплекс PBAF, входящий в семейство SWI/SNF комплексов. Комплекс PBAF состоит из коровых субъединиц (Brg1, BAF155/BAF170, BAF47 и др.) и субъединиц специфического модуля (PHF10, BAF200, BAF180 и BRD7). Коровые субъединицы – это структурные субъединицы и АТФаза, специфические субъединицы – субъединицы, необходимые для связывания хроматина. Субъединичный состав комплекса не является постоянным. В процессе развития и дифференцировки клеток организма субъединицы комплекса заменяются гомологичными, что обуславливает специфичность работы комплекса на различных генах. Белок PHF10 – субъединица модуля PBAF комплекса, он играет важную роль в регуляции генов млекопитающих. В клетках и тканях человека и мыши PHF10 представлен четырьмя изоформами. Изоформы PHF10 имеют разную доменную структуру N- и С-концов, что определяет их свойства – различную клеточную локализацию, стабильность и модификационные паттерны. Две изоформы PHF10 (PHF10-P) экспрессируются на высоком уровне в нейрональных и миелоидных предшественниках и необходимы для пролиферации клеток. Эти изоформы содержат домены типа «PHD-пальцев», необходимые для связывания нуклеосом, и рекрутируют РНК-полимеразу II на промоторы генов клеточного цикла. Две другие изоформы (PHF10-S) вместо PHD доменов на С-конце имеют мотив PDSM для конъюгации SUMO1. Белок PHF10 представляет собой наиболее нестабильную субъединицу комплекса PBAF. Стабильность изоформ может регулировать скорость замены субъединиц в PBAF комплексе. Все PHF10 изоформы деградируют посредством убиквитинирования,

осуществляемого B-TrCP убиквитин-лигазой, и дальнейшего расщепления 26-S протеасомой. Изоформы PHF10 содержат кластер серинов (X-кластер), подвергающийся интенсивному фосфорилированию казеин-киназой. Это фосфорилирование защищает B-TrCP дегрон от узнавания B-TrCP убиквитин-лигазой и последующей деградации, что приводит к большей стабильности PHF10-S форм по сравнению с PHF10-P формами. Таким образом, включение в PBAF изоформ PHF, обладающих различными паттернами фосфорилирования и различной стабильностью, влияет на функции целого PBAF комплекса и определяет спектр ремоделируемых генов.
Ключевые слова: ремоделирование хроматина; PBAF комплекс; изоформы PHF10.

Introduction

One of the most widely represented families of chromatin-remodelling complexes is the family of ATP-dependent complexes SWI/SNF (SWItch/Sucrose Non-Fermentable). These complexes use the energy of ATP hydrolysis to locally disrupt or alter the association of histones with DNA (Vignali et al., 2000) and play crucial role in chromatin assembly. This family is highly conserved in eukaryotes. At the same time, in the process of evolution, it acquires an increasing complexity in the structure, following the demand for increasing complexity of chromatin organization and chromatin regulation in multicellular organisms. For example, in yeast (*Saccharomyces cerevisiae*) chromatin does not contain methylated DNA or linker histones and the complexes perform the function of a transcriptional activator only. In the transition to multicellular organisms, the complexes acquire new components and a greater variation in composition and additional functions. Higher eukaryotes exhibit an increased genome size and a more complex genomic organization.

The complexes act as transcriptional activators and repressors and the large combinatorial diversity, which is associated with the incorporation of homologous subunits and different isoforms in the complexes, provides assemblies that are tissue-specific or specific to a certain developmental stage (Ho, Crabtree, 2010). In mammals, SWI/SNF is represented by two subfamilies of complexes – BAF (BRG-/BRM-associated factor) and PBAF that differ by subunit composition (Tang et al., 2010). Each complex contains subunits that form the core part and several specific subunits that form the module relevant to a particular type of complex. The core part of SWI/SNF complex consists of ATPase which is encoded by homologous *Brg1* or *Brm* genes, and several structural subunits – BAF155/BAF170, BAF47, BAF60, actin. Specific subunits (BAF200, BAF180, BRD7, PHF10 of PBAF and BAF250a/b of BAF) are responsible for interaction with chromatin. Almost all subunits of the complex can be replaced by paralogs. In particular, this leads to greater variability of subunit composition and, subsequently, to a wider range of functions of this complex. This also determines the pattern of the expression of target genes and specifies the regulation of nucleosome remodelling. Post-translational modifications of SWI/SNF complex subunits impact on the regulation of its functions even more (Simone, 2006).

The chromatin-remodelling complex of the PBAF family plays an essential role in the regulation of gene expression during embryogenesis. Mutations in the subunits of this complex are often lethal or lead to developmental defects (Hodges et al., 2016). Also, many mutations were characterized as secondary mutations during oncogenesis, leading to the advanced course of disease (Masliah-Planchon et al., 2015). One of the most modificated and regulated subunits is the protein PHF10, characterized by our group (Brechalov et al., 2014; Tatarskiy et al., 2017).

Structural domains of PHF10 are evolutionary conserved

The two main structural domains of PHF10, SAY (supporter of activation of yellow) and PHD, are evolutionarily conserved and are represented in all multicellular organisms. The length of the N- and C- terminal unstructured regions may vary (Fig. 1) (Vorobyeva et al., 2009).

The most thoroughly studied homologue of PHF10 in *Drosophila* is the protein SAYP (supporter of activation of yellow protein). The homozygous SAYP knockout in *Drosophila*,

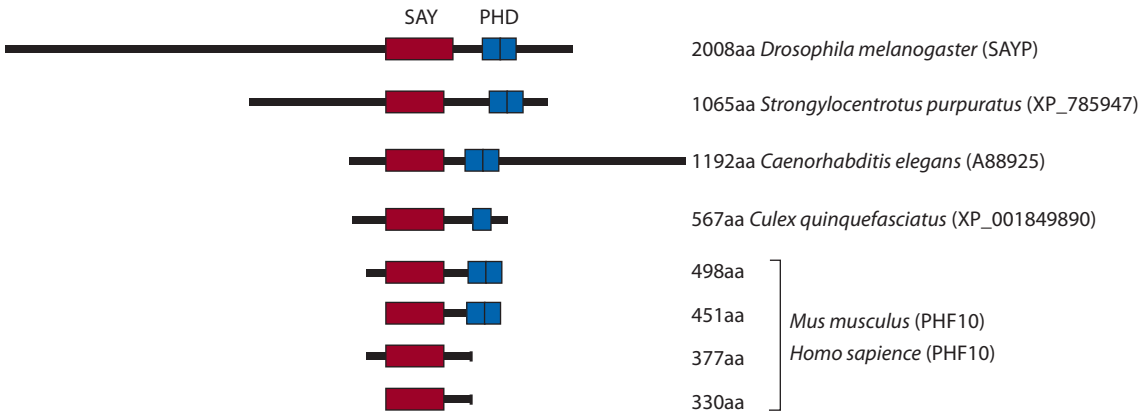


Fig. 1. PHF10 isoforms and their homologues in different species.

The total lengths of the proteins are indicated *Drosophila melanogaster*, SAYP; *Strongylocentrotus purpuratus*, XP_785947 locus; *Caenorhabditis elegans*, A88925 locus; *Culex quinquefasciatus*, XP_001849890 locus; *Mus musculus*, PHF10; *Homo sapiens*, PHF10. The total lengths of the proteins are indicated. SAY domain is depicted as a red box. PHD finger domain is depicted as a blue box. Adopted from (Vorobyeva et al., 2009).

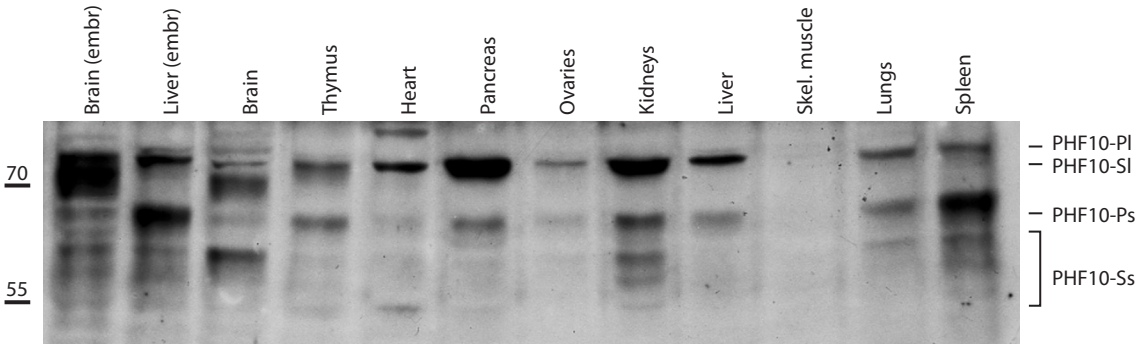


Fig. 2. Equal amount of protein (40 μg) was loaded on each lane.

Each lane indicates different PHF10 isoforms contained in various tissues. Detection of PHF10 in extracts from corresponding tissue by western blotting with affinity purified anti-PHF10 antibodies. Set of protein bands in a molecular weight range of about 60–80 kDa corresponds to different isoforms. The presence of multiple (more than four) bands and the difference between the predicted and the real molecular weight is due to the presence of post-translational modifications.

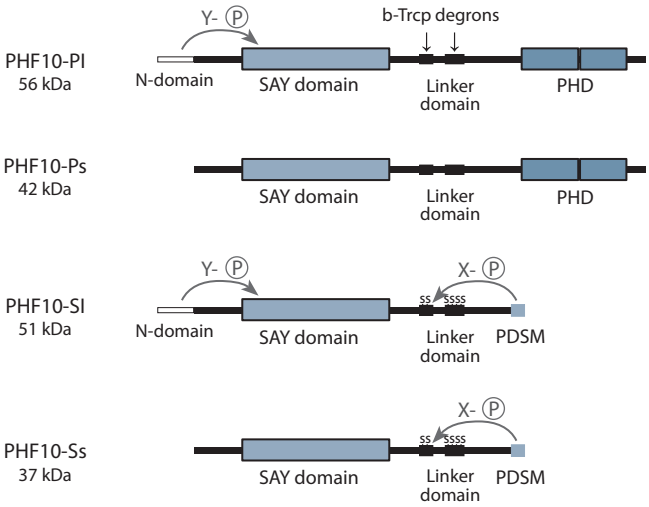


Fig. 3. Domain structure and phosphorylation patterns of PHF10 isoforms.

PHF10-P isoform contains two PHD domains at the C-terminus. PHF10-S isoform contains PDSM. Also both isoforms differ in N-terminal parts by the presence or absence of 46 amino acids: "l" – long and "s" – short. X- and Y-phosphorylation clusters are depicted. β-TrCP degrons depicted as thick black lines. Adopted from (Tatarskiy et al., 2017).

as well as the homozygous PHF10 knockout in a mouse is embryonic lethal (Krasteva et al., 2017). That suggests an important role of this protein in the expression of developmental genes. In the adult *Drosophila*, SAYP is mainly expressed in female ovaries (Vorobyeva et al., 2009), while PHF10 is expressed in almost all tissues of the mammalian organism (Brechalov et al., 2016), (Tissue expression of PHF10…, 2018) (Fig. 2).

Unlike *Drosophila*, where SAYP is represented by a polypeptides of 2006–2012 amino acids that have similar domain structure (FlyBase n.d., 2019), four isoforms of PHF10 that have a different domain structure are expressed in mammals. These isoforms are evolutionarily highly conserved in different mammalian species (Brechalov et al., 2014), which supports the existence of important functions of each isoform (Fig. 3).

Domain organisation of PHF10 isoforms

All four isoforms of PHF10 are the products of the same gene, and are formed as a result of alternative splicing of transcripts and differ in N- and C-terminal domains (see Fig. 3). The first isoform that was described contains two PHD domains at the C-terminus and was designated by us as the isoform PHF10-P (containing the PHD domain). Another isoform lacks a PHD domain, instead of it six C-terminal amino acids form PDSM (phosphorylation-dependent sumoylation motif) that regulates phosphorylation-dependent sumoylation. We designated this isoform as PHF10-S (see Fig. 3). Also, PHF10-P and PHF10-S isoforms may differ in N-terminal parts by the presence or absence of 46 amino acids: "l" – long and "s" – short (see Fig. 3). Molecular weight is 56, 42, 51, 37 kDa for isoforms PHF PL/Ps/SL/Ss respectively.

All PHF10 isoforms contain an evolutionary conserved SAY domain, which is responsible for the incorporation of protein in the PBAF complex (see Fig. 3). In mammals, the SAY domain seems to be uniquely represented only in the structure of the PHF10 isoforms, although in a number of works some homology of the SAY domain with the N-terminal domain of BAF47 and Kruppel-like BAF45b/c/d subunits of BAF complexes was found (Lessard et al., 2007; Allen et al., 2015).

The PHF10-P and PHF10-S isoforms are included in the PBAF complex in a mutually exclusive manner. In other words, there are PBAF complexes that contain PHF10-P isoforms, and PBAF complexes that contain PHF10-S isoforms (Brechalov et al., 2014). The incorporation of different isoforms into the complex does not lead to a change in the subunit composition of the complex (Brechalov et al., 2014). However, different isoforms may interact with different activators or transcriptional coactivator complexes, leading to modulation of mechanisms of functioning of the PBAF complexes containing different PHF10 isoforms.

The PHF10-P isoforms contain two PHD zinc finger domains. It has been shown that these domains can bind the N-terminal "tails" of histones, especially trimethylated histone H3 lysine 4 (Hyun et al., 2017). The specific ligand has not yet been found for PHD domains of the PHF10-P isoforms, but it is likely that the incorporation of PHF10-P isoform in

the complex may determine the binding of the PBAF complex to specifically modified nucleosomes.

PHF10-S isoforms do not contain PHD domains that recognise nucleosomes. However, they can covalently attach a small ubiquitin-related modifier protein, SUMO1, to the C-terminus (Fig. 4) (Brechalov et al., 2014). The role of the sumoylation of PHF10-S isoforms is not yet clear, but it is known that protein sumoylation can affect their turnover, localisation and determine other proteins interacting with them (Seeler, Dejean, 2017). Thus, the incorporation of various isoforms in the PBAF complex can significantly change the functions of the entire complex, influence the range of protein binding partners, and position the PBAF complex on chromatin. Therefore, this determines the patterns of genes that the PBAF complex remodels (Brechalov et al., 2014).

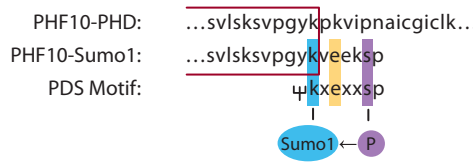


Fig. 4. The consensus sequence for phosphorylation-dependent SUMO modification (PDSM) at the C terminus of PHF10-Sl/Ss isoforms and scheme of amino acids sumoylation.

Red box indicates the common part of PHF10-P and PHF10-S isoforms. Adopted from (Brechalov et al., 2014).

Functions of PHF10 isoforms

It has been shown that PBAF complexes that contain PHF10 isoforms with PHD domains have a positive effect on the transcription of genes that are involved in cell proliferation, comparing to the PHF10-S isoform. The percentage of proliferation genes that changed their expression relative to the total number of differentially expressed genes during overexpression of PHF10-Pl and PHF10-Sl isoforms in HEK293 cells compared to control, non-transfected cells, was 9.23 % and 4.28 % respectively. These changes in gene expression led to an increase in the number of cell divisions per unit of time (from 3.4 to 4.3 in 24 hours and from 8.8 to 11.4 in 48 hours) in cells overexpressing the PHF10-Pl isoform, compared to control. The number of divisions in the cells expressing the PHF10-Sl isoform was comparable to the control (Brechalov et al., 2014).

Using chromatin immunoprecipitation, we showed that the PHF10-Pl and the PHF10-Sl isoforms are present in the PBAF complexes on the promoters of genes involved in cell proliferation *ZMIZ1* and *NOV*. However, only PHF10-P overexpression resulted in an increased level of RNA polymerase II on the promoter, which correlated with an increase in level of *ZMIZ1* and *NOV* mRNA (Brechalov et al., 2014). That is consistent with a model in which PBAF complexes containing the PHF10-Pl isoform, regulate gene expression differently than PBAF complexes containing the PHF10-Sl isoform.

In the study of mouse brain development, J. Lessard (2007) showed that PHF10-Pl is highly expressed during the process of embryonic brain development, with increased expression in proliferating neurons. During the transition from proliferating progenitor cells to postmitotic neurons, expression of the PHF10-Pl isoforms is stopped along with the increase in expression of short PDSM-containing isoforms PHF10-Ss (Azieva et al., 2018).

PBAF complexes that contain the PHF10-Pl isoform were localised on the promoters of the SHH-Olig and Notch signalling genes, responsible for the development of the mouse brain (Lessard et al., 2007).

Also, the important role of PHF10 isoforms in the development of blood cells was shown on mouse models. In the hematopoietic stem cells, the PHF10-Pl isoforms are expressed at a high level. High expression of PHF10-Pl isoforms maintains a pool of stem cells. Conditional knockout of *Phf10* in the hematopoietic tissue of an adult mouse leads to depletion of the pool of blood stem cells and reduces the number of myeloid progenitors, mature macrophages, monocytes and granulocytes in bone tissue (Krasteva et al., 2017). We have also shown that during myeloid differentiation, the PHF10-Pl isoform in the PBAF complex is replaced by the PHF10-Ss isoform, which affects the patterns of actively transcribed genes.

Post-translational modifications of PHF10 isoforms

According to the phosphosite.org database (Phosphosite Knowledgebase n.d., 2018), the subunits of the PBAF complex are intensively phosphorylated. A particularly large number of sites (more than 20 per protein) are located in the BRG1, BAF155, BAF200, BAF180, BRD7, and PHF10 subunits. Phosphorylation by various kinases provides a complex regulation of the functions of the PBAF complex and sensitivity to various signalling pathways (Simone, 2006). Phosphorylation of subunits alters the charge of the complex and allows it to bind to various proteins, thus providing specificity in the reprogramming of gene expression (Cui et al., 2016).

Different isoforms of PHF10 have different phosphorylation patterns that are dependent on their domain structure (see Fig. 3) (Tatarskiy et al., 2017). The additional N-terminal domain of 47 amino acids that exists in the long PHF10-Pl/Sl isoforms contributes to the presence of the phosphorylation pattern, which we designated as Y (see Fig. 3). The absence of PHD domains inherent to the isoforms of PHF10-Sl/Ss contributes to the phosphorylation of a number of serines in the linker domain. We designated this cluster as X-phosphorylation (see Fig. 3). Also, PHF10-Sl and PHF10-Ss have a six amino acids PDSM motif at the C-terminus. We have determined that the lysine can covalently bind the SUMO1 modifying protein. Moreover, sumoylation occurs more intensively if the subsequent serine is phosphorylated in this motif (see Fig. 4) (Brechalov et al., 2014).

Another function of phosphorylation is the regulation of protein stability. Due to the different stability of the subunits, the additional regulation of the functions of the complex is provided by the replacement of a subunit via degradation of the unnecessary one (Sears et al., 2000; Liu et al., 2002; Hafumi et al., 2011).

PHF10 is one of the most phosphorylated and unstable subunits of the PBAF complex (Brechalov et al., 2014, 2016; Tatarskiy et al., 2017; Azieva et al., 2018). Our studies revealed that PHF10 isoforms contain the non-canonical β-TrCP ubiquitin ligase recognition motif and are degraded by β-TrCP

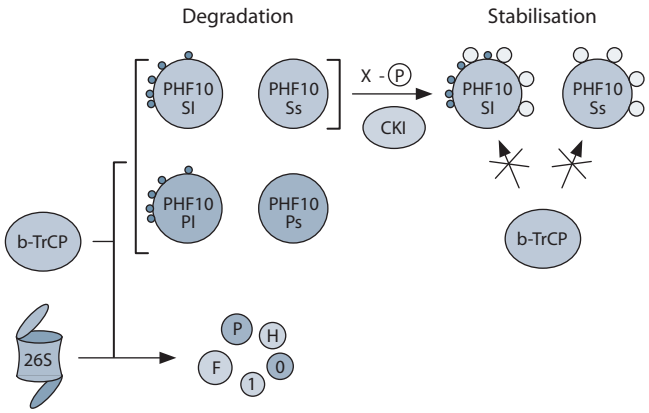


Fig. 5. Diagram of PHF10 degradation.

PHF10 isoforms contain the non-canonical β-TrCP ubiquitin ligase recognition motif and are degraded by β-TrCP in a phospho-dependent manner through 26S proteasome. Phosphorylation by CKI (casein kinase I) of serines in the X-cluster in the linker domain of the PHF10-Sl/Ss isoforms impairs interaction of isoforms with β-TrCP and opposes their degradation. Large light circles – X-cluster phosphorylation; small dark – Y-cluster phosphorylation. Adopted from (Tatarskiy et al., 2017).

in a phospho-dependent manner. The non-canonical recognition motif for ligase is localised in the linker domain and is represented in all isoforms (Tatarskiy et al., 2017).

The function of phosphorylation of serines organized in the X-cluster in the linker domain of the isoforms PHF10-Sl/Ss was thoroughly studied by our group (Tatarskiy et al., 2017). X-cluster serines are phosphorylated by casein kinase 1. X-cluster contains positively charged amino acids – β-TrCP degron. Normally β-TrCP recognises this degron and contributes to the degradation of PHF10 isoforms. The X-cluster phosphorylation block the positive charges and opposes the interaction β-TrCP ubiquitin ligase with its degron (Tatarskiy et al., 2017). Thus, phosphorylation of X-cluster serines leads to a significant stabilization of the PHF10-Sl and PHF10-Ss isoforms, compared to the stability of PHF10-Pl and PHF10-Ps isoforms (Fig. 5) (Tatarskiy et al., 2017).

Conclusions

Thus, the isoforms of PHF10, being part of the PBAF complex, are highly regulated through post-translational modifications providing a broad variety of their functions through the ability to accept messages of signalling pathways. Also, the incorporation into the PBAF complexes of PHF10 isoforms with different phosphorylation patterns and different stability alters the functions of the entire PBAF complex and determines the range of genes undergoing remodelling.

References

Allen M.D., Freund S.M.V., Zinzalla G., Bycroft M. the SWI/SNF subunit INI1 contains an N-terminal winged helix DNA binding domain that is a target for mutations in schwannomatosis. Structure. 2015;23(7):1344-1349. DOI 10.1016/j.str.2015.04.021.

Azieva A.M., Sheinov A.A., Galkin F.A., Georgieva S.G., Soshnikova N.V. Stability of chromatin remodeling complex subunits is determined by their phosphorylation status. Dokl. Biochem. Biophys. 2018;479(1):66-68. DOI 10.1134/S1607672918020035.

Brechalov A.V., Georgieva S.G., Soshnikova N.V. Mammalian cells contain two functionally distinct PBAF complexes incorporating different isoforms of PHF10 signature subunit. Cell Cycle. 2014;13(12):1970-1979. DOI 10.4161/cc.28922.

Brechalov A.V., Valieva M.E., Georgieva S.G., Soshnikova N.V. PHF10 isoforms are phosphorylated in the PBAF mammalian chromatin remodeling complex. Mol. Biol. 2016;50(2):278-283. DOI 10.1134/S0026893316010039.

Cui H., Schlesinger J., Schoenhals S., Tönjes M., Dunkel I., Meierhofer D., Cano E., Schulz K., Berger M.F., Haack T., Abdelilah-Seyfried S., Bulyk M. L., Sauer S., Sperling S.R. Phosphorylation of the chromatin remodeling factor DPF3a induces cardiac hypertrophy through releasing HEY repressors from DNA. Nucleic Acids Res. 2016;44(6):2538-2553. DOI 10.1093/nar/gkv1244.

FlyBase. n.d. "FlyBase Gene Report: Dmel\e(y)3." Accessed January 18, 2019. http://flybase.org/reports/FBgn0087008.

Hafumi N., Hashimoto K., Panchenko A.R. Phosphorylation in protein-protein binding: effect on stability and function. Structure. 2011; 19(12):1807-1815. DOI 10.1016/j.str.2011.09.021.

Ho L., Crabtree G.R. Chromatin remodelling during development. Nature. 2010;463(7280):474-484. DOI 10.1038/nature08911.

Hodges C., Kirkland J.G., Crabtree G.R. The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer. Cold Spring Harb. Perspect. Med. 2016;6(8):a026930. DOI 10.1101/cshperspect.a026930.

Hyun K., Jeon J., Park K., Kim J. Writing, erasing and reading histone lysine methylations. Exp. Mol. Med. 2017;49(4): e324. DOI 10.1038/emm.2017.11.

Krasteva V., Crabtree G.R., Lessard J.A. The BAF45a/PHF10 subunit of SWI/SNF-like chromatin remodeling complexes is essential for hematopoietic stem cell maintenance. Exp. Hematol. 2017;48:58-71.e15. DOI 10.1038/emm.2017.11.

Lessard J., Wu J.I., Ranish J.A.,Wan M.,Winslow M.M., Staahl B.T., Wu H., Aebersold R., Graef I.A., Crabtree G.R. An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron. 2007;55(2):201-215. DOI 10.1016/j.neuron.2007.06.019.

Liu C., Li Y., Semenov M., Han C., Baeg G.H., Tan Y., Zhang Z., Lin X., He X. Control of β-catenin phosphorylation/degradation by a dual-kinase mechanism. Cell. 2002;108(6):837-847. DOI 10.1016/s0092-8674(02)00685-2.

Masliah-Planchon J., Bièche I., Guinebretière J.M., Bourdeaut F., Delattre O. SWI/SNF chromatin remodeling and human malignancies. Annu. Rev. Pathol. 2015;10:145-171. DOI 10.1146/annurev-pathol-012414-040445.

Phosphosite Knowledgebase. n.d. Accessed October 22, 2018. http://phosphosite.org.

Sears R., Nuckolls F., Haura E., Taya Y., Tamaid K., Nevins J.R. Multiple Ras-dependent phosphorylation pathways regulate myc protein stability. Genes Dev. 2000;14(19):2501-2514. https://www.ncbi.nlm.nih.gov/pubmed/11018017.

Seeler J.S., Dejean A. SUMO and the robustness of cancer. Nat. Rev. Cancer. 2017;17(3):184-197. DOI 10.1038/nrc.2016.143.

Simone C. SWI/SNF: the crossroads where extracellular signaling pathways meet chromatin. J. Cell. Physiol. 2006;207(2):309-314. DOI 10.1002/jcp.20514.

Tang L., Nogales E., Ciferri C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. Prog. Biophys. Mol. Biol. 2010;102(2-3):122-128.

Tatarskiy V.V., Simonov Y.P., Shcherbinin D.S., Brechalov A.V., Georgieva S.G., Soshnikova N.V. Stability of the PHF10 subunit of PBAF signature module is regulated by phosphorylation: role of β-TrCP. Sci. Rep. 2017;7(1):5645. DOI 10.1038/s41598-017-05944-3.

Tissue expression of PHF10 – Summary – The Human Protein Atlas. n.d. Accessed October 22,2018.

Vignali M.A., Hassan H., Neely K.E., Workman J.L. ATP-dependent chromatin-remodeling complexes. Mol. Cell. Biol. 2000;20(6):1899-1910. https://www.ncbi.nlm.nih.gov/pubmed/10688638.

Vorobyeva N.E., Soshnikova N.V., Kuzmina J.L., Kopantseva M.R., Nikolenko J.V., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. The novel regulator of metazoan development SAYP organizes a nuclear coactivator supercomplex. Cell Cycle. 2009;8(14):2152-2156. DOI 10.4161/cc.8.14.9115.

https://www.proteinatlas.org/ENSG00000130024-PHF10/tissue.

ORCID ID

A.A. Sheynov orcid.org/0000-0002-3302-2797
V.V. Tatarskiy orcid.org/0000-0002-9080-5683
A.M. Azieva orcid.org/0000-0003-2222-6302
S.G. Georgieva orcid.org/0000-0003-3665-0390
N.V. Soshnikova orcid.org/0000-0001-9260-3068

Acknowledgements. This study was supported by the Russian Science Foundation (grant No. 18-14-00303 to S.N.). The experiments were carried out using infrastructure of Center for collective use of the Institute of Gene Biology, Russian Academy of Science: "Biology of the living cell and drug biomedical nanotransporters".

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised January 18, 2019. Accepted January 24, 2019.

Вавиловский журнал генетики и селекции. 2019;23(2):190-198
DOI 10.18699/VJ19.481

Non3 is an essential *Drosophila* gene required for proper nucleolus assembly

E.N. Andreyeva[1] ✉, A.A. Ogienko[1, 2], A.A. Yushkova[1, 2], J.V. Popova[1, 3], G.A. Pavlova[1], E.N. Kozhevnikova[1, 3], A.V. Ivankin[1], M. Gatti[1, 4], A.V. Pindyurin[1, 2, 3] ✉

[1] Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
[4] IBPM CNR and Department of Biology and Biotechnology, Sapienza University of Rome, Rome, Italy
✉ e-mail: andreeva@mcb.nsc.ru; a.pindyurin@mcb.nsc.ru

The nucleolus is a dynamic non-membrane-bound nuclear organelle, which plays key roles not only in ribosome biogenesis but also in many other cellular processes. Consistent with its multiple functions, the nucleolus has been implicated in many human diseases, including cancer and degenerative pathologies of the nervous system and heart. Here, we report the characterization of the *Drosophila Non3* (*Novel nucleolar protein 3*) gene, which encodes a protein homologous to the human Brix domain-containing Rpf2 that has been shown to control ribosomal RNA (rRNA) processing. We used imprecise *P*-element excision to generate four new mutant alleles in the *Non3* gene. Complementation and phenotypic analyses showed that these *Non3* mutations can be arranged in an allelic series that includes both viable and lethal alleles. The strongest lethal allele ($Non3^{\Delta 600}$) is a genetically null allele that carries a large deletion of the gene and exhibits early lethality when homozygous. Flies heterozygous for $Non3^{\Delta 600}$ occasionally exhibit a mild reduction in the bristle size, but develop normally and are fertile. However, heteroallelic combinations of viable *Non3* mutations ($Non3^{197}$, $Non3^{310}$ and $Non3^{259}$) display a *Minute*-like phenotype, consisting in delayed development and short and thin bristles, suggesting that they are defective in ribosome biogenesis. We also demonstrate that the Non3 protein localizes to the nucleolus of larval brain cells and it is required for proper nucleolar localization of Fibrillarin, a protein important for post-translational modification and processing of rRNAs. In summary, we generated a number of genetic and biochemical tools that were exploited for an initial characterization of *Non3*, and will be instrumental for future functional studies on this gene and its protein product.
Key words: *Drosophila melanogaster*; nucleolus; *Minute*-like phenotype; Non3; Fibrillarin; Rpf2; Brix domain; ribosome biogenesis.

For citation: Andreyeva E.N., Ogienko A.A., Yushkova A.A., Popova J.V., Pavlova G.A., Kozhevnikova E.N., Ivankin A.V., Gatti M., Pindyurin A.V. *Non3* is an essential *Drosophila* gene required for proper nucleolus assembly. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):190-198. DOI 10.18699/VJ19.481

Ген *Non3* необходим для формирования ядрышка у *Drosophila*

Е.Н. Андреева[1] ✉, А.А. Огиенко[1, 2], А.А. Юшкова[1, 2], Ю.В. Попова[1, 3], Г.А. Павлова[1], Е.Н. Кожевникова[1, 3], А.В. Иванкин[1], М. Гатти[1, 4], А.В. Пиндюрин[1, 2, 3] ✉

[1] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[4] Институт молекулярной биологии и патологии Национального исследовательского совета и Отделение биологии и биотехнологии, Римский университет Ла Сапиенса, Рим, Италия
✉ e-mail: andreeva@mcb.nsc.ru; a.pindyurin@mcb.nsc.ru

Ядрышко представляет собой динамичную немембранную внутриядерную органеллу, которая играет ключевую роль как в биогенезе рибосом, так и в других клеточных процессах. Нарушение функции ядрышка ассоциировано со многими заболеваниями человека, в том числе с дегенеративными патологиями нервной и сердечно-сосудистой систем, а также с образованием злокачественных опухолей. В данной работе нами впервые охарактеризована функция гена *Non3* (*Novel nucleolar protein 3*) у плодовой мушки *Drosophila melanogaster*, который кодирует гомолог Brix домен-содержащего белка человека Rpf2, участвующего в процессе созревания рибосомной РНК (рРНК). С помощью метода неточной эксцизии *P*-элемента мы получили набор из четырех мутаций по гену *Non3*. Эти мутации формируют аллельный ряд, который включает в себя как жизнеспособные, так и летальные аллели. $Non3^{\Delta 600}$ – это нуль-аллель, который несет делецию большей части гена и является ранней рецессивной леталью. Мухи генотипа $Non3^{\Delta 600}$/+ демонстрируют очень слабое уменьшение длины и толщины торакальных щетинок, но при этом развиваются нормально и фертильны. Гетероаллельные комбинации жизнеспособных мутаций гена *Non3* ($Non3^{197}$, $Non3^{310}$ и $Non3^{259}$) имеют укоро-

E.N. Andreyeva and A.A. Ogienko contributed equally to this work.

ченные и тонкие торакальные щетинки, также у них наблюдается некоторое замедление онтогенеза. Такой же паттерн нарушений был ранее описан в литературе как *Minute*-like фенотип, который характерен при дефектах биогенеза рибосом. Кроме того, мы обнаружили, что белок Non3 является компонентом ядрышка в клетках вентрально-мозгового ганглия личинок третьего возраста и необходим для локализации белка Fibrillarin, важного для пост-трансляционной модификации и процессинга рРНК, в этом немембранном внутриядерном субкомпартменте. Полученный нами набор генетических и биохимических инструментов, использованных в ходе данной работы для первичной характеристики гена *Non3*, будет полезен также для исследований функции этого гена и его белкового продукта в будущем.
Ключевые слова: *Drosophila melanogaster*; ядрышко; фенотип *Minute*-like; Non3; Fibrillarin; Rpf2; Brix домен; биогенез рибосом.

Introduction

Ribosomes are highly conserved macromolecular machines that organize and catalyze mRNA translation in all organisms. In higher eukaryotes, mature cytoplasmic ribosomes include 4 ribosomal RNA molecules (rRNAs; 18S, 25S/28S, 5.8S and 5S) and 79 ribosomal proteins (RPs) (Marygold et al., 2007; Xue, Barna, 2012; Genuth, Barna, 2018). *Drosophila melanogaster* has a 2S rRNA instead of 5S rRNA (Stage, Eickbush, 2007). Coordinated processing of 45S/47S precursor rRNA (pre-rRNA) into the mature 18S, 5.8S and 25S/28S rRNAs is a central process in the highly orchestrated ribosome assembly (Henras et al., 2015). Ribosomal biosynthesis includes progressive association of individual ribosomal proteins with maturating rRNAs, export of pre-ribosome particles from the nucleolus to the cytoplasm and their assembly into mature ribosomes. More than 200 different factors are required for ribosome biosynthesis and quality control of ribosome maturation on the path from the nucleolus to the cytoplasm (Kressler et al., 2017).

Depletion of eukaryotic RPs could be associated with disturbances in pre-rRNA processing and nucleolar organization (Neumüller et al., 2013; Farley-Barnes et al., 2018). In *Drosophila*, mutations in 75 % of the genes encoding protein components of mature cytoplasmic ribosomes dominantly cause a *Minute* phenotype. The *Minute* syndrome includes short and thin bristles, delayed development, and reduced viability and fertility (Marygold et al., 2007). In addition, decreased levels of some *Drosophila* RPs result in overgrowth of specific tissues and melanotic tumors (Goudarzi, Lindström, 2016). Mutations in genes encoding RPs and ribosome biogenesis factors lead to a series of congenital human disorders collectively called ribosomopathies, and predispose to cancer (Narla, Ebert, 2010; Henras et al., 2015; Mills, Green, 2017; Núñez Villacís et al., 2018). In some cases, these diseases are also caused by haploinsufficiency for these genes caused by somatic mutations (Narla, Ebert, 2010; Núñez Villacís et al., 2018).

Although ribosomes have been considered for many years as ancient and rather invariable molecular machines, several lines of evidence indicate that they are instead heterogeneous in both RPs and rRNA composition (Genuth, Barna, 2018). Indeed, mutations in RP coding genes lead to tissue-specific pleiotropic phenotypes in multicellular organisms (Kongsuwan et al., 1985; Draptchinskaia et al., 1999; Marygold et al., 2005; Gupta, Warner, 2014; Shi, Barna, 2015). Pleiotropic phenotypes could be only partially explained by the fact that several of these genes have paralogues (Xue, Barna, 2012). For example, in humans there are three paralogous RP genes (*RPS4X*, *RPS4Y1* and *RPSY2*) encoding the RPS4 protein, and one of them (*RPSY2*) is specifically expressed in testis and prostate (Fisher et al., 1990; Lopes et al., 2010). In *Drosophila*, a total of 88 genes encoding 79 different RPs have been identified; nine of these genes are present as duplicates (Marygold et al., 2007). These duplicated genes are expressed at different levels, and in some cases with a tissue-specific pattern. For instance, *RpL22L*, *RpS5b*, *RpS19a*, *RpL10Aa* and *RpL37b* exhibit enhanced expression in the testes compared to their paralogues, suggesting specific composition of testis ribosomes (Marygold et al., 2007; Kearse et al., 2011).

A complex sequence of processing steps, involving several protein factors, is required to gradually release the mature rRNAs from precursor pre-rRNA (Henras et al., 2015). Many of these factors contain putative RNA-binding domains (e. g., GAR, RRM, KH, Brix, S1, dsRBD, and Zinc finger) and/or protein-protein interaction domains (WD40, HEAT, TPR, and HAT), but the enzymatic activities of most proteins required for ribosome biosynthesis have not been determined (Henras et al., 2015). One of the Brix domain-containing proteins is the *Saccharomyces cerevisiae* ribosome assembly factor Rpf2 (Ribosome production factor 2), which in complex with Rrs1 plays a role in the early steps of the 60S ribosome subunit maturation (Zhang et al., 2007; Henras et al., 2015; Kressler et al., 2017). Rpf2 binds the Rpl5 and Rpl11 ribosomal proteins and the Rrs1 protein, forming both the 5S ribonucleoprotein particle necessary for 25S rRNA maturation and the large 60S ribosomal subunit (Tutuncuoglu et al., 2016). Depletion of Rpf2 results in defects in pre-rRNA processing (Wehner, Baserga, 2002).

The Rpf2 proteins are highly conserved. The *Drosophila* Rpf2 orthologous protein is encoded by the *Novel nucleolar protein 3* (*Non3*) gene (*CG7993*). Non3, which contains a Brix domain, exhibits 66 % similarity and 47 % sequence identity with human Rpf2 (Gramates et al., 2017). Thus far, most studies on Non3 were carried in tissue culture cells. Most interestingly, an RNAi-based screen showed that Non3 depletion results in short mitotic spindles. In addition, it has been shown that in interphase cells the GFP-tagged Non3 protein localizes to the nucleolus (Moutinho-Pereira et al., 2013). These phenotypes, and the finding that RNAi-mediated depletion of other nucleolar proteins results in short spindles, suggest that the short-spindle phenotype observed in *Non3* RNAi cells is due to limited translation of tubulin and/or other spindle components. However, it is also possible that Non3 has a direct role in spindle formation (Moutinho-Pereira et al., 2013).

To address the mitotic role of Non3, and to provide a *Drosophila* model for the study of the Rpf2 function in nucleolus assembly and ribosome maturation, we characterized the *Non3* gene. We generated several allelic mutations in the gene and

showed that these mutations affect viability, fertility and bristle formation, resulting in a *Minute*-like phenotype. We also show that Non3 localizes to the nucleolus and is required for proper formation of this organelle. We believe that the mutations and the reagents generated in this study will be instrumental to define the role of *Non3* in living flies.

Materials and methods

Fly stocks. Flies were raised and crossed on standard cornmeal agar media at 25 °C. The fly stocks used in this study are from the Bloomington Stock Center (Bloomington, IN, USA; flystocks.bio.indiana.edu): #30094 (w^{1118}; P{w^{+mC} = EP} $Non3^{G4706}$/TM6C, Sb^{1}) (hereafter $Non3^{G4706}$) and #4368 (*y*, *w*; *Ki*, P{ry^{+}, Δ2-3}99B) (hereafter Δ2-3). The y^{1}, P{$y^{+t7.7}$ = nosphiC31\int.NLS}X; P{$y^{+t7.7}$ = CaryP}attP40 fly line was kindly provided by Sergei A. Demakov (Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia).

Generation of new *Non3* alleles. The *Non3* mutants were generated by imprecise excision (O'Brochta et al., 1991) of the P{EP} transposon located 48 bp downstream of the predicted transcription start site of *Non3* in the $Non3^{G4706}$ line (Fig. 1). The Δ2-3 strain was used as a source of transposase (Robertson et al., 1988), and the F_1 progeny of $Non3^{G4706}$/Δ2-3 flies was screened for loss of the mini-*white* (w^{+mC}) marker carried by the P{EP} transposon. 512 independent excision lines were established and analyzed by PCR using primers Ins-non3-F2 (5′-CGGTTGTTTTCACATCCCTAAC-3′) and Ins-non3-R (5′-CGTCTGTGCTAATGTTCTTCTTCTTG-3′) (primer positions are shown in Fig. 1). We identified one line with a large deletion internal to the *Non3* gene, three lines with remnants of the P{EP} transposon, and several lines in which the transposon was most probably precisely excised. Only one of the latter lines was further analyzed and confirmed to be generated by precise excision of the P{EP} transposon; it was designated as $Non3^{ex}$. The PCR products obtained from all new *Non3* mutant lines were cloned in a plasmid vector and sequenced (Suppl. Fig. 1)[1].

Construction of rescue plasmid and germ-line transformation. To make a rescue construct, we cloned a 2.76-kb genomic DNA fragment [chr3R:18222722–18225482; the coordinates are from Release 6 of the *Drosophila melanogaster* genome assembly (Hoskins et al., 2015)] carrying the *Non3* gene into the pUASTattB vector (Bischof et al., 2007) by substituting a UAS promoter, a multiple cloning site and a SV40 terminator. Details of plasmid construction are available upon request. The cloned genomic DNA fragment contains 27 known single nucleotide polymorphisms (SNPs) (Mackay et al., 2012; Huang et al., 2014) (Suppl. Fig. 2). The rescue construct was injected at the concentration of 300 ng/μl into embryos carrying attP40 landing site (Markstein et al., 2008) and expressing the phiC31 integrase in the germline (Bischof et al., 2007).

Anti-Non3 antibody production. The full-length *Non3* coding sequence (corresponding to nucleotides 98–1060 of GenBank accession no. NM_142437.3, but with the synonymous nucleotide substitution 1000T>A) was PCR-amplified from 0–24 h *Drosophila* wild-type (*Canton-S*) embryonic cDNA library. The amplified DNA fragment was cloned in-frame into the pGEX-4T-1 plasmid vector (GE Healthcare) downstream of the glutathione S-transferase (GST) coding sequence to produce pGEX-4T-Non3 construct. Details of plasmid construction are available upon request. The GST-Non3 fusion protein was expressed in *Escherichia coli* and subsequently purified as described previously (Chalkley, Verrijzer, 2004). The purified protein was used to immunize mice. Polyclonal antibodies were affinity purified from serum as previously described (Chalkley, Verrijzer, 2004).

Double-stranded RNA (dsRNA) production. A 730-bp fragment of the *Non3* coding sequence was PCR-amplified with primers CG7993-rnaF1 (5′-TAATACGACTCACTATAGGGAGGTGTTGCTGGCCAG-3′) and CG7993-rnaR1 (5′-TAATACGACTCACTATAGGGAGGGCGTCTGTGCTAATG-3′) (underlined is the added T7 promoter sequence) from the pGEX-4T-Non3 plasmid. The purified PCR product was used as a template to synthesize dsRNA as described earlier (Somma et al., 2002), with the minor modifications: the phenol/chloroform extraction step was omitted and DNaseI treatment was performed at the end of the procedure.

S2 cell culture and RNA interference (RNAi). S2 cells (for details, see Strunov et al., 2016) were cultured at 25 °C in 39.4 g/L Shields and Sang M3 Insect medium (Sigma, S8398) supplemented with 0.5 g/L $KHCO_3$ and 20 % heat-inactivated fetal bovine serum (FBS) (Thermo Scientific, 10270106). RNAi treatments were carried out as described previously (Somma et al., 2008), with the following modifications: 25 μg of dsRNA was added to the cells three times (on the first, the third and the fifth day of incubation), and cells were harvested for analyses after 7 days of RNAi.

Reverse transcription followed by quantitative PCR (RT-qPCR). RNAi efficiency in S2 cells was assessed by RT-qPCR. Total RNA was isolated from control and dsRNA-treated cells using RNAzol® RT reagent (Molecular Research Center, RN 190); genomic DNA was eliminated using the Rapid-Out DNA Removal Kit (Thermo Fisher Scientific, K2981) according to the manufacturer's instructions. Synthesis of cDNA and qPCR were performed as described previously (Ogienko et al., 2018), using the following gene-specific primers: RT-Non3-Fw2 (5′-CGCTTTTACGCATCAGGAAACC-3′) and RT-Non3-Rev2 (5′-CTTCCTTCCGTCCAAAAACAGC-3′) for *Non3* (this study), and RPL32-realtime-F (5′-CTAAGCTGTCGCACAAATGG-3′) and RPL32-realtime-R (5′-AGGAACTTCTTGAATCCGGTG-3′) for *RpL32* (Yang et al., 2013), which was used as a reference gene.

Western blotting. S2 cells were harvested by centrifugation at 200 *g* for 5 min at room temperature, washed with phosphate-buffered saline (PBS) and centrifuged again. Pellets were lysed in RIPA buffer (Sigma, R0278) containing 1× Halt™ Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific, 1861282), and the lysates were clarified by centrifugation at 15.000 *g* for 15 min at 4 °C. The samples were normalized to the total amount of protein using DC Protein Assay (Bio-Rad, 5000116). Each normalized sample was mixed with an equal volume of 2× Laemmli buffer and incubated at 95 °C for 5 min prior to loading on a SDS-PAGE gel. Larval tissues were dissected, homogenized in 1× Laemmli buffer with a pestle and incubated at 95 °C for 5 min prior to loading on a SDS-PAGE gel. The following primary

[1] Supplementary Figures 1, 2 are available in the online version of the paper: http://www.bionet.nsc.ru/vogis/download/pict-2019-23/appx3.pdf

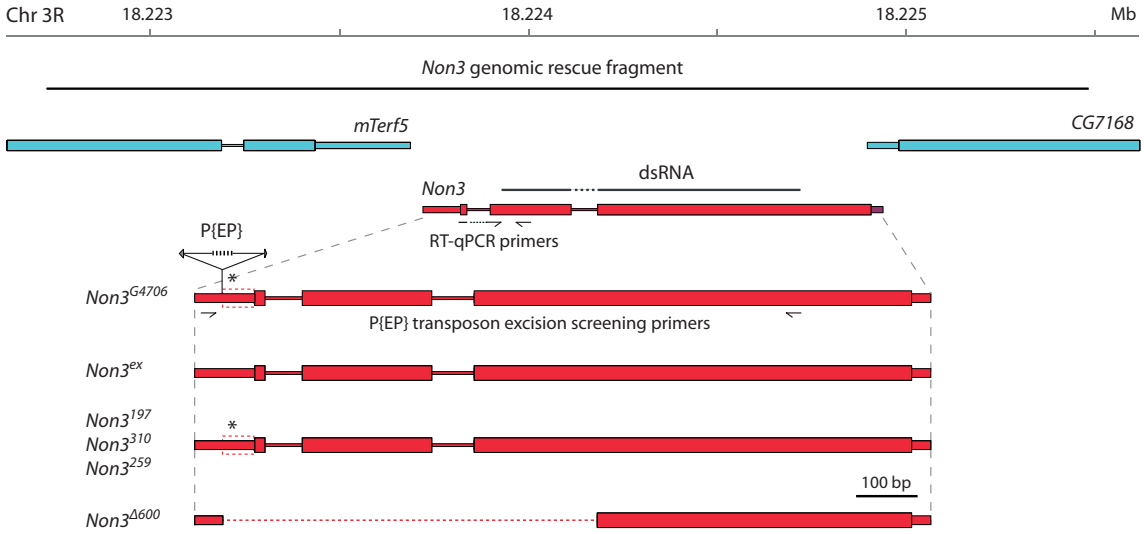


Fig. 1. Schematic representation of the genomic locus containing *Non3* and its flanking genes.

The *Non3* gene, which is on the forward strand, is shown in red. The *mTerf5* and *CG7168* genes, which are on the reverse strand, are shown in light blue. Coding sequences, UTRs and introns are represented by wide bars, narrow bars and lines, respectively. The $Non3^{G4706}$ mutation is caused by an insertion of the P{EP} transposon (which is not shown to scale) 48 bp downstream of the predicted transcription start of the gene. $Non3^{\Delta 600}$ is a 617-bp deletion removing the region (indicated by a dotted line) downstream from the P{EP} insertion site. $Non3^{197}$, $Non3^{310}$ and $Non3^{259}$ are hypomorphic mutations carrying remnants of *P*-element ends, which were generated by imprecise excision of the P{EP} transposon. $Non3^{ex}$ has been generated by a precise excision of the P{EP} transposon without any disruption of the gene. The arrows indicate the position of primer pairs used for identification and characterization of the mutations. The alternative coding sequence in the $Non3^{G4706}$, $Non3^{197}$, $Non3^{310}$ and $Non3^{259}$ alleles, which is likely to start from an ATG codon located within the 5′ *P*-element remnants, is delimited by dotted rectangles and marked with asterisks (for additional details see Suppl. Fig.1). The DNA segments used for the rescue construct and for the dsRNA synthesis are shown above the *Non3* gene.

antibodies were used: mouse anti-Lamin Dm0 (1:300; DSHB, ADL67.10), mouse anti-α-Tubulin (1:5000; Sigma, T6199), mouse anti-β-Tubulin [1:800; BX69 (Tavares et al., 1996), kindly provided by Harald Saumweber (Humboldt University Berlin, Institute of Biology, Berlin, Germany)] and mouse anti-Non3 (1:5000; this study). The primary antibodies were detected with HRP-conjugated goat α-mouse IgG (1:3.500; Life Technology, G-21040) and images were captured using an Amersham Imager 600 System (GE Healthcare).

Immunofluorescence (IF) staining. Brains from late third instar larvae were dissected and immunostained as described earlier (Bonaccorsi et al., 2000). The following primary antibodies were used: mouse anti-Non3 (1:250; this study) and rabbit anti-Fibrillarin (1:300; Abcam, ab5821). They were detected with goat anti-mouse IgG conjugated to Alexa Fluor 488 (1:500; Invitrogen, A-11001) and goat anti-rabbit IgG conjugated to Alexa Fluor 568 (1:350; Invitrogen, A-11036), respectively. DAPI was used to stain DNA. IF images were acquired with a Zeiss Axio Imager M2 fluorescence microscope equipped with an Axiocam 506 mono (D) camera and a NeoFluar 100× /1.3 Oil objective using the ZEN 2012 software.

Results and discussion

Generation and characterization of the novel *Non3* mutations

D. melanogaster Non3 gene (*CG7993*) maps to the 90F8 region of polytene chromosomes, includes two small introns, and has a total size of 1221 bp. It is ubiquitously expressed, with the highest expression levels in embryos (0–12 h), nervous ganglia, ovaries and testes (Gramates et al., 2017). When we began this work, only a *P*-element-induced mutation in the *Non3* gene ($Non3^{G4706}$) was available. Animals homozygous for this mutation are lethal and die at the late pupal stages. We sequenced $Non3^{G4706}$ and confirmed that this mutation carries a P{EP} transposon inserted 48 bp downstream of the predicted transcription start of the gene (see Fig. 1, Suppl. Fig. 1). We then used imprecise *P*-element excision to generate a set of additional mutations in *Non3*. We isolated three hypomorphic alleles ($Non3^{197}$, $Non3^{310}$ and $Non3^{259}$) and one putative null allele ($Non3^{\Delta 600}$) (see Material and Methods for details). As a control, we also generated a chromosome bearing a precise excision of the P{EP} transposon ($Non3^{ex}$).

Flies homozygous for $Non3^{ex}$ are fully viable and fertile, excluding a possible influence of background mutations on the "starting" chromosome (see Fig. 1). All mutations generated by imprecise excision were sequenced and their precise locations and molecular structures are shown in Figure 1 and Suppl. Fig. 1. The $Non3^{\Delta 600}$ allele carries a deletion of 617 bp of the *Non3* coding region and contains 13 residual bp of 3′ end of the *P*-element. The $Non3^{197}$, $Non3^{310}$ and $Non3^{259}$ mutations carry differently sized remnants of the 5′ and 3′ *P*-element ends (373, 367 and 453 bp, respectively) that do not disrupt the coding region of the gene. $Non3^{ex}$ did not show any sequence variation compared to wild type.

Non3 mutants exhibit a *Minute*-like phenotype

We next performed a phenotypic analysis and a functional categorization of the new mutations. This investigation was possible because the "starting" $Non3^{G4706}$ chromosome does not carry background mutations, as shown by the charac-

Complementation analysis of *Non3* mutant alleles

♂ \ ♀	$Non3^{ex}$	$Non3^{197}$	$Non3^{310}$	$Non3^{259}$	$Non3^{G4706}$	$Non3^{\Delta 600}$
$Non3^{ex}$	V, F, NB	V, F, NB	V, F, NB	V, F, NB	V, F, NB	V, F, NB
$Non3^{197}$		V, F, db	V, F, db	V, F, db	V, ss, DB	PLE, S, DB
$Non3^{310}$			V, F, db	V, F, db	V, ss, DB	PL
$Non3^{259}$				V, ss, db	V, S, DB	PL
$Non3^{G4706}$					PL	PL
$Non3^{\Delta 600}$						EL

Note: V, viable both sexes; F, fertile both sexes; EL, early lethal (L1/L2); PL, pupal lethal; PLE, pupal lethal with escapers; S, sterile in both sexes; ss, semi sterile; NB, normal bristles; DB, defective bristles; db, slightly defective bristles.

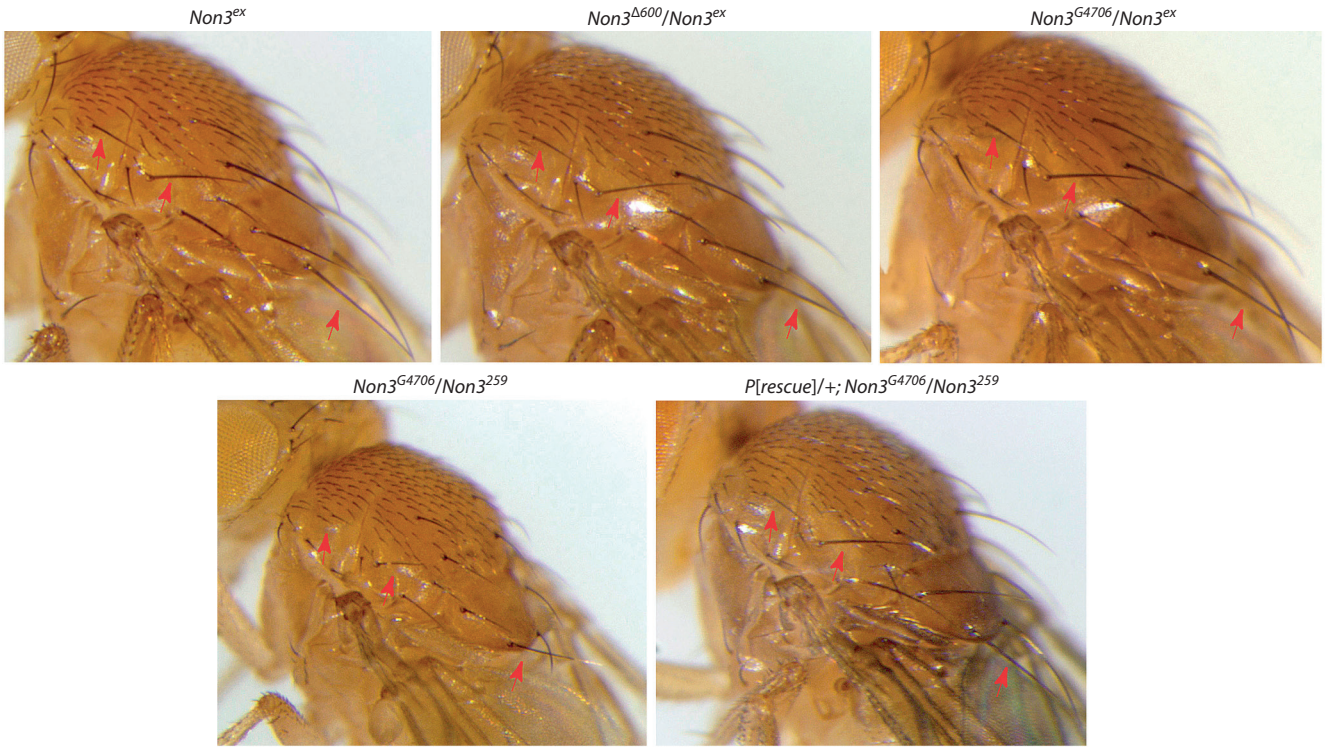


Fig. 2. *Non3* mutants exhibit a recessive *Minute*-like bristle phenotype.

$Non3^{G4706}/Non3^{259}$ flies have shorter and thinner bristles than $Non3^{ex}$ used as control. This is most clearly seen by comparing the notum bristles (red arrows). $Non3^{\Delta 600}/Non3^{ex}$ and $Non3^{G4706}/Non3^{ex}$ flies exhibit a very weak dominant bristle phenotype. Addition of a wild-type copy of *Non3* (*P*[*rescue*]) to $Non3^{G4706}/Non3^{259}$ flies leads to a complete rescue of the *Minute*-like phenotype.

terization of $Non3^{ex}$ precise excision line. We performed a complementation analysis between all extant *Non3* alleles, including $Non3^{G4706}$ (see the Table). This analysis showed that the $Non3^{\Delta 600}$ allele is a homozygous larval lethal that dies at the first instar larval (L1) stage, a finding that suggests that $Non3^{\Delta 600}$ is a genetically null mutation. The $Non3^{G4706}$ homozygotes and the $Non3^{\Delta 600}/Non3^{G4706}$ heterozygotes are also lethal but die during pupal development. All the remaining *Non3* mutations ($Non3^{197}$, $Non3^{310}$ and $Non3^{259}$) are homozygous viable. However, $Non3^{\Delta 600}/Non3^{259}$ and $Non3^{\Delta 600}/Non3^{310}$ heterozygotes are mostly pupal lethal. $Non3^{\Delta 600}/Non3^{197}$ also exhibit pupal lethality but produce some escapers with reduced fertility. In addition, viable mutant combinations bearing the Non^{310} allele are more fertile than those carrying $Non3^{259}$. These results suggest that the *Non3* mutations can be ordered in an allelic series with $Non3^{\Delta 600} > Non3^{G4706} > Non3^{259} > Non3^{310} > Non3^{197}$.

In *Drosophila*, mutations in 75 % of the RP coding genes dominantly produce a *Minute* phenotype characterized by prolonged development, short and thin bristles, and reduced viability and fertility, often accompanied by additional patterning and growth defects such as roughened eyes, abnormal wings, defective abdominal segmentation, and small body size (Marygold et al., 2005). Bristle production and gametogenesis require maximal protein synthesis and are therefore particularly sensitive to a reduction in the translational capacity of

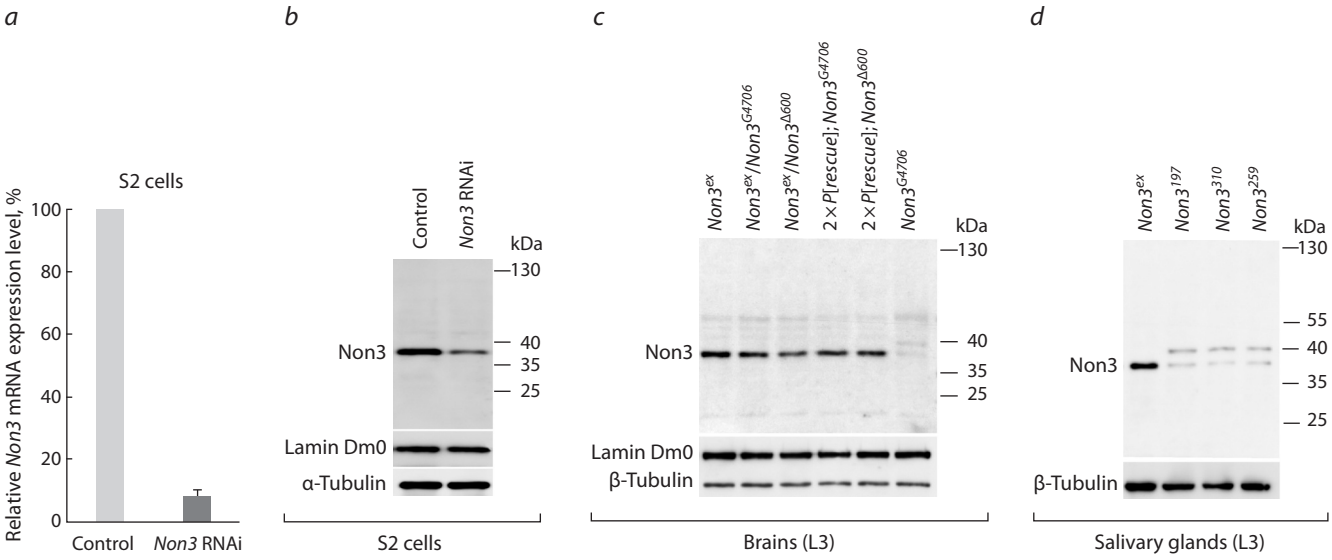


Fig. 3. Expression of the Non3 protein in S2 cells and in larval tissues from *Non3* mutants.

(*a*) RT-qPCR showing that in *Non3* RNAi S2 cells the target mRNA level is strongly reduced compared to control cells. Each bar represents the average of three separate biological replicates (RNA isolations); error bar, standard error of the mean (SEM). (*b*) Western blot of S2 cell extracts showing that the anti-Non3 antibody specifically recognizes one band of the predicted size (~37 kDa), which is substantially reduced in *Non3* RNAi cells. Lamin Dm0 and α-Tubulin are loading controls. (*c*) Western blot from larval brain extracts showing a clear Non3 band in all samples expressing a wild-type ($Non3^{ex}$) copy of the *Non3* gene; this band is strongly reduced in $Non3^{G4706}$ homozygous mutants. Lamin Dm0 and β-Tubulin are loading controls. $Non3^{\Delta 600}$ homozygotes could not be analyzed because they are early lethals. (*d*) Western blot from larval salivary gland extracts showing that in $Non3^{259}$, $Non3^{310}$ and $Non3^{197}$ homozygous mutants the level of the wild-type Non3 protein is substantially reduced. In addition, these mutants exhibit aberrant proteins of 39–40 kDa, which are probably transcribed from the ATG codons located within the remnants of the P{EP} transposon present in these mutants (see text, Fig. 1 and Suppl. Fig. 1 for detailed explanation). β-Tubulin is a loading control.

the cell (Marygold et al., 2005). It is thus generally accepted that the *Minute* syndrome reflects a reduced protein synthesis resulting from insufficient ribosome function (Morata, Ripoll, 1975; Marygold et al., 2005). Mutations in several *Drosophila* genes involved in ribosome biosynthesis do not have dominant effects but, when homozygous, cause the same defective traits elicited by the dominant *Minute* mutations. The phenotype produced by these mutations is commonly designated as *Minute*-like (see for example, Cui, DiMario, 2007).

$Non3^{\Delta 600}$/+ and $Non3^{G4706}$/+ heterozygous flies occasionally exhibit a limited reduction in bristle size but develop normally and are fertile. Thus, *Non3* mutations have only a minimal dominant effect. However, the combination of the $Non3^{G4706}$ allele with any other weaker allele results in a clear *Minute*-like phenotype: prolonged development, poor viability and fertility, and abnormally short and thin bristles (Table, Fig. 2) (some flies had crumpled wings). To verify that all *Non3* phenotypes were due to a decrease in the Non3 protein level, we performed rescue experiments using a transgene that carries a full genomic copy of *Non3* with its adjacent regulatory regions (see Fig. 1 and Suppl. Fig. 2). Mutant *Non3* flies carrying one copy of this rescue construct (*P*[*rescue*]) showed normal viability and bristle length, and restored fertility (Fig. 2 and data not shown). Taken together, these findings indicate that *Non3* mutations behave like *Minute*-like mutations (Cui, DiMario, 2007). The finding that mutants in *Non3* exhibit a *Minute*-like phenotype is not surprising. The *S. cerevisiae* Rpf2 protein is one of the many factors necessary for the assembly of the pre-60S subunits (Beidka et al., 2018) and it is likely that its *Drosophila* homologue Non3 participates in the same processes.

Non3 hypomorphic mutations reduce the Non3 protein level

To further characterize the *Non3* mutant alleles, we raised a polyclonal anti-Non3 antibody in mice (see Materials and Methods for details). This antibody recognizes a band of the expected molecular weight (~37 kDa) in Western blots from S2 cell protein extracts; this band is substantially reduced in the extracts from *Non3* RNAi cells, confirming the specificity of the antibody (Fig. 3, *a*, *b*). It was suggested earlier that the short-spindle phenotype observed in *Non3* RNAi S2 cells is caused by limited translation of tubulin and/or other spindle components (Moutinho-Pereira et al., 2013). Our finding that *Non3* RNAi-treated cells exhibit a normal tubulin level (see Fig. 3, *b*) indicates that this phenotype is not caused by limited tubulin availability.

We next analyzed the levels of the Non3 protein in larval brains and salivary glands from the different *Non3* mutants (see Fig. 3, *c*, *d*). In larval brain protein extracts from $Non3^{G4706}$ homozygotes, the Non3 protein is drastically reduced, supporting the view that $Non3^{G4706}$ is the strongest among the *Non3* hypomorphic mutations. Two copies of the rescue construct (2 × *P*[*rescue*]) restored the Non3 protein level in both $Non3^{G4706}$/$Non3^{G4706}$ and $Non3^{\Delta 600}$/$Non3^{\Delta 600}$ mutants, approximately up to the level of $Non3^{ex}$ homozygous brains (see Fig. 3, *c*). We also found that Non3 is substantially reduced in salivary glands from $Non3^{197}$, $Non3^{310}$ and $Non3^{259}$ homozygotes. In addition to a reduction of the wild-type Non3 protein, the salivary glands of these mutants displayed immunoreactive proteins of ~39–40 kDa (see Fig. 3, *d*). Because the $Non3^{197}$, $Non3^{310}$ and $Non3^{259}$ mutations are carrying remnants of *P*-element ends, we believe that the aberrant Non3 proteins

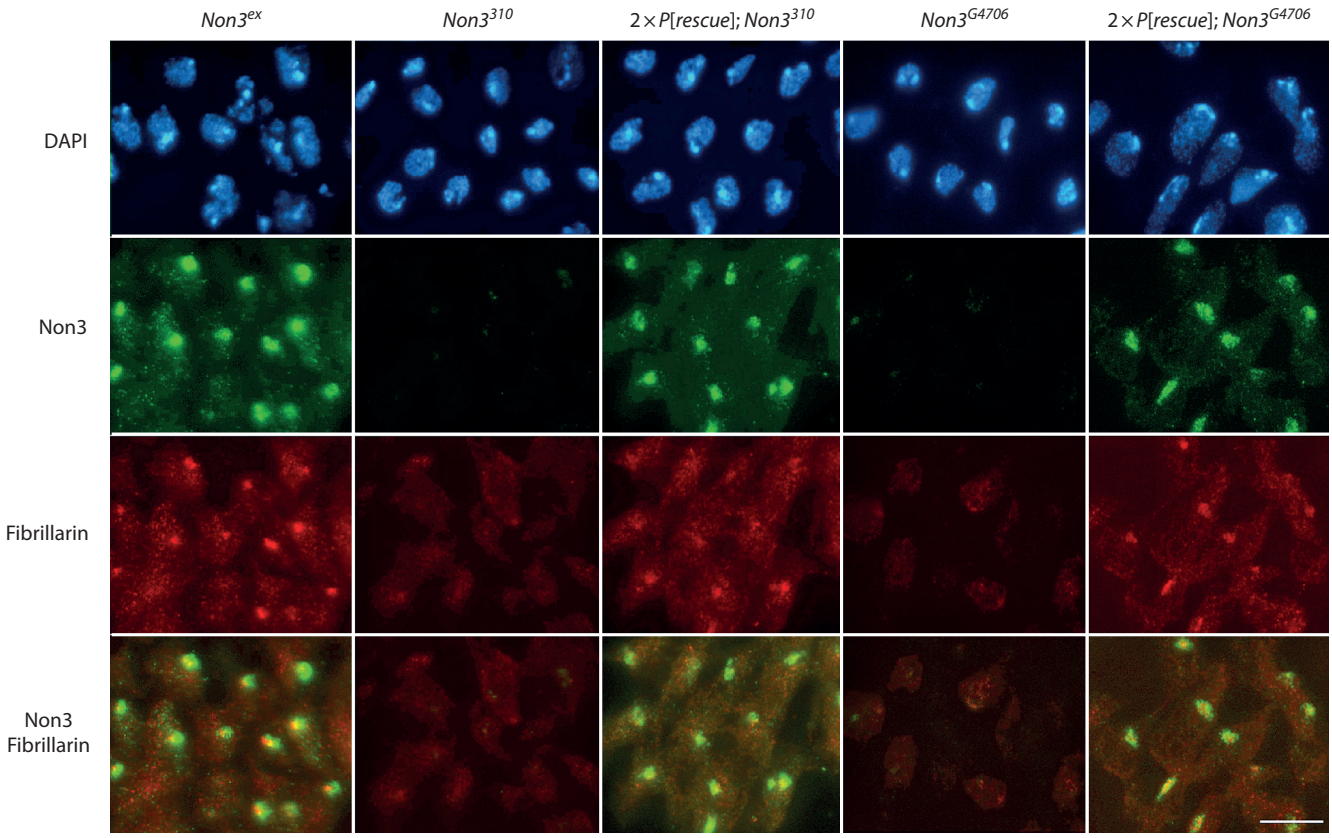


Fig. 4. Non3 localizes to the nucleoli of larval brain cells.

In wild-type (*Non3*^{ex}) brain cells, Non3 largely co-localizes with the nucleolus marker Fibrillarin. Notably, brain cells from *Non3*^{G4706} and *Non3*^{310} homozygous larvae not only show a strong decrease of Non3 signals but also of Fibrillarin signals. Scale bar is 10 μm.

are transcribed starting from ATG codons located within these *P*-element fragments (see Fig. 3, *d*; Suppl. Fig. 1).

Nucleolar Non3 protein is required for proper Fibrillarin localization

It has been previously shown that S2 cells transiently transfected with a plasmid encoding a GFP-Non3 fusion protein show a specific localization of the protein in the nucleolus (Moutinho-Pereira et al., 2013). We tried to confirm this Non3 localization by immunostaining S2 cells with our antibody, but were unable to see a clear signal. However, we successfully immunostained the nucleoli of larval brain cells (Fig. 4), possibly because nervous ganglia are one of the tissues with the highest level of *Non3* expression (Gramates et al., 2017). In wild-type brain cells, Fibrillarin, the main component of the active nucleolus (Neumüller et al., 2013), co-localizes with Non3 in the nucleolus. In brain cells of homozygous *Non3*^{G4706} and *Non3*^{310} mutants, we observed a drastic reduction of both Non3 and Fibrillarin in the nucleolus, suggesting that Non3 is required for Fibrillarin localization in nucleoli. The presence of the *Non3* rescue transgenes (2×*P*[*rescue*]) restored the normal nucleolar signals of both proteins (see Fig. 4).

These results indicate that Non3 is a nucleolar protein that is required for proper Fibrillarin localization in the nucleolus. Fibrillarin is a major nucleolar protein with methyltransferase activity, playing roles in rRNA biogenesis and function (Rodriguez-Corona et al., 2015). Loss of nucleolar proteins such as Nopp140, Nop56, and Nop5 leads to mislocalization of Fibrillarin, compromises several nucleolar functions (Pederson, 1998; Olson, 2004), and causes developmental abnormalities (Cui, DiMario, 2007). The role of Non3 in Fibrillarin recruitment to the nucleolus is currently unknown and will be addressed in future studies.

Conclusions

We have generated several mutant alleles of the *Non3* gene and shown that viable combinations of these alleles exhibit a *Minute*-like phenotype, suggesting a role of Non3 in ribosome biogenesis. We have also shown that the Non3 protein localizes to the nucleolus and, most importantly, it is required for Fibrillarin recruitment to this organelle. Fibrillarin is a multifunctional protein that mediates methylation of several RNA species and plays roles in tumorigenesis and stem cell differentiation (Rodriguez-Corona et al., 2015; van Nues, Watkins, 2017). We believe that the Non3-related tools generated in our study will be instrumental to define the role of Non3 in RNA metabolism and to elucidate how it affects mitotic spindle formation. These tools will also help to investigate the mechanisms and biological significance of the interaction between Non3 and Fibrillarin.

References

Biedka S., Micic J., Wilson D., Brown H., Diorio-Toth L., Woolford J.L., Jr. Hierarchical recruitment of ribosomal proteins and assembly factors remodels nucleolar pre-60S ribosomes. J. Cell Biol. 2018;217(7):2503-2518. DOI 10.1083/jcb.201711037.

Bischof J., Maeda R.K., Hediger M., Karch F., Basler K. An optimized transgenesis system for *Drosophila* using germ-line-specific φC31 integrases. Proc. Natl. Acad. Sci. USA. 2007;104(9):3312-3317. DOI 10.1073/pnas.0611511104.

Bonaccorsi S., Giansanti M.G., Gatti M. Spindle assembly in *Drosophila* neuroblasts and ganglion mother cells. Nat. Cell Biol. 2000; 2(1):54-56. DOI 10.1038/71378.

Chalkley G.E., Verrijzer C.P. Immuno-depletion and purification strategies to study chromatin-remodeling factors *in vitro*. Methods Enzymol. 2004;377:421-442. DOI 10.1016/S0076-6879(03)77028-1.

Cui Z., DiMario P.J. RNAi knockdown of Nopp140 induces *Minute*-like phenotypes in *Drosophila*. Mol. Biol. Cell. 2007;18(6):2179-2191. DOI 10.1091/mbc.e07-01-0074.

Draptchinskaia N., Gustavsson P., Andersson B., Pettersson M., Willig T.N., Dianzani I., Ball S., Tchernia G., Klar J., Matsson H., Tentler D., Mohandas N., Carlsson B., Dahl N. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. Nat. Genet. 1999;21(2):169-175. DOI 10.1038/5951.

Farley-Barnes K.I., McCann K.L., Ogawa L.M., Merkel J., Surovtseva Y.V., Baserga S.J. Diverse regulators of human ribosome biogenesis discovered by changes in nucleolar number. Cell Rep. 2018; 22(7):1923-1934. DOI 10.1016/j.celrep.2018.01.056.

Fisher E.M.C., Beer-Romero P., Brown L.G., Ridley A., McNeil J.A., Lawrence J.B., Willard H.F., Bieber F.R., Page D.C. Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. Cell. 1990;63(6):1205-1218.

Genuth N.R., Barna M. The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. Mol. Cell. 2018;71(3):364-374. DOI 10.1016/j.molcel.2018.07.018.

Goudarzi K.M., Lindström M.S. Role of ribosomal protein mutations in tumor development. Int. J. Oncol. 2016;48(4):1313-1324. DOI 10.3892/ijo.2016.3387.

Gramates L.S., Marygold S.J., Santos G.D., Urbano J.M., Antonazzo G., Matthews B.B., Rey A.J., Tabone C.J., Crosby M.A., Emmert D.B., Falls K., Goodman J.L., Hu Y., Ponting L., Schroeder A.J., Strelets V.B., Thurmond J., Zhou P., the FlyBase Consortium. FlyBase at 25: looking to the future. Nucleic Acids Res. 2017;45(D1):D663-D671. DOI 10.1093/nar/gkw1016.

Gupta V., Warner J.R. Ribosome-omics of the human ribosome. RNA. 2014;20(7):1004-1013. DOI 10.1261/rna.043653.113.

Henras A.K., Plisson-Chastang C., O'Donohue M.F., Chakraborty A., Gleizes P.E. An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip. Rev. RNA. 2015;6(2):225-242. DOI 10.1002/wrna.1269.

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. DOI 10.1101/gr.185579.114.

Huang W., Massouras A., Inoue Y., Peiffer J., Ràmia M., Tarone A.M., Turlapati L., Zichner T., Zhu D., Lyman R.F., Magwire M.M., Blankenburg K., Carbone M.A., Chang K., Ellis L.L., Fernandez S., Han Y., Highnam G., Hjelmen C.E., Jack J.R., Javaid M., Jayaseelan J., Kalra D., Lee S., Lewis L., Munidasa M., Ongeri F., Patel S., Perales L., Perez A., Pu L., Rollmann S.M., Ruth R., Saada N., Warner C., Williams A., Wu Y.Q., Yamamoto A., Zhang Y., Zhu Y., Anholt R.R., Korbel J.O., Mittelman D., Muzny D.M., Gibbs R.A., Barbadilla A., Johnston J.S., Stone E.A., Richards S., Deplancke B., Mackay T.F.C. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. Genome Res. 2014;24(7):1193-1208. DOI 10.1101/gr.171546.113.

Kearse M.G., Chen A.S., Ware V.C. Expression of ribosomal protein L22e family members in *Drosophila melanogaster*: *rpL22-like* is differentially expressed and alternatively spliced. Nucleic Acids Res. 2011;39(7):2701-2716. DOI 10.1093/nar/gkq1218.

Kongsuwan K., Yu Q., Vincent A., Frisardi M.C., Rosbash M., Lengyel J.A., Merriam J. A *Drosophila Minute* gene encodes a ribosomal protein. Nature. 1985;317(6037):555-558.

Kressler D., Hurt E., Baßler J. A puzzle of life: crafting ribosomal subunits. Trends Biochem. Sci. 2017;42(8):640-654. DOI 10.1016/j.tibs.2017.05.005.

Lopes A.M., Miguel R.N., Sargent C.A., Ellis P.J., Amorim A., Affara N.A. The human RPS4 paralogue on Yq11.223 encodes a structurally conserved ribosomal protein and is preferentially expressed during spermatogenesis. BMC Mol. Biol. 2010;11:33. DOI 10.1186/1471-2199-11-33.

Mackay T.F., Richards S., Stone E.A., Barbadilla A., Ayroles J.F., Zhu D., Casillas S., Han Y., Magwire M.M., Cridland J.M., Richardson M.F., Anholt R.R., Barrón M., Bess C., Blankenburg K.P., Carbone M.A., Castellano D., Chaboub L., Duncan L., Harris Z., Javaid M., Jayaseelan J.C., Jhangiani S.N., Jordan K.W., Lara F., Lawrence F., Lee S.L., Librado P., Linheiro R.S., Lyman R.F., Mackey A.J., Munidasa M., Muzny D.M., Nazareth L., Newsham I., Perales L., Pu L.L., Qu C., Ràmia M., Reid J.G., Rollmann S.M., Rozas J., Saada N., Turlapati L., Worley K.C., Wu Y.Q., Yamamoto A., Zhu Y., Bergman C.M., Thornton K.R., Mittelman D., Gibbs R.A. The *Drosophila melanogaster* Genetic Reference Panel. Nature. 2012;482(7384):173-178. DOI 10.1038/nature10811.

Markstein M., Pitsouli C., Villalta C., Celniker S.E., Perrimon N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat. Genet. 2008;40(4):476-483. DOI 10.1038/ng.101.

Marygold S.J., Coelho C.M.A., Leevers S.J. Genetic analysis of *RpL38* and *RpL5*, two *Minute* genes located in the centric heterochromatin of chromosome 2 of *Drosophila melanogaster*. Genetics. 2005;169(2):683-695. DOI 10.1534/genetics.104.034124.

Marygold S.J., Roote J., Reuter G., Lambertsson A., Ashburner M., Millburn G.H., Harrison P.M., Yu Z., Kenmochi N., Kaufman T.C., Leevers S.J., Cook K.R. The ribosomal protein genes and *Minute* loci of *Drosophila melanogaster*. Genome Biol. 2007;8(10):R216. DOI 10.1186/gb-2007-8-10-r216.

Mills E.W., Green R. Ribosomopathies: there's strength in numbers. Science. 2017;358(6363):eaan2755. DOI 10.1126/science.aan2755.

Morata G., Ripoll P. Minutes: mutants of Drosophila autonomously affecting cell division rate. Dev. Biol. 1975;42(2):211-221.

Moutinho-Pereira S., Stuurman N., Afonso O., Hornsveld M., Aguiar P., Goshima G., Vale R.D., Maiato H. Genes involved in centrosome-independent mitotic spindle assembly in *Drosophila* S2 cells. Proc. Natl. Acad. Sci. USA. 2013;110(49):19808-19813. DOI 10.1073/pnas.1320013110.

Narla A., Ebert B.L. Ribosomopathies: human disorders of ribosome dysfunction. Blood. 2010;115(16):3196-3205. DOI 10.1182/blood-2009-10-178129.

Neumüller R.A., Gross T., Samsonova A.A., Vinayagam A., Buckner M., Founk K., Hu Y., Sharifpoor S., Rosebrock A.P., Andrews B., Winston F., Perrimon N. Conserved regulators of nucleolar size revealed by global phenotypic analyses. Sci. Signal. 2013;6(289):ra70. DOI 10.1126/scisignal.2004145.

Núñez Villacís L., Wong M.S., Ferguson L.L., Hein N., George A.J., Hannan K.M. New roles for the nucleolus in health and disease. BioEssays. 2018;40(5):e1700233. DOI 10.1002/bies.201700233.

O'Brochta D.A., Gomez S.P., Handler A.M. P element excision in *Drosophila melanogaster* and related drosophilids. Mol. Gen. Genet. 1991;225(3):387-394.

Ogienko A.A., Yarinich L.A., Fedorova E.V., Lebedev M.O., Andreyeva E.N., Pindyurin A.V., Baricheva E.M. New *slbo*-Gal4 driver lines for the analysis of border cell migration during *Drosophila* oogenesis. Chromosoma. 2018;127(4):475-487. DOI 10.1007/s00412-018-0676-7.

Olson M.O.J. Sensing cellular stress: another new function for the nucleolus? Sci. STKE. 2004;2004(224):pe10. DOI 10.1126/stke.2242004pe10.

Pederson T. The plurifunctional nucleolus. Nucleic Acids Res. 1998; 26(17):3871-3876.

Robertson H.M., Preston C.R., Phillis R.W., Johnson-Schlitz D.M., Benz W.K., Engels W.R. A stable genomic source of *P* element transposase in *Drosophila melanogaster*. Genetics. 1988;118(3):461-470.

Rodriguez-Corona U., Sobol M., Rodriguez-Zapata L.C., Hozak P., Castano E. Fibrillarin from Archaea to human. Biol. Cell. 2015; 107(6):159-174. DOI 10.1111/boc.201400077.

Shi Z., Barna M. Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. Annu. Rev. Cell Dev. Biol. 2015;31:31-54. DOI 10.1146/annurev-cellbio-100814-125346.

Somma M.P., Ceprani F., Bucciarelli E., Naim V., De Arcangelis V., Piergentili R., Palena A., Ciapponi L., Giansanti M.G., Pellacani C., Petrucci R., Cenci G., Vernì F., Fasulo B., Goldberg M.L., Di Cunto F., Gatti M. Identification of *Drosophila* mitotic genes by combining co-expression analysis and RNA interference. PLoS Genet. 2008;4(7):e1000126. DOI 10.1371/journal.pgen.1000126.

Somma M.P., Fasulo B., Cenci G., Cundari E., Gatti M. Molecular dissection of cytokinesis by RNA interference in *Drosophila* cultured cells. Mol. Biol. Cell. 2002;13(7):2448-2460. DOI 10.1091/mbc.01-12-0589.

Stage D.E., Eickbush T.H. Sequence variation within the rRNA gene loci of 12 *Drosophila* species. Genome Res. 2007;17(12):1888-1897. DOI 10.1101/gr.6376807.

Strunov A., Boldyreva L.V., Pavlova G.A., Pindyurin A.V., Gatti M., Kiseleva E. A simple and effective method for ultrastructural analysis of mitosis in *Drosophila* S2 cells. MethodsX. 2016;3:551-559. DOI 10.1016/j.mex.2016.10.003.

Tavares A.A.M., Glover D.M., Sunkel C.E. The conserved mitotic kinase polo is regulated by phosphorylation and has preferred microtubule-associated substrates in *Drosophila* embryo extracts. EMBO J. 1996;15(18):4873-4883.

Tutuncuoglu B., Jakovljevic J., Wu S., Gao N., Woolford J.L., Jr. The N-terminal extension of yeast ribosomal protein L8 is involved in two major remodeling events during late nuclear stages of 60S ribosomal subunit assembly. RNA. 2016;22(9):1386-1399. DOI 10.1261/rna.055798.115.

van Nues R.W., Watkins N.J. Unusual C′/D′ motifs enable box C/D snoRNPs to modify multiple sites in the same rRNA target region. Nucleic Acids Res. 2017;45(4):2016-2028.

Wehner K.A., Baserga S.J. The σ^{70}-like motif: a eukaryotic RNA binding domain unique to a superfamily of proteins required for ribosome biogenesis. Mol. Cell. 2002;9(2):329-339. DOI 10.1016/S1097-2765(02)00438-0.

Xue S., Barna M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat. Rev. Mol. Cell Biol. 2012;13(6): 355-369. DOI 10.1038/nrm3359.

Yang X., Mao F., Lv X., Zhang Z., Fu L., Lu Y., Wu W., Zhou Z., Zhang L., Zhao Y. *Drosophila* Vps36 regulates Smo trafficking in Hedgehog signaling. J. Cell Sci. 2013;126(Pt.18):4230-4238. DOI 10.1242/jcs.128603.

Zhang J., Harnpicharnchai P., Jakovljevic J., Tang L., Guo Y., Oeffinger M., Rout M.P., Hiley S.L., Hughes T., Woolford J.L., Jr. Assembly factors Rpf2 and Rrs1 recruit 5S rRNA and ribosomal proteins rpL5 and rpL11 into nascent ribosomes. Genes Dev. 2007;21(20):2580-2592. DOI 10.1101/gad.1569307.

ORCID ID

E.N. Andreyeva orcid.org/0000-0001-6997-0492
A.A. Ogienko orcid.org/0000-0002-0896-1899
A.A. Yushkova orcid.org/0000-0002-9037-6866
J.V. Popova orcid.org/0000-0001-8800-8645
G.A. Pavlova orcid.org/0000-0001-7361-5039
A.V. Ivankin orcid.org/0000-0003-3309-6363
M. Gatti orcid.org/0000-0003-3777-300X
A.V. Pindyurin orcid.org/0000-0001-6959-0641

Acknowledgements. We thank E.V. Fedorova for technical assistance with generation of *Non3* mutants, G. Reuter for valuable discussion, H. Saumweber for anti-β-Tubulin antibody, and S.A. Demakov for providing the fly line carrying the attP40 landing site. This work was supported by a grant from the Russian Foundation for Basic Research (18-34-00699), by a grant from the Ministry of Education and Science of the Russian Federation (14.Z50.31.0005) for raising antibodies specific to Non3, and by a project of the Fundamental Scientific Research Program of the Siberian Branch of the Russian Academy of Sciences (0310-2019-0005) for generation of mutant *Non3* alleles. The microscopy and DNA sequencing were performed using resources provided by the Molecular and Cellular Biology core facility of the Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences. Antibodies were raised at the Center for Genetic Resources of Laboratory Animals, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (RFMEFI61914X0005 and RFMEFI61914X0010).

Conflict of interest. The authors declare no conflict of interest.

Received December 6, 2018. Revised December 18, 2018. Accepted January 10, 2019.

Вавиловский журнал генетики и селекции. 2019;23(2):199-202
DOI 10.18699/VJ19.482

The study of the regulatory region of the *Drosophila melanogaster Notch* gene by new methods of directed genome editing

O.V. Andreyenkov ✉, E.I. Volkova, N.G. Andreyenkova, S.A. Demakov

Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
✉ e-mail: andreenkov@mcb.nsc.ru

The *Notch* gene plays a key role in the development of organs and tissues of neuroectodermic origin, including the nervous system. In eukaryotic organisms, the *Notch* pathway is involved in cell fate determination. The *Notch* gene was first discovered in *Drosophila melanogaster*. In mammals, the family of Notch receptors includes four homologues. In humans, mutations in the *Notch* gene cause several hereditary diseases and carcinogenesis. Studies of the regulatory zone of the *Notch* gene in *D. melanogaster* have been conducted for several decades. We review their results and methods. The regulatory zone of the *Notch* gene is in the region of open chromatin state that corresponds to the 3C6/3C7 interband on the cytological map of polytene chromosomes of *D. melanogaster* salivary glands. The development of new methods for directed genome editing made it possible to create a system for introducing directed changes into the regulatory zone of the gene. Using the CRISPR/Cas9 system, we obtained a directed 4-kilobase deletion including the 5′-regulatory zone, promoter, and the first exon of the *Notch* gene and introduced the *attP* site into the first intron of the *Notch* gene. This approach enabled targeted changes of the sequence of the regulatory and promoter regions of the gene. Thus, it provided a new powerful tool for studies of *Notch* gene regulation and the organization of the open chromatin state.

Key words: *Drosophila melanogaster*; *Notch* gene; CRISPR/Cas9 system; gene activity regulation; insulator protein; open chromatin state; interbands.

For citation: Andreyenkov O.V., Volkova E.I., Andreyenkova N.G., Demakov S.A. The study of the regulatory region of the *Drosophila melanogaster Notch* gene by new methods of directed genome editing. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):199-202. DOI 10.18699/VJ19.482

Исследования регуляторной зоны гена *Notch* у *Drosophila melanogaster* с использованием новых подходов направленного геномного редактирования

О.В. Андреенков ✉, Е.И. Волкова, Н.Г. Андреенкова, С.А. Демаков

Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ e-mail: andreenkov@mcb.nsc.ru

Ген *Notch* играет ключевую роль в развитии органов и тканей нейроэктодермального происхождения, в том числе нервной системы. У эукариотических организмов сигнальный каскад *Notch* вовлечен в процессы клеточной детерминации. Впервые ген *Notch* был открыт у *Drosophila melanogaster*. У млекопитающих семейство Notch рецепторов включает четыре гомолога. У человека мутации в гене *Notch* приводят к развитию ряда наследственных заболеваний, а также связаны с канцерогенезом. Исследования регуляторной зоны гена *Notch* на *D. melanogaster* насчитывают уже несколько десятилетий. В статье сделан обзор результатов и методов этих исследований. Регуляторная зона гена находится в районе открытого состояния хроматина, соответствующем междиску 3C6/3C7 на цитологической карте политенных хромосом слюнных желез *D. melanogaster*. Развитие новых методов направленного геномного редактирования позволило создать систему для введения направленных изменений в регуляторную зону гена. Используя систему CRISPR/Cas9, мы получили делецию в регуляторной и промоторной зоне гена *Notch*, включая его первый экзон, и ввели сайт *attP* в первый интрон гена *Notch*. Это сделало возможным направленное изменение последовательностей регуляторной и промоторной зон гена.

Ключевые слова: *Drosophila melanogaster*; ген *Notch*; система CRISPR/Cas9; регуляция активности генов; инсуляторные белки; открытое состояние хроматина; междиски.

Introduction

The *Notch* gene is the central element of the evolutionarily conservative gene network that controls embryonic development. It is involved in determining neuroectodermal cell fates through the mechanism of lateral inhibition. This mechanism is based on the ligand-receptor interaction between neighboring cells. Accordingly, the *Notch* gene product is a transmembrane receptor protein. Its ligands are Delta and Serrate. The extracellular domain of the protein, which includes 36 EGF repeats and 3 lin12/Notch repeats, is responsible for binding

to the ligand (Wharton et al., 1985). The protein also includes transmembrane and intracellular domains. Processing of this receptor protein involves several stages of cleavage by specific proteases. Interaction with the ligand also triggers the irreversible separation of the intracellular domain. The latter is transmitted to the nucleus, where it forms a part of a complex that activates the transcription of target genes. Thus, each stage of the signal transmission involves several proteins. One of the key features of the signal transduction system is that the Notch receptor is irreversibly split in the course of signal transduction. The signal transmission, therefore, is conducted without amplification at this stage.

The *Notch* signaling cascade plays a key role in lateral inhibition. This term refers to intercellular interactions in the differentiation of peripheral nervous system cells, leading to the separation of proneuronal cluster cells through the neuronal development pathway and to the suppression of proneuronal gene expression in neighboring cells. This process suppresses the neuronal differentiation potential of neighboring cells and switches them to the epidermal pathway of development (Artavanis-Tsakonas et al., 1999). However, there is evidence that the gene functions are not limited to the processes of lateral inhibition; in particular, the gene also takes part in the formation of organs of neuronal origin (Brennan et al., 1997).

The molecular approaches to study of the Notch regulatory region

A very large number of works have been devoted to the functioning of the gene network as a whole and various stages of signal transmission (Greenwald, 2012). However, it is obvious that the regulation of expression of the gene itself can also influence the functioning of the *Notch* gene network profoundly.

This statement is supported by the existence of numerous mutations caused by various disorders in the *Notch* locus. In *Drosophila melanogaster*, the *Notch* gene occupies a locus of more than 37 kb in size located on the X chromosome. Its position on the cytological map of salivary gland polytene chromosomes corresponds to the 3C6/3C7 region. The entire structural region of the gene, including exons and introns, is in the 3C7 band, and the 5′ regulatory region of the gene upstream of the transcription initiation sites is in the open chromatin of the 3C6/3C7 interband (Rykowski et al., 1988). The long length of the locus is due to several large introns in the gene structure. In general, the *Notch* locus is one of the most fully characterized *D. melanogaster* genes. It has long been the subject of molecular and cytogenetic studies.

In accordance with the function of the gene, the mutations associated with this locus are manifested in disorders emerging during neuroectoderm differentiation in organs of neuroectodermal origin. Phenotypically, they include the eye, wing, gonad, and bristle aberrations. However, the presence of several independent functions of a gene at different developmental stages makes it difficult to analyze its mutations by studying their phenotypic consequences (Brennan et al., 1997).

Along with mutations affecting the structural part of the *Notch* gene, a lot of them being associated with EGF repeats, there are mutations in introns and the regulatory region of the gene that affect the fly development. The described mutations in the gene introns are associated with mobile element insertions. They include, for example, mutations of the *facet* group (*fa*, fa^3, fa^g, $fa^{g\text{-}2}$, fa^{fx} and fa^{sw}) (Markopoulou, 1989), which are associated with insertions of *copia*-like mobile elements clustered on a 4-kb segment in the second intron of the gene. All mutations in this group hinder the correct formation of the facet eye. In the severest cases, the correct hexagonal pattern of facets is lost, the order and number of bristles are disturbed, and additional interstitial tissue separating the facets appears. Some mutations of the *facet* group also cause notches on the wing.

Along with insertions into the introns of the gene, a deletion in the regulatory region of the *Notch* gene is well known. The *strawberry* (fa^{swb}) phenotype of rough eyes in males is associated with a deletion of 880 bp in the 5′ regulatory region of the *Notch* gene in the immediate vicinity of the transcription initiation sites. Together with the deletion, the 3C6/3C7 interband disappears from the salivary gland polytene chromosome cytological preparations, and the disappearance is accompanied by the replacement of the 3C6 band with a longer band, probably resulting from the fusion of the 3C6 and 3C7 bands.

The authors of works concerning the cytology of this region interpret the interband disappearance either as a consequence of deletion of all sequences that normally comprise the interband, or as the result of the disappearance of a cis-activating sequence inside the deleted sequence. This sequence is meant to be necessary for the formation of the decondensed chromatin state, and its disappearance leads to closure of a large interband sequence into a more condensed band. “It would be interesting to know whether a band and an interband accept their state actively, or one of these states is initial and is realized in the absence of sequences necessary for the other state” (Rykowski et al., 1988).

The subdivision into several structurally and functionally independent units is characteristic of eukaryotic chromosomes. The 5′ regulatory region of the *Notch* gene performs barrier functions. According to the high-resolution Hi-C data, the boundary of topological domains is mapped to this locus (Stadler et al., 2017). It also comprises the binding peaks of insulator proteins CTCF, CP190, and GAF (the *trl* gene product), which may be responsible for the insulator activity of promoters (Ohtsuki, Levine, 1998).

The most detailed work on the insulator function of the sequence directly adjacent to the promoter region of the *Notch* gene was carried out by J. Vazquez and P. Schedl (Vazquez, Schedl, 2000). The authors focused on the study of the fa^{swb} deletion effects. The conclusion that the sequence deleted by fa^{swb} normally protects the promoter of the *Notch* gene from the position effect is based on the observation of the impact of chromosome rearrangements that save the mutant phenotype. In this work, the insulator activity of the sequence deleted by fa^{swb} was checked in transgenic constructs.

First, the distribution of DNase-I hypersensitive sites (DHSs) in the promoter region of the *Notch* gene was investigated. In the regulatory region of the gene outside the promoter region, three hypersensitive sites were found. They occupied a total of about 200 nucleotide pairs. One of the sites overlapped the proximal break point of the fa^{swb} deletion, and two others

were mapped to the region removed by this deletion. It is known that the positions of DHSs correlate with the locations of regulatory DNA sequences in eukaryotes, and this fact was used for their mapping (Thomas et al., 2011).

The insulator activity of the sequence removed by the fa^{swb} deletion was tested in two regards: the ability to block enhancer-promoter interactions and the ability to protect reporter genes from the position effect (Vazquez, Schedl, 2000). The cloned sequence of 2.3 kb in length included all the transcription initiation sites of the gene, its promoter region, the sequence removed by the fa^{swb} deletion, and about 1 kb of more distal DNA sequence. As part of the *P*-transposon, it disturbed the reporter gene expression when the cloned sequence was placed between the enhancer and promoter, and it did not affect the expression when the cloned sequence was situated upstream of the enhancer. With the second approach applied in the work, the same sequence as part of a transgenic construct protected the reporter gene from the position effect. Expression of the reporter gene with the presence of the sequence from the 5′ region of the *Notch* gene was constant regardless of the genetic environment at the site of the transgenic construct insertion in the genome. With the absence of the sequence in question, the rate of the expression of the reporter gene depended on the position effect.

In all the tests described above, the removal of the region corresponding to the fa^{swb} deletion from the transgenic sequences resulted in disappearance or substantial weakening of the insulator properties of the sequence studied. Further analysis of the insulator activity of some restriction fragments of the sequence under study within transposons brought the authors to the conclusion that the insulator activity was associated with the sequence removed by the fa^{swb} deletion; specifically, a sequence including about 60 bp localized proximal to the gene might also be responsible for this activity.

Thus, in the work by J. Vazquez and P. Schedl (2000), the approach proposed thirty years ago was practically used for the first time: "Sequencing the material deleted by fa^{swb} have revealed a 47 bp repeat that has no supposed regulatory motifs. Experiments on transformation with the removal of short sequences could help determine whether a 47 bp repeat or some other part of the 880 bp sequence is necessary to determine the interband state" (Ramos et al., 1989). Here it is necessary to clarify that the tandem repeat having a complex organization of two sections, 47-bp long each, occupying a total of 97 bp is one of the characteristic features of the sequence deleted by fa^{swb}. The authors of early works devoted to the *Notch* gene regulatory region had certain expectations of it.

The obvious continuation of using this approach was the search of the *Notch* gene regulatory region for the shortest nucleotide sequence that would be necessary and sufficient for the formation of the interband chromatin state. The work began with the design of the pICon genetic construct, which allowed the studied sequences of the *Notch* gene regulatory region to be embedded into the same given position in the genome (Zimin et al., 2004). To do this, an FRT site was introduced into the construct. With the presence of FLP recombinase, every sequence studied could be integrated as part of a special donor construct into the FRT site introduced into the genome within pICon. The sequences studied in the donor construct were flanked by FRT sites and, as a result of an FRT/FLP recombination event, integrated into the genome.

However, at the second stage of the work, a 1504-bp long sequence including the region removed by the fa^{swb} deletion and more distal regions was integrated into the genome as part of the pICon construct by means of *P*-element-mediated transformation (Semeshin et al., 2008). The transforming construct was called p3C-Icon after the cytological region of the origin of the sequence under study. Thus, the potential of the site-specific transformation system was not used at this stage. Seven fly stocks were obtained, and four of them, carrying the insertion of the sequence from the regulatory region of the *Notch* gene, were examined. In all cases, the insertions of p3C-Icon were characterized by emergence of a new interband. The FRT/FLP system was then used to cut the material of the sequence under study, and in all cases the interband disappeared from the sites of p3C-Icon insertion. It was also shown that when transferred to a new genetic environment as part of the construct, the DHSs from the *Notch* 5′ regulatory region retained their properties.

At the third stage of the work, we used the FRT site integrated in the pICon(dv) genetic construct situated in the 84F region of chromosome 3 to insert the sequences from the *Notch* regulatory region. This time, an integrated FRT site provided an opportunity to test different sequences in the same genetic environment and then to cut out the integrated sequences. For this purpose, the tested sequences as part of the pFRTV vector were introduced into various sites in the *Drosophila* genome and then transferred site-specifically to the 84F region. The previously tested 1504-bp sequence was used as a basic one to design a shorter sequence of 914 bp including the whole fragment deleted by the fa^{swb} deletion. Also, a second test sequence of 1258 bp in length was obtained by removing a part of the 1504 bp sequence proximal to the *Notch* promoter. It was assumed that the DHSs found in (Vazquez, Schedl, 2000) were located in the removed part. We showed that the deletion of 246 bp led to the complete disappearance of the interband from the 84F region, whereas the presence of the 914-bp sequence completely including the fragment deleted by the fa^{swb} deletion caused the formation of an interband (Andreenkov et al., 2010). Thus, the shortest sequence required for the formation of decondensed chromatin state of the 3C6/3C7 interband was identified.

After obtaining these data, we cloned this minimal sequence into pFRTV, inserted it into the genome, and transferred to the acceptor construct pICon(dv) in the 84F region of chromosome 3 (a slightly larger 276-bp fragment was used). The presence of this fragment in the 84F region led to the appearance of an interband. It also turned out that the appearance of the interband was accompanied by the emergence of a DHS in the insertion region. Specifically, the insertion of sequences of 914 bp and 276 bp in size caused the formation of a DHS and an interband and the insertion of the 1258-bp sequence did not lead to the formation of these elements. Similarly, we cloned three tandem repeated sequences of 914 bp in length (fragment deleted by fa^{swb}) into the 84F region. As a result, an interband and three DHSs formed.

To summarize, studies of the *Notch* gene regulatory region were carried out using different approaches united by the idea of transferring individual sequences from the regula-

tory region to ectopic genetic environment. The autonomous ability of individual short fragments to cause the formation of decondensed chromatin state was shown.

However, the possibility of introducing changes directly to the sequence of the *Notch* gene regulatory region offers much greater opportunities. One of them arises with the development of directed genome editing methods. Previously, we used homologous recombination for the directional deletion of the *Drosophila RNaseZ* gene. The gene sequence was replaced by the attP site (Andreenkov et al., 2016). Such sites are used to integrate sequences into the genome by using the *ϕC31*-mediated recombination method. The attP/attB method of directed recombination permits one to embed then a sequence that will completely rescue the deletion, as well as any other artificially constructed sequence, at the attP site introduced at the deletion site.

Recently, this method has been modified. Now the approach employs the CRISPR/Cas9 system to obtain a directed deletion and insertion of the attP site, which allows more efficient construction of systems for making directed changes in a genome (Gratz et al., 2014).

We used this approach to obtain a directed 4-kilobase deletion in the *Notch* gene. With the CRISPR/Cas9 system, the regulatory and promoter zones, the first exon, and a part of the first intron of the *Notch* gene were removed and replaced by the attP site. For this purpose, we created a genetic construct that contained sequences homologous to the 3′ and 5′ sites flanking the deletion, the attP site, the *miniwhite* reporter gene, and the GMR enhancer. The *miniwhite* reporter gene of the construction was flanked by lox sites, which allowed us to remove them after inserting the attP site. In this way, a fly stock carrying the attP site and a 4-kilobase deletion in the *Notch* gene was obtained. To check the functionality of the created system, we inserted the 4-kb sequence deleted when creating an attP site into this attP site and thereby restored the functions of the gene.

Conclusion

Thus, a system that allows introducing any directed changes into the regulatory region of the *Notch* gene was first created. It will be used to obtain directed deletions, to replace functionally significant sites, and to introduce single nucleotide substitutions. Now we have an opportunity to accurately reproduce the natural fa^{swb} deletion and to remove the minimal sequence found in the regulatory region of the *Notch* gene, which has the properties of interband formation when transferred as part of a transposon to an ectopic environment. The system allows removing various *Notch* gene transcription initiation sites and analyzing their influence on gene activity. It can also be used to replace the binding site of the CTCF and CP190 insulator proteins with an alternative BEAF-32 binding site. It will allow us to make inferences as to the functionality of various insulator proteins. We conclude that a fundamentally new and powerful tool has been developed and it can be applied to studies of the *Notch* gene regulation and the organization of the open chromatin state.

References

Andreenkov O.V., Volkova E.I., Demakov S.A., Semeshin V.F., Zhimulev I.F. The decompact state of interchromomeric chromatin from the 3C6/C7 region of *Drosophila melanogaster* is determined by short DNA sequence. Doklady Biochemistry and Biophysics. 2010; 431:57-59. DOI 10.1134/S1607672910020018.

Andreenkov O.V., Volkova E.I., Demakov S.A., Xie X., Dubrovsky E.B., Zhimulev I.F. Targeted mutagenesis of *Drosophila RNaseZ* gene by homologous recombination. Doklady Biochemistry and Biophysics. 2016;471(1):399-402. DOI 10.1134/S1607672916060065.

Artavanis-Tsakonas S., Rand M.D., Lake R.J. Notch signaling: cell fate control and signal integration in development. Science. 1999;284: 770-776. DOI 10.1126/science.284.5415.770.

Brennan K., Tateson R., Lewis K., Martinez Arias A. A functional analysis of *Notch* mutations in Drosophila. Genetics. 1997;147:177-188.

Gratz S.J., Ukken F.P., Rubinstein C.D., Thiede G., Donohue L.K., Cummings A.M., O'Connor-Giles K.M. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. Genetics. 2014;196:961-971. DOI 10.1534/genetics.113.160713.

Greenwald I. *Notch* and the awesome power of genetics. Genetics. 2012;191:655-669. DOI 10.1534/genetics.112.141812.

Markopoulou K. Phenotypic and molecular analysis of the facet, a group of intronic mutations at the Notch locus of *Drosophila melanogaster* which affect postembryonic development. Genetics. 1989; 122:417-428.

Ohtsuki S., Levine M. GAGA mediates the enhancer blocking activity of the *eve* promoter in the *Drosophila* embryo. Genes Dev. 1998; 12:3325-3330.

Ramos R.G.P., Grimwade B.G., Wharton K., Scottgale T.N., Artavanis-Tsakonas S. Physical and functional definition of the *Drosophila Notch* locus by *P* element transformation. Genetics. 1989;123:337-348.

Rykowski M.C., Parmelee S.J., Agard D.A., Sedat J.W. Precise determination of the molecular limits of a polytene chromosome band: regulatory sequences for the *Notch* gene are in the interband. Cell. 1988;54:461-472.

Semeshin V.F., Demakov S.A., Shloma V.V., Vatolina T.Y., Gorchakov A.A., Zhimulev I.F. Interbands behave as decompacted autonomous units in *Drosophila melanogaster* polytene chromosomes. Genetica. 2008;132:267-279. DOI 10.1007/s10709-007-9170-5.

Stadler M.R., Haines J.E., Eisen M.B. Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early *Drosophila melanogaster* embryo. ELife. 2017;6:e29550. DOI 10.7554/eLife.29550.

Thomas S., Li X.-Y., Sabo P.J., Sandstrom R., Thurman R.E., Canfield T.K., Giste E., Fisher W., Hammonds A., Celniker S.E., Biggin M.D., Stamatoyannopoulos J.A. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. Genome Biology. 2011;12:R43. DOI 10.1186/gb-2011-12-5-r43.

Vazquez J., Schedl P. Deletion of an insulator element by the mutation *facet-strawberry* in *Drosophila melanogaster*. Genetics. 2000;155: 1297-1311.

Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S. Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats. Cell. 1985;43:567-581.

Zimin P.I., Gortchakov A.A., Demakov S.A., Zhimulev I.F. A new construct for cloning DNA and modeling the structure of *Drosophila melanogaster* polytene chromosomes. Mol. Biology. 2004;38(2): 205-209.

Acknowledgements. This work is supported by the Russian Science Foundation, project 14-14-00934.

Conflict of interest. The authors declare no conflict of interest.

Received November 24, 2018. Revised December 17, 2018. Accepted December 21, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):203-211
DOI 10.18699/VJ19.483

Generation of barcoded plasmid libraries for massively parallel analysis of chromatin position effects

M.O. Lebedev[1, 2], L.A. Yarinich[1, 2], A.V. Ivankin[1], A.V. Pindyurin[1, 2]

[1] Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
e-mail: a.pindyurin@mcb.nsc.ru

The discovery of the position effect variegation phenomenon and the subsequent comprehensive analysis of its molecular mechanisms led to understanding that the local chromatin composition has a dramatic effect on gene activity. To study this effect in a high-throughput mode and at the genome-wide level, the Thousands of Reporters Integrated in Parallel (TRIP) approach based on the usage of barcoded reporter gene constructs was recently developed. Here we describe the construction and quality checks of high-diversity barcoded plasmid libraries supposed to be used for high-throughput analysis of chromatin position effects in *Drosophila* cells. First, we highlight the critical parameters that should be considered in the generation of barcoded plasmid libraries and introduce a simple method to assess the diversity of random sequences (barcodes) of synthetic oligonucleotides using PCR amplification followed by Sanger sequencing. Second, we compare the conventional restriction-ligation method with the Gibson assembly approach for cloning barcodes into the same plasmid vector. Third, we provide optimized parameters for the construction of barcoded plasmid libraries, such as the vector : insert ratio in the Gibson assembly reaction and the voltage used for electroporation of bacterial cells with ligation products. We also compare different approaches to check the quality of barcoded plasmid libraries. Finally, we briefly describe alternative approaches that can be used for the generation of such libraries. Importantly, all improvements and modifications of the techniques described here can be applied to a wide range of experiments involving barcoded plasmid libraries.
Key words: barcode; reporter construct; plasmid library; DNA cloning; Gibson assembly; massively parallel analysis; regulation of gene expression; regulatory DNA elements; chromatin position effects; cultured *Drosophila* cells.

For citation: Lebedev M.O., Yarinich L.A., Ivankin A.V., Pindyurin A.V. Generation of barcoded plasmid libraries for massively parallel analysis of chromatin position effects Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):203-211. DOI 10.18699/VJ19.483.

Создание штрихкодированных плазмидных библиотек для множественного одновременного анализа эффекта положения гена

М.О. Лебедев[1, 2], Л.А. Яринич[1, 2], А.В. Иванкин[1], А.В. Пиндюрин[1, 2]

[1] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
e-mail: a.pindyurin@mcb.nsc.ru

Открытие явления мозаичного эффекта положения гена и последующий тщательный анализ его молекулярных механизмов привели к пониманию того, что локальный состав хроматина оказывает существенное влияние на активность генов. Разработанный недавно метод Thousands of Reporters Integrated in Parallel (TRIP) основан на использовании штрихкодированных генов-репортеров и позволяет выполнять высокопроизводительный анализ эффекта положения гена в масштабе всего генома. В настоящей работе мы описываем конструирование и проверку качества штрихкодированных плазмидных библиотек высокой сложности, которые предполагается использовать для высокопроизводительного анализа эффекта положения гена в клетках дрозофилы. Во-первых, рассмотрены наиболее важные параметры, которые следует учитывать при создании штрихкодированных плазмидных библиотек, и предложен простой метод оценки сложности вырожденного фрагмента (штрихкода) в синтетических олигонуклеотидах с использованием ПЦР-амплификации и последующего секвенирования по методу Сэнгера. Далее проведено сравнение традиционного метода клонирования посредством рестрикции–лигирования с подходом сборки по методу Гибсона для клонирования штрихкодов в один и тот же плазмидный вектор. Кроме того, описаны оптимизированные параметры для создания штрихкодированных плазмидных библиотек, такие как соотношение вектор : вставка в реакции сборки методом Гибсона и напряжение, используемое для электропорации бактериальных клеток продуктами лигирования. Сравниваются также различные подходы для проверки качества штрихкодированных плазмидных библиотек. Наконец, кратко описаны альтернативные подходы,

M.O. Lebedev and L.A. Yarinich contributed equally to this work.

которые можно использовать для создания таких библиотек. Важно отметить, что все улучшения и модификации методов, приведенные в данной работе, могут быть применены к широкому кругу экспериментов, в которых используются штрихкодированные плазмидные библиотеки.
Ключевые слова: штрихкод; репортерная конструкция; плазмидная библиотека; клонирование ДНК; сборка методом Гибсона; множественный одновременный анализ; регуляция экспрессии генов; регуляторные элементы ДНК; эффект положения гена; культивируемые клетки дрозофилы.

Introduction

In a eukaryotic cell, an enormous number of specialized proteins are responsible for dense packaging of very long genomic DNA molecules into a nucleus. These proteins also play a crucial role in the maintenance and replication of the genome as well as in gene expression. These DNA-protein complexes are referred to as chromatin, which is classically subdivided into loosely packed, transcriptionally active euchromatin and more tightly packed, transcriptionally inactive heterochromatin (Babu, Verma, 1987; Grewal, Moazed, 2003; Huisinga et al., 2006). Recent genome-wide localization studies of chromatin components revealed up to 15 principal chromatin types depending on the nature of cells. These chromatin types differ in composition and gene activity levels. Furthermore, the activity of genes greatly varies even within the same chromatin type (Filion et al., 2010; Ernst et al., 2011; Kharchenko et al., 2011; Riddle et al., 2011).

To exclude the input of unique DNA regulatory elements associated with each gene and thus to identify the pure influence of the local chromatin environment on gene activity, it was proposed to insert the same reporter construct (as a sensor) in different genomic loci. The idea of this assay was based on the position effect variegation (PEV) phenomenon, originally described in *Drosophila* (Elgin, Reuter, 2013). Such analysis performed in cells of different species substantially contributed to our understanding of the mechanisms controlling gene activity throughout the genome, although only up to a few hundred/thousand of unique genomic sites were tested in each study due to the laborious and time-consuming nature of the assay (Gierman et al., 2007; Babenko et al., 2010; Ruf et al., 2011; Chen et al., 2013). Subsequently, usage of unique identifiers (barcodes, or tags, or indexes, or unique identification DNA sequences) (Kinde et al., 2011; Blundell, Levy, 2014; Quail et al., 2014; Vvedenskaya et al., 2015; Kebschull, Zador, 2018) for labeling the reporters enabled the development of a high-throughput mode of the analysis. The modified method provided a possibility of studying thousands integrated into the genome reporters in one relatively simple and short experiment and eliminated the bias towards selection of reporter insertions in active chromatin regions. The approach was named Thousands of Reporters Integrated in Parallel (TRIP), and its pilot application to mouse embryonic stem cells (mESCs) allowed characterization of chromatin position effects at more than 27,000 distinct genomic loci (Akhtar et al., 2013, 2014).

From a technical point of view, TRIP is a specialized variant of massively parallel reporter assays (Patwardhan et al., 2009, 2012; Melnikov et al., 2012), so its performance relies on barcoded plasmid libraries. Such libraries consist of reporter plasmid constructs, which are identical except random and *a priori* unknown barcode sequences. Within a library, all barcodes are typically of the same length, which can vary from few bp up to several dozen bp depending on the experimental setup. Thus, as soon as the plasmid vector carrying all necessary components except the barcode (e. g., functional transposon terminal repeats, a promoter, a coding DNA sequence, a transcriptional terminator, etc.) is available, the next technical task is to add the barcodes into such plasmid. Several different cloning strategies can be used for this purpose, but the easiest way to get a DNA fragment carrying barcodes (hereafter barcoded fragment or insert) is to PCR-amplify it using synthetic oligonucleotides containing a fragment of random sequence (the barcode).

Theoretically, the maximum complexity of a barcoded plasmid library or, in other words, the total number of unique molecules or clones in the library is determined by the barcode length and its nucleotide composition. More specifically, this parameter depends on whether all four nucleotides or just a subset of them are used at each position within the barcode. For example, a random 4-nucleotide 18-bp sequence can generate over $\sim 6.8 \times 10^{10}$ (4^{18}) different barcodes, while a random 3-nucleotide 20-bp sequence can give rise to only $\sim 3.5 \times 10^{9}$ (3^{20}) unique sequences. Although such high numbers are never achieved in practice due to the necessity to propagate plasmids in bacterial cells that typically limits the complexity to 10^6–10^7 clones, the diversity of synthetic barcoded oligonucleotides is nevertheless crucial for the construction of high-quality barcoded plasmid libraries. Indeed, a strong overrepresentation of one or two particular nucleotides within the barcode in synthetic oligonucleotides extremely reduces the complexity of the final plasmid library. In addition, it leads to the appearance of very similar barcode sequences, which can substantially complicate unambiguous identification of barcodes during the processing and analysis of raw TRIP datasets generated by high-throughput DNA sequencing (HTS). Furthermore, note that barcoded plasmid libraries frequently contain some amounts of the original plasmid vectors due to limitations of cloning techniques. Such contamination should be kept at the lowest possible level or, ideally, completely eliminated to increase the proportion of useful reads in raw TRIP datasets. All these parameters, the total number of unique barcodes, the randomness and diversity of their sequences, and the level of the original vector contamination, determine the quality of barcoded plasmid libraries.

We aimed to find out the simplest and most efficient way to construct high-quality barcoded plasmid libraries for subsequent TRIP experiments. In doing this, we first developed a fast and simple test to check whether the quality of synthetic oligonucleotides containing a barcode meets the expectations before the generation of plasmid libraries. Next, we found that the Gibson assembly approach was superior over the conventional restriction-ligation cloning for the construction of barcoded plasmid libraries in terms of both the total number of unique clones obtained per the same amount of the vector used and the level of original vector contamination. Finally, we provide a set of quality checks of barcoded

plasmid libraries to verify that they are suitable for subsequent TRIP experiments.

Materials and methods

Barcoded oligonucleotide primers and quality check of the randomness of their barcode sequence. Synthetic oligonucleotides containing 18-nt barcodes, within which all four nucleotides (A, C, G, and T) were supposed to be equally represented at each position, were ordered either from IDT (USA) with standard desalting purification (the PB-Barcode1-SalI-F primer) or from DNK-Sintez (Moscow, Russia) with PAGE purification (the PB-Barcode1-SalI-F primer) or from Biosynthesis (Novosibirsk region, Russia) with PAGE purification (the PB-Barcode-PI-X-Gibson-F primers). Barcoded PCR products were amplified with the following combinations of the forward barcoded/reverse primers – PB-Barcode1-SalI-F/3′-TR-EagI-R or PB-Barcode-PI-X-Gibson-F/3′-TR-EagI-R (sequences of all primers used in the study are listed in the Suppl. Table S1)[1] – using plasmids with the appropriate library indexes (Suppl. Fig. S1) as templates. Five microliters of the PCR products were incubated with Exonuclease I (NEB, Cat. No. M0293S) and Shrimp Alkaline Phosphatase (NEB, Cat. No. M0371S) at 37 °C for 30 min to remove primers and nucleotides. The treated products were subjected to direct Sanger sequencing with the 3′-TR-EagI-R primer.

Preparation of PCR products for construction of barcoded plasmid libraries. All DNA fragments used in the ligation or Gibson assembly reactions were amplified using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Cat. No. F530L) according to manufacturer's recommendations at the following cycling conditions: 98 °C for 60 sec followed by 35 cycles with 98 °C for 30 sec, 58 °C for 30 sec, 72 °C for 1 min and a final extension step of 72 °C for 5 min. Amplified fragments were then column-purified with GeneJET PCR Purification Kit (Thermo Fisher Scientific, Cat. No. K0702) before use in subsequent reactions.

Construction of barcoded plasmid libraries by restriction enzyme cloning. To generate the barcoded plasmid libraries, the 3′-terminal repeat (TR) of the piggyBac transposon was amplified with primers PB-Barcode1-SalI-F and 3′-TR-EagI-R using the pPB-Promoter1-eGFP-LI plasmid as a template. The PCR product was digested with SalI-HF (NEB, Cat. No. R3138L) and EagI-HF (NEB, Cat. No. R3505L). Then 0.9 pmol (240 ng) of PCR product were ligated with 0.3 pmol (1000 ng) of pPB-Promoter1-eGFP-LI vector digested with SalI-HF and NotI (Thermo Fisher Scientific, Cat. No. FD0593) restriction enzymes, using T4 DNA ligase (Roche, Cat. No. 10799009001). The ligation mixture was digested with NotI restriction enzyme at 37 °C for 4 h in a volume of 100 μL to destroy the original pPB-Promoter1-eGFP-LI vector molecules and then purified with MinElute PCR Purification Kit (Qiagen, Cat. No. 28004); DNA was eluted with 10 μL of pre-warmed elution buffer. Five microliters of the purified ligation mixture were used to transform *Escherichia coli* TOP10 electrocompetent cells.

Gibson assembly of barcoded plasmid libraries. pPB-PromoterX-eGFP-LI vectors (where X is 0 through 6) were digested with SalI (Thermo Fisher Scientific, Cat. No. FD0644) and NotI (Thermo Fisher Scientific, Cat. No. FD0593) restriction enzymes to prepare linearized vector backbones. The 3′-TR of the piggyBac transposon was amplified with primers PB-Barcode-PI-X-Gibson-F and PB-Gibson-R1 using a pPB-Promoter0-eGFP-LI plasmid as a template. The PCR products were digested with DpnI restriction enzyme (NEB, Cat. No. R0176) at 37 °C for 4 h in a total volume of 100 μL in order to destroy plasmid template and thus minimize non-barcoded vector contamination in the libraries. Next, 60 fmol (200 ng) of the linearized vector and 0.30 pmol (~80 ng) of each DpnI-treated and column-purified PCR product were incubated with Gibson Assembly Master Mix (NEB, Cat. No. M5510AA) at 50 °C for 1 h. Then, the reaction mixture was diluted 10-fold with nuclease-free water and purified with MinElute PCR Purification Kit; DNA was eluted with 10 μL of pre-warmed elution buffer. Five microliters of the purified Gibson reaction products were used to transform *E. coli* TOP10 electrocompetent cells.

Electrotransformation of bacterial cells. *E. coli* TOP10 strain (F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac* ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G λ$^-$) was used as the host to clone plasmid libraries. Electrocompetent cells with transformation efficiency of ~1.5×10^9 colony forming units (CFU) per 1 μg of supercoiled plasmid DNA were prepared using the previously reported protocol (Morrison, 2001). For transformation, 50 μL of the electrocompetent cells were mixed with 5 μL of purified DNA sample, transferred into a 0.1 cm cuvette (Bio-Rad), and electroporated at the following settings: 1.8 kV, 25 mF, and 200 Ω using the Gene Pulser electroporator with the capacitance extender (Bio-Rad) unless otherwise stated. Next, cells were grown in 1 mL of pre-warmed Luria–Bertani (LB) medium at 37 °C, 220 rpm for 1 h. Then, aliquots of cells (1/5,000) were spread on LB-ampicillin plates (LB medium supplemented with 15 g/L bacto agar and 100 μg/mL ampicillin) and allowed to grow overnight at 37 °C before manual counting of colonies, whereas the rest of cells were transferred into 500 mL of LB medium supplemented with 100 μg/mL ampicillin and cultured at 37 °C and 220 rpm overnight before plasmid isolation with GeneJET Plasmid Maxiprep Kit (Thermo Fisher Scientific, Cat. No. K0491).

Analysis of the complexity and quality of barcoded plasmid libraries. The complexity of each plasmid library was defined as the average number of colonies on LB-ampicillin plates (CFU per plate; see above) multiplied by 5,000. To measure the contamination of each library by the original plasmid vector, the appropriate control ligation or Gibson assembly reaction lacking the barcoded insert DNA fragment was always set up and processed in parallel with the main samples, except the plasmid isolation step. The contamination is reported as the proportion (percentage) of the number of colonies obtained for the control sample with respect to the number of colonies obtained for the main samples. The DNA sequences of barcodes in the plasmid libraries and randomly selected individual clones were analyzed by Sanger sequencing with the EGFP-int3 primer.

Preparation of barcoded plasmid libraries for Illumina sequencing. An equimolar mixture of barcoded plasmid libraries was used to transform *E. coli* TOP10 cells in order to obtain several thousand clones. The isolated mixture of such

[1] Supplementary Materials are available in the online version of the paper: http://www.bionet.nsc.ru/vogis/download/pict-2019-23/appx4.pdf

low-complexity plasmid libraries was used as a template for amplification of barcode sequences for their subsequent Illumina HTS. The amplification was done in duplicate by two rounds of PCR with Phusion High-Fidelity DNA Polymerase and 200 μM of each dNTP in a total volume of 25 μL. In the first round, 2 ng of the plasmid mixture and primers Libr-cDNA-for/Libr-cDNA-AN-rev (where N was 18 or 19) were used at the following cycling conditions: 98 °C for 60 sec followed by 15 cycles with 98 °C for 30 sec, 70 °C for 30 sec, 72 °C for 30 sec and a final extension step of 72 °C for 5 min. In this round, each of the duplicates was marked by a unique HTS index present within the Libr-cDNA-AN-rev primer. In the second round, 0.5 μL of the unpurified first round PCR products and primers Libr-P5-for and Libr-P7-rev were used at the following cycling conditions: 98 °C for 60 sec followed by 23 cycles with 98 °C for 30 sec, 61 °C for 30 sec, 72 °C for 30 sec and a final extension step at 72 °C for 5 min. The samples were purified with Monarch PCR & DNA Cleanup Kit (NEB, Cat. No. T1030S), quantified using the KAPA Library Quantification Kit (Roche, Cat. No. 07960255001), and sequenced on an Illumina MiSeq sequencer using a 150 cycle MiSeq Reagent Kit v3 (Illumina, Cat. Nos. 15043893 and 15043894). The expected structure of the DNA fragments subjected to HTS is shown in Suppl. Fig. S2.

HTS data analysis. The FASTQ files were analyzed with custom Python scripts. First, only reads having the expected structure (see Suppl. Fig. S2) were filtered for the downstream analysis. Briefly, no mismatches were allowed within sequences of HTS and library indexes, while up to 4 mismatches were allowed within the first constant part (CGCC AGGGTTTTCCCAGTCACAAGGGCCGGCCACAACTC) and 1 mismatch was allowed within the second constant part (CTCGATC). The barcodes were defined as sequences of any length present between (i) the first constant part and (ii) the sequence consisting of the CCTC motif, the library index, and the second constant part. Next, the reverse complement sequences of barcodes present in the filtered reads were divided into individual pools according to the associated HTS and library indexes. For each pool, unique barcode sequences were identified and subjected to length distribution analysis. To analyze the frequency distribution of the four nucleotides per position within barcodes, only unique 18-nt barcodes were considered.

Results and discussion

Quality control of barcoded oligonucleotides

The main problems associated with preparations of long (>30 nt) synthetic barcoded oligonucleotides, which are the key element in the cloning of barcoded plasmid libraries, are (i) single-base deletions/insertions arising from incomplete coupling of nucleotide monomers during oligonucleotide synthesis and (ii) unequal representation of different nucleotides at each position along the barcode. PAGE purification of the synthesized oligonucleotides can partially reduce the first problem, but, to the best of our knowledge, no fast, simple, and cheap approach to assess the randomness of barcode sequences in an oligonucleotide preparation has been described so far. Instead, Sanger sequencing of plasmids isolated from dozens of randomly selected bacterial colonies, as well as

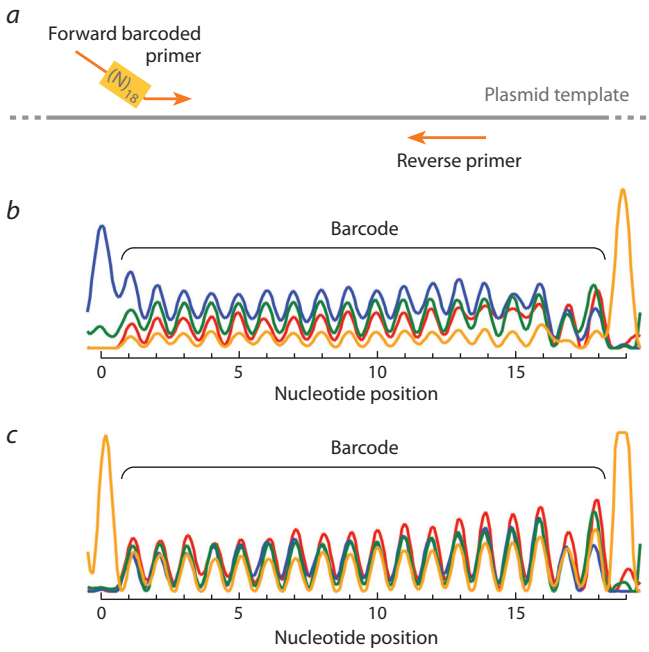


Fig. 1. The fast and simple approach to assess the diversity of barcoded oligonucleotides.

(*a*) Scheme of the DNA template and oligonucleotide primers used to prepare the barcoded PCR product. Note that the forward primer contains a random sequence of 18-nt long, the barcode. The obtained PCR product is purified and Sanger sequenced from the reverse primer. (*b*, *c*) Sanger sequencing chromatograms of the representative PCR products generated using (*b*) low-quality and (*c*) high-quality barcoded primers.

Sanger or HTS of barcoded plasmid libraries were used in previous studies (Akhtar et al., 2013; Wong et al., 2016). In addition to being laborious, these approaches require a ready barcoded plasmid library.

In this study, we developed a simple and fast protocol to check the randomness of barcode sequences within synthesized primers by using Sanger sequencing prior to the generation of plasmid libraries. The procedure includes amplification of the barcoded PCR product from a non-barcoded plasmid template followed by enzymatic removal of nucleotides and primers from the PCR mixture by Exo-SAP (Exonuclease I with Shrimp Alkaline Phosphatase) treatment, and direct Sanger sequencing of the purified product (Fig. 1). If the barcode is designed to consist of the four nucleotides equally represented at each position, then in the case of ideally synthesized oligonucleotides, the height of Sanger sequencing chromatogram peaks for each of nucleotides along the barcode should be equal with only slight variations.

Generation of barcoded plasmid libraries

Many approaches have been developed for cloning or assembly of barcoded plasmid libraries, including traditional cloning by restriction enzyme digestion, self-ligation of PCR products, polymerase cycling assembly (PCA), dA/dT ligation, Gibson assembly and others (see Table 1 and references therein). All these approaches consist of several steps: (i) preparation of a plasmid vector and barcoded insert, (ii) ligation of the vector and insert, (iii) highly efficient transformation of bacterial cells, (iv) plasmid propagation in bacteria and its subsequent

Table 1. Common methods to construct barcoded plasmid libraries

Method	Advantages "+"/Disadvantages "–"	Requirements	Efficiency*	References
Cloning by restriction enzyme digestion and ligation	"+" (1) Simple primer design; (2) Cost effective "–" (1) Restriction site-depending; (2) Time-consuming; (3) Assembly of multiple fragments is complicated/ inefficient	Restriction endonucleases, High-fidelity DNA polymerase, Antarctic phosphatase, T4 DNA ligase, DNA purification columns	1.5×10^3	Akhtar et al., 2014
Self-ligation of barcoded PCR product	"+" (1) Simple primer design; (2) Cost effective "–" (1) PCR-induced mutations in the vector sequence; (2) Not applicable for large DNA molecules (>8 kb); (3) Time-consuming; (4) Formation of concatemers	Phosphorylated oligonucleotide, High-fidelity DNA polymerase, T4 DNA polymerase, DpnI restriction enzyme, T4 DNA ligase, DNA purification columns	5.8×10^7	Brueckner et al., 2016; Chen et al., 2017
dA/dT ligation	"+" (1) Simple primer design; (2) Sequence-independent "–" (1) Nondirectional; (2) Expensive; (3) Time-consuming	T-tailed vector, DNA polymerase, T4 DNA ligase, DNA purification columns	20	Seguin-Orlando et al., 2013; Davidsson et al., 2016
Golden Gate assembly	"+" (1) Multiple fragments assembly; (2) Single step; (3) Possible transfer of a DNA fragment library from one vector to another; (4) Fast; (5) Tolerant to repetitive sequences "–" (1) Complicated primer design; (2) Restriction site-depended; (3) Expensive	Nonpalindromic restriction endonucleases, High-fidelity DNA polymerase, Special assembly kit, DNA purification columns	2.7×10^4	Woodruff et al., 2017
Gateway cloning	"+" (1) Single step; (2) Possible transfer of a DNA fragment library from one vector to another; (3) Fast "–" (1) Complicated primer design; (2) Sequence-dependent; (3) Necessity to use specific plasmid backbones; (4) Expensive; (5) Scar att regions	Special assembly kit, High-fidelity DNA polymerase, DNA purification columns	2.0×10^4	Dickel et al., 2014; Lehtonen et al., 2015; Davidsson et al., 2016
Gibson assembly	"+" (1) Single step; (2) Multiple fragments assembly; (3) Sequence-independent; (4) Seamless; (5) Fast "–" (1) Complicated primer design; (2) Self-ligation of linearized vector backbone, if only one restriction site is used for the linearization; (3) Expensive; (4) Not tolerant to repetitive sequences	Special assembly kit, Restriction endonucleases, DNA purification columns	6.9×10^3	Hill et al., 2018

* Efficiency is defined as the number of unique bacterial clones obtained per 1 ng of vector DNA used in an assembly/ligation reaction.

isolation, and (v) assessment of the plasmid library quality. Each approach has its own advantages and disadvantages (see Table 1 and Suppl. Table S2).

For example, cloning by restriction enzyme digestion and ligation is simple and widely used, but it has the following limitations. First, since restriction enzyme recognition sites are short sequences, they are certainly present in a subset of randomly synthesized barcodes. During the cloning, such barcodes will be cut into pieces, thus leading to generation of cloned barcodes of undesired shorter lengths. Accordingly, the number of plasmid molecules with unique barcode sequences of the expected length will be reduced. The longer is the barcode, the higher is the probability of restriction site occurrence in its sequence. For example, a random 18-nt barcode is expected to contain one of two 6-bp restriction sites used for cloning at a frequency of ~1/160. Second, restriction enzymes are not hundred-percent effective, and some of them are extremely ineffective in cutting their recognition sites located close to the ends of linear DNA fragments (in particular, typical PCR products). This results in reduced ligation efficiency, whose compensation requires large amounts of the prepared vector (~1 μg) and insert (~0.2 μg) at the ligation step.

The aim of the present study was to make a set of barcoded plasmid libraries for high-throughput analysis of chromatin position effects (Elgin, Reuter, 2013; Yankulov, 2013) on different types of promoter sequences in *Drosophila* cells. Since our ultimate goal requires integration of barcoded reporter constructs at randomly chosen genomic loci in the studied cells, we employed the piggyBac transposon-based gene delivery system (Handler, Harrell, 1999; Cadinanos, Bradley, 2007), which was successfully used for the same purpose in cultured mouse embryonic stem cells (mESCs) (Akhtar et al., 2013). We selected six different *Drosophila* promoters (here referred to as Promoter1 through 6; their details will be described somewhere else), and cloned each of them upstream of the coding sequence for enhanced green fluorescent protein (eGFP) placed between the piggyBac TRs. The constructs also contained polyadenylation signals immediately before and within the piggyBac 3′-TR. As the control, we used a similar construct without any promoter element (hereafter Promoter0; see Suppl. Fig. S1). Next, two unique 5-bp long library indexes were separately cloned downstream of the eGFP coding sequence into each of the aforementioned construct to generate 14 different indexed plasmids, which were

Table 2. Barcoded plasmid libraries constructed by Gibson assembly

#	Barcoded plasmid library	Library index	Efficiency*
1	pPB-Promoter0-eGFP-Libr1	TCGCT	2.2×10^4
2	pPB-Promoter0-eGFP-Libr2	CCGAG	2.6×10^3
3	pPB-Promoter1-eGFP-Libr3	TTGAG	3.0×10^3
4	pPB-Promoter1-eGFP-Libr4	GATGG	5.7×10^3
5	pPB-Promoter2-eGFP-Libr5	GGCTT	1.8×10^3
6	pPB-Promoter2-eGFP-Libr6	TGTGT	5.2×10^3
7	pPB-Promoter3-eGFP-Libr7	AGTCA	3.0×10^3
8	pPB-Promoter3-eGFP-Libr8	TGGCC	3.0×10^3
9	pPB-Promoter4-eGFP-Libr9	CTAGT	3.1×10^3
10	pPB-Promoter4-eGFP-Libr10	TTATC	4.7×10^3
11	pPB-Promoter5-eGFP-Libr11	CTGCT	4.0×10^2
12	pPB-Promoter5-eGFP-Libr12	CAATT	2.9×10^3
13	pPB-Promoter6-eGFP-Libr13	GTCAA	8.4×10^3
14	pPB-Promoter6-eGFP-Libr14	GGGAT	3.2×10^3

* Efficiency is defined as the number of unique bacterial clones obtained per 1 ng of vector DNA used in the assembly/ligation reaction.

named pPB-PromoterX-eGFP-LI vectors (for exact sequences of library indexes, see Table 2). Such indexing of plasmids allows combinations of (even all) barcoded constructs to be studied simultaneously in a single TRIP experiment.

At the final step, it was necessary to clone random barcode sequences immediately downstream of library indexes to generate barcoded plasmid libraries. We chose the length of the barcodes to be 18 bp, which meant that, theoretically, there might be up to $4^{18}\approx68$ billion unique barcode sequences. First, we tried to construct a set of barcoded plasmid libraries using traditional cloning by restriction enzyme digestion and subsequent ligation following the previously described protocol (Akhtar et al., 2014) with minor modifications. Specifically, we designed the barcoded oligonucleotide PB-Barcode1-SalI-F, which contains the SalI restriction site at its 5′ end and is suitable to construct all 14 barcoded plasmid libraries due to the presence of a unique SalI restriction site immediately downstream of library indexes in pPB-PromoterX-eGFP-LI vectors. Next, we amplified the DNA fragment containing piggyBac 3′-TR using this barcoded primer and the reverse primer 3′-TR-EagI-R containing the EagI restriction site. We digested barcoded PCR products with SalI and EagI and cloned them into the pPB-Promoter1-eGFP-LI vector cut with SalI and NotI. We used the property of the EagI and NotI restriction sites, which produce compatible cohesive ends restoring only the EagI site, to remove the original (non-barcoded) plasmid molecules from the barcoded plasmid library by NotI digestion. Overall, we found this way of barcode cloning simple, cheap, and convenient. However, we could not produce plasmid libraries with complexity of more than 30 clones per 1 ng of the digested vector due to incomplete digestion of the barcoded PCR product presumably by the SalI restriction enzyme. The protocol consisted of six steps (PCR, restriction digestion, dephosphorylation of 5′-ends, DNA ligation, treatment of the ligation mixture by restriction endonuclease, and electrotransformation of bacteria) with DNA purification after each step. As a result, using about 1000 ng of the digested vector and 10 individual electrotransformations, we obtained a plasmid library consisting of ~300,000 clones. Thus, the conventional restriction-ligation method appeared to be too laborious especially for the generation of large numbers of different barcoded plasmid libraries, while its cloning accuracy, calculated as the percentage of the number of colonies obtained for the "vector with barcoded insert" ligation with respect to the number of colonies obtained by the control "vector only" ligation, was satisfactory (about 0.6 %).

To optimize the construction of barcoded plasmid libraries, we decided to use the Gibson assembly approach, which employs three commonly used enzymes: (i) 5′ exonuclease that digests the ends of dsDNA, exposing the ssDNA, (ii) high-fidelity DNA polymerase, and (iii) T4 DNA ligase (Gibson et al., 2009). The assembly fragments were designed so that they contained at their ends overlapping sequences of 23 bp in length, which were expected to anneal at the temperature 60–65 °C (Fig. 2, *a*). We amplified the barcoded DNA fragments containing piggyBac 3′-TR using PB-Barcode-PI-X-Gibson-F/PB-Gibson-R1 primers. The primer PB-Barcode-PI-X-Gibson-F contains a random 18-bp barcode but lacks the SalI restriction site in order to allow elimination of the original plasmid molecules from the barcoded plasmid library by SalI digestion. As vector fragments, we used NotI/SalI-linearized pPB-PromoterX-eGFP-LI plasmids. Next, to enhance the efficiency of the Gibson assembly reaction, we tested a vector to insert ratios of 1:3, 1:5, and 1:10 and found that the maximum number of individual bacterial colonies was obtained with the 1:3 ratio (570,000 clones versus 470,000 and 380,000 clones for the 1:5 and 1:10 ratios, respectively). Thus, an excess of the insert reduced the Gibson assembly efficiency and consequently yielded fewer clones possibly due to competition of DNA molecules for enzymes. Furthermore, we revealed that electroporation of *E. coli* cells at 1.6 kV increased the number of bacterial colonies more than fourfold and greatly reduced arcing compared to the predefined 1.8 kV in the Gene Pulser electroporator (Bio-Rad) (see Fig. 2, *b*).

A total of 14 barcoded plasmid libraries were generated by the Gibson assembly approach (see Table 2), which consists of four steps (PCR, vector linearization, 1-h assembly reaction, and electrotransformation of bacteria) with DNA purification after each step. As a result, using about 200 ng of the linearized vector and 2 individual electrotransformations, we obtained plasmid libraries consisting of ~300,000–2,000,000 clones.

The complexity and quality of barcoded plasmid libraries

To generate populations of cells carrying thousands of uniquely barcoded reporter constructs integrated at different genomic loci, it is desirable to obtain appropriate plasmid libraries with the highest possible complexity. This would minimize the chances of integration of the identically barcoded reporter constructs in different genomic locations. Note that the presence of the same barcode at different unique genomic sites (we do not consider here integration of a reporter within repetitive elements) makes this barcode useless for the downstream analysis, since its characteristics (e. g., the expression

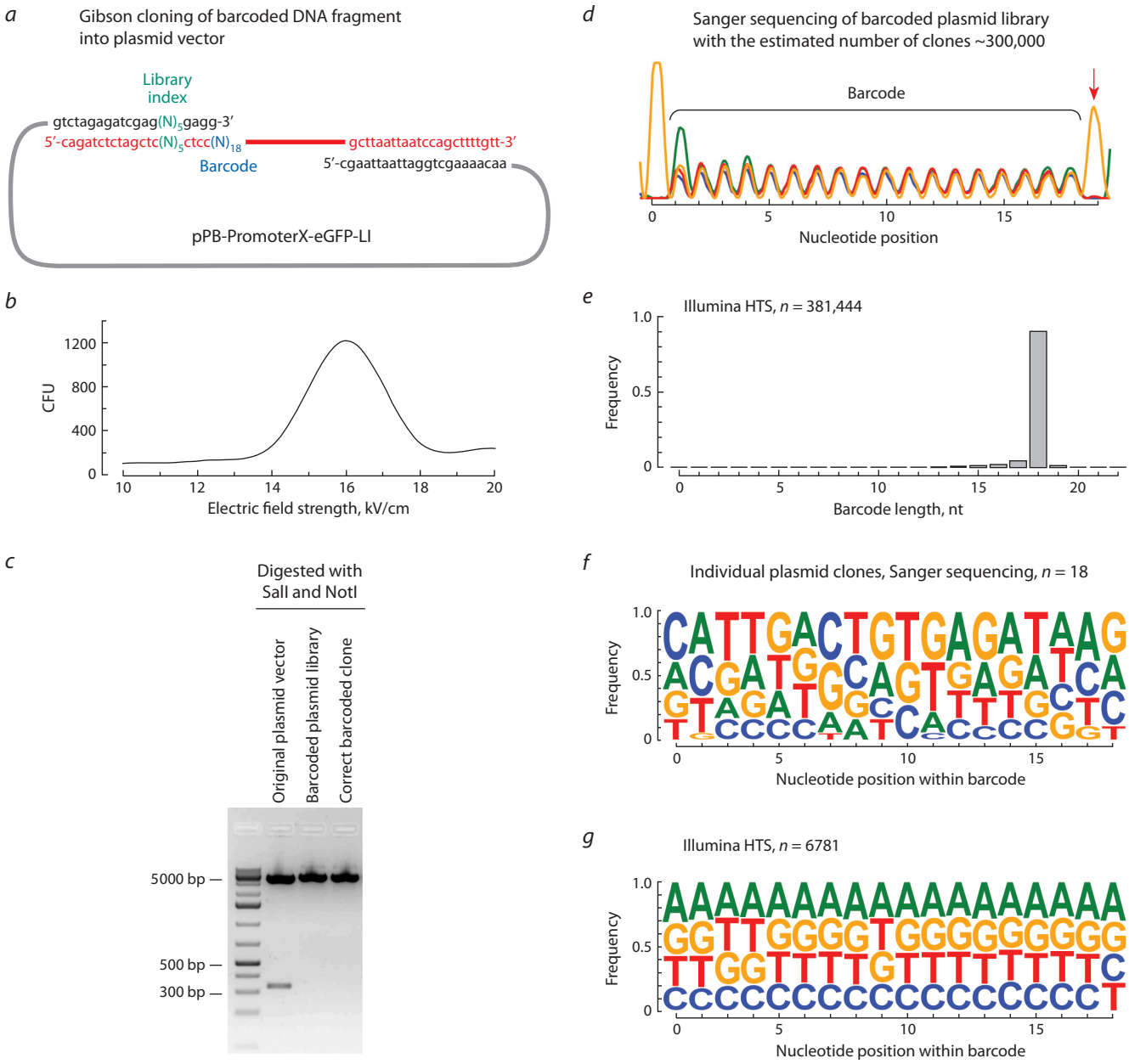


Fig. 2. Generation and quality control of barcoded plasmid libraries.

(*a*) Schematic presentation of the Gibson assembly of barcoded plasmid libraries. (*b*) Dependence of the number of bacterial colonies obtained on the electric field strength used in the electrotransformation. (*c–g*) Characterization of barcode diversity: (*c*) Agarose gel electrophoresis showing a negligible amount, if any, of the original vector contamination in the barcoded plasmid library. The original plasmid vector and correct barcoded clone serve as positive and negative controls of the presence of the original vector, respectively. (*d*) Representative Sanger sequencing chromatogram of a barcoded plasmid library generated by Gibson assembly approach. Note the presence of only guanine (orange line), but not thymine (red line) expected in the case of extensive original vector contamination at the position indicated by the red arrow. (*e*) Distribution of the barcode length revealed by HTS; *n*, number of reads analyzed. (*f*, *g*) Frequency distributions of the four nucleotides per position within the barcode revealed by (*f*) Sanger sequencing of individual plasmids and (*g*) by HTS. Weblogo 3, a web based application (http://weblogo.threeplusone.com/create.cgi), was used for visualization of the frequencies. In panel (*f*), *n* is the number of clones, and in panel (*g*), unique barcodes analyzed.

level) cannot be unambiguously associated with a genomic locus. Also, such ambiguous barcodes reduce the number of informative HTS reads used to identify the genomic locations of reporter constructs and their expression levels. In addition, barcode sequences should be easily distinguishable from each other to avoid mistakes in the interpretation of the results. Such diversity among barcode sequences can be achieved by equal representation of each nucleotide at each position along a barcode in the plasmid library (not to mention the effects of barcode length and the number of different nucleotides constituting the barcode). Moreover, it is crucial to avoid substantial levels of contamination of barcoded plasmid libraries with the original (non-barcoded) vector molecules, as the latter would also reduce the number of informative HTS reads.

To assess the quality of the barcoded plasmid libraries obtained in the present study, we performed a set of experiments.

First, to evaluate the level of contamination by the original vector, we digested one of the barcoded plasmid libraries with the NotI and SalI restriction enzymes. Recognition sites for both these enzymes were present in the original vectors but not in plasmids with properly cloned barcodes. The analysis did not reveal any substantial amount of the original vector molecules within the library (see Fig. 2, *c*). Consistently, the transformation of bacterial cells with the control "vector only" samples also showed that the level of the original vector contamination in the libraries was no higher than 0.5 %. Second, we analyzed plasmids isolated from 20 randomly selected bacterial colonies of one barcoded plasmid library by Sanger sequencing. It showed that all the 20 plasmids had the expected structure, although one plasmid carried a 17 bp long barcode and another colony appeared to contain two plasmids with different barcodes. The barcodes of the remaining 18 plasmids were of the expected length (18 bp) and demonstrated satisfactory representation of different nucleotides at each position (see Fig. 2, *f*). Third, we subjected all barcoded plasmid libraries to Sanger sequencing to assess the diversity of the barcode sequences and possible non-barcoded plasmid contamination. This analysis demonstrated almost equal representation of all four nucleotides at the most positions within the barcode and confirmed the absence of contamination by the original vector (see Fig. 2, *d*). Fourth, we performed Illumina HTS for more precise characterization of barcoded plasmid libraries. For this purpose, we mixed all prepared barcoded plasmid libraries and reduced the total number of clones by restrictive transformation of *E. coli* cells. Analysis of the HTS reads showed that (i) the length of vast majority of barcodes was 18 bp (see Fig. 2, *e*), (ii) the representation of different nucleotides along the barcodes was almost perfect (see Fig. 2, *g*), and (iii) the contamination of the barcoded plasmid libraries by the original vectors was about 0.4 % on the average.

Finally, as described above, we quantified the complexity of each barcoded plasmid library by plating a small aliquot of the transformation mixture in duplicate on LB-ampicillin plates and counting the resulting number of bacterial colonies that supposedly corresponded to the number of individual plasmid molecules. Unfortunately, we noticed that this method was not always accurate due to its dependence on bacterial concentration in liquid culture, the quality of LB-ampicillin plates, and plating techniques. Thus, alternative simple methods for the estimation of the complexity of barcoded plasmids libraries need to be developed.

Conclusions

Generation of barcoded plasmid libraries is challenging. The main difficulties are associated with the diversity of barcode sequences and their efficient cloning in a plasmid vector, including obtaining a sufficient number of individual bacterial colonies. We describe a low-cost, fast, and simple technique to check the quality of barcoded oligonucleotides by broadly used Sanger sequencing. We also present an improved method to create barcoded plasmid libraries in a single step with the commercially available Gibson assembly mixture. It is worth mentioning that the protocols developed in this study are suitable for generation of barcoded plasmid libraries for a wide range of applications, including but not limited to functional analysis of DNA regulatory elements in different species.

References

Akhtar W., de Jong J., Pindyurin A.V., Pagie L., Meuleman W., de Ridder J., Berns A., Wessels L.F., van Lohuizen M., van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell. 2013;154:914-927. DOI 10.1016/j.cell.2013.07.018.

Akhtar W., Pindyurin A.V., de Jong J., Pagie L., Ten Hoeve J., Berns A., Wessels L.F., van Steensel B., van Lohuizen M. Using TRIP for genome-wide position effect analysis in cultured cells. Nat. Protoc. 2014;9:1255-1281. DOI 10.1038/nprot.2014.072.

Babenko V.N., Makunin I.V., Brusentsova I.V., Belyaeva E.S., Maksimov D.A., Belyakin S.N., Maroy P., Vasil'eva L.A., Zhimulev I.F. Paucity and preferential suppression of transgenes in late replication domains of the *D. melanogaster* genome. BMC Genomics. 2010;11:318. DOI 10.1186/1471-2164-11-318.

Babu A., Verma R.S. Chromosome structure: euchromatin and heterochromatin. Int. Rev. Cytol. 1987;108:1-60.

Blundell J.R., Levy S.F. Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. Genomics. 2014;104:417-430. DOI 10.1038/nature14279.

Brueckner L., van Arensbergen J., Akhtar W., Pagie L., van Steensel B. High-throughput assessment of context-dependent effects of chromatin proteins. Epigenetics Chromatin. 2016;9:43. DOI 10.1186/s13072-016-0096-y.

Cadinanos J., Bradley A. Generation of an inducible and optimized piggyBac transposon system. Nucleic Acids Res. 2007;35:e87. DOI 10.1093/nar/gkm446.

Chen H.C., Martinez J.P., Zorita E., Meyerhans A., Filion G.J. Position effects influence HIV latency reversal. Nat. Struct. Mol. Biol. 2017;24:47-54. DOI 10.1038/nsmb.3328.

Chen M., Licon K., Otsuka R., Pillus L., Ideker T. Decoupling epigenetic and genetic effects through systematic analysis of gene position. Cell Rep. 2013;3:128-137. DOI 10.1016/j.celrep.2012.12.003.

Davidsson M., Diaz-Fernandez P., Schwich O.D., Torroba M., Wang G., Bjorklund T. A novel process of viral vector barcoding and library preparation enables high-diversity library generation and recombination-free paired-end sequencing. Sci. Rep. 2016;6:37563. DOI 10.1038/srep37563.

Dickel D.E., Zhu Y., Nord A.S., Wylie J.N., Akiyama J.A., Afzal V., Plajzer-Frick I., Kirkpatrick A., Gottgens B., Bruneau B.G., Visel A., Pennacchio L.A. Function-based identification of mammalian enhancers using site-specific integration. Nat. Methods. 2014;11:566-571. DOI 10.1038/nmeth.2886.

Elgin S.C., Reuter G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. Cold Spring Harb. Perspect. Biol. 2013;5:a017780. DOI 10.1101/cshperspect.a017780.

Ernst J., Kheradpour P., Mikkelsen T.S., Shoresh N., Ward L.D., Epstein C.B., Zhang X., Wang L., Issner R., Coyne M., Ku M., Durham T., Kellis M., Bernstein B.E. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473:43-49. DOI 10.1038/nature09906.

Filion G.J., van Bemmel 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;143:212-224. DOI 10.1016/j.cell.2010.09.009.

Gibson D.G., Young L., Chuang R.Y., Venter J.C., Hutchison C.A. 3rd, Smith H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods. 2009;6:343-345. DOI 10.1038/nmeth.1318.

Gierman H.J., Indemans M.H., Koster J., Goetze S., Seppen J., Geerts D., van Driel R., Versteeg R. Domain-wide regulation of gene expression in the human genome. Genome Res. 2007;17:1286-1295. DOI 10.1101/gr.6276007.

Grewal S.I., Moazed D. Heterochromatin and epigenetic control of gene expression. Science. 2003;301:798-802. DOI 10.1126/science.1086887.

Handler A.M., Harrell R.A. 2nd. Germline transformation of *Drosophila melanogaster* with the piggyBac transposon vector. Insect Mol. Biol. 1999;8:449-457. DOI 10.1046/j.1365-2583.1999.00139.x.

Hill A.J., McFaline-Figueroa J.L., Starita L.M., Gasperini M.J., Matreyek K.A., Packer J., Jackson D., Shendure J., Trapnell C. On the design of CRISPR-based single-cell molecular screens. Nat. Methods. 2018;15:271-274. DOI 10.1038/nmeth.4604.

Huisinga K.L., Brower-Toland B., Elgin S.C. The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma. 2006;115:110-122. DOI 10.1007/s00412-006-0052-x.

Kebschull J.M., Zador A.M. Cellular barcoding: lineage tracing, screening and beyond. Nat. Methods. 2018;15:871-879. DOI 10.1038/s41592-018-0185-x.

Kharchenko P.V., Alekseyenko A.A., Schwartz Y.B., Minoda A., Riddle N.C., Ernst J., Sabo P.J., Larschan E., Gorchakov A.A., Gu T., Linder-Basso D., Plachetka A., Shanower G., Tolstorukov M.Y., Luquette L.J., Xi R., Jung Y.L., Park R.W., Bishop E.P., Canfield T.K., Sandstrom R., Thurman R.E., MacAlpine D.M., Stamatoyannopoulos J.A., Kellis M., Elgin S.C., Kuroda M.I., Pirrotta V., Karpen G.H., Park P.J. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. Nature. 2011;471:480-485. DOI 10.1038/nature09725.

Kinde I., Wu J., Papadopoulos N., Kinzler K.W., Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc. Natl. Acad. Sci. USA. 2011;108:9530-9535. DOI 10.1073/pnas.1105422108.

Lehtonen S.I., Taskinen B., Ojala E., Kukkurainen S., Rahikainen R., Riihimaki T.A., Laitinen O.H., Kulomaa M.S., Hytonen V.P. Efficient preparation of shuffled DNA libraries through recombination (Gateway) cloning. Protein Eng. Des. Sel. 2015;28:23-28. DOI 10.1093/protein/gzu050.

Melnikov A., Murugan A., Zhang X., Tesileanu T., Wang L., Rogov P., Feizi S., Gnirke A., Callan C.G., Jr., Kinney J.B., Kellis M., Lander E.S., Mikkelsen T.S. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. 2012;30:271-277. DOI 10.1038/nbt.2137.

Morrison S.L. Transformation of *E. coli* by electroporation. Curr. Protoc. Immunol. 2001;21(1):A.3N.1-A.3N.4. DOI 10.1002/0471142735.ima03ns21.

Patwardhan R.P., Hiatt J.B., Witten D.M., Kim M.J., Smith R.P., May D., Lee C., Andrie J.M., Lee S.I., Cooper G.M., Ahituv N., Pennacchio L.A., Shendure J. Massively parallel functional dissection of mammalian enhancers *in vivo*. Nat. Biotechnol. 2012;30:265-270. DOI 10.1038/nbt.2136.

Patwardhan R.P., Lee C., Litvin O., Young D.L., Pe'er D., Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat. Biotechnol. 2009;27:1173-1175. DOI 10.1038/nbt.1589.

Quail M.A., Smith M., Jackson D., Leonard S., Skelly T., Swerdlow H.P., Gu Y., Ellis P. SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. BMC Genomics. 2014;15:110. DOI 10.1186/1471-2164-15-110.

Riddle N.C., Minoda A., Kharchenko P.V., Alekseyenko A.A., Schwartz Y.B., Tolstorukov M.Y., Gorchakov A.A., Jaffe J.D., Kennedy C., Linder-Basso D., Peach S.E., Shanower G., Zheng H., Kuroda M.I., Pirrotta V., Park P.J., Elgin S.C., Karpen G.H. Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. Genome Res. 2011;21:147-163. DOI 10.1101/gr.110098.110.

Ruf S., Symmons O., Uslu V.V., Dolle D., Hot C., Ettwiller L., Spitz F. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat. Genet. 2011;43:379-386. DOI 10.1038/ng.790.

Seguin-Orlando A., Schubert M., Clary J., Stagegaard J., Alberdi M.T., Prado J.L., Prieto A., Willerslev E., Orlando L. Ligation bias in Illumina next-generation DNA libraries: implications for sequencing ancient genomes. PLoS One. 2013;8:e78575. DOI 10.1371/journal.pone.0078575.

Vvedenskaya I.O., Zhang Y., Goldman S.R., Valenti A., Visone V., Taylor D.M., Ebright R.H., Nickels B.E. Massively Systematic Transcript End Readout, "MASTER": transcription start site selection, transcriptional slippage, and transcript yields. Mol. Cell. 2015;60:953-965. DOI 10.1016/j.molcel.2015.10.029.

Wong A.S., Choi G.C., Cui C.H., Pregernig G., Milani P., Adam M., Perli S.D., Kazer S.W., Gaillard A., Hermann M., Shalek A.K., Fraenkel E., Lu T.K. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc. Natl. Acad. Sci. USA. 2016;113:2544-2549. DOI 10.1073/pnas.1517883113.

Woodruff L.B.A., Gorochowski T.E., Roehner N., Mikkelsen T.S., Densmore D., Gordon D.B., Nicol R., Voigt C.A. Registry in a tube: multiplexed pools of retrievable parts for genetic design space exploration. Nucleic Acids Res. 2017;45:1553-1565. DOI 10.1093/nar/gkw1226.

Yankulov K. Dynamics and stability: epigenetic conversions in position effect variegation. Biochem. Cell Biol. 2013;91:6-13. DOI 10.1139/bcb-2012-0048.

ORCID ID

M.O. Lebedev orcid.org/0000-0001-7461-0977
L.A. Yarinich orcid.org/0000-0003-0469-0371
A.V. Ivankin orcid.org/0000-0003-3309-6363
A.V. Pindyurin orcid.org/0000-0001-6959-0641

Acknowledgements. This work was supported by the Russian Science Foundation, project 16-14-10288. DNA sequencing was performed with equipment provided by the Molecular and Cellular Biology core facility, Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences. We are grateful to P.P. Laktionov and D.A. Maksimov for assistance with Illumina sequencing.

Conflict of interest. The authors declare no conflict of interest.

Received December 10, 2018. Revised February 7, 2019. Accepted February 7, 2019.

Вавиловский журнал генетики и селекции. 2019;23(2):212-218
DOI 10.18699/VJ19.484

Mechanisms of transcriptional regulation of ecdysone response

M.Yu. Mazina, N.E. Vorobyeva ✉

Institute of Gene Biology, RAS, Group of transcriptional complexes dynamics, Moscow, Russia
✉ e-mail: vorobyeva@genebiology.ru

The mechanisms of ecdysone-dependent expression have been studied for many decades. Initially, the activation of individual genes under the influence of ecdysone was studied on the model of polythene chromosomes from salivary glands of *Drosophila melanogaster*. These works helped to investigate the many aspects of the *Drosophila* development. They also revealed plenty of valuable information regarding the fundamental mechanisms controlling the genes' work. Many years ago, a model describing the process of gene activation by ecdysone, named after the author – Ashburner model – was proposed. This model is still considered an excellent description of the ecdysone cascade, which is implemented in the salivary glands during the formation of the *Drosophila* pupa. However, these days there is an opinion that the response of cells to the hormone ecdysone can develop with significant differences, depending on the type of cells. The same genes can be activated or repressed under the influence of ecdysone in different tissues. Likely, certain DNA-binding transcription factors that are involved in the ecdysone-dependent response together with the EcR/Usp heterodimer are responsible for cell-type specificity. A number of transcriptional regulators involved in the ecdysone response have been described. Among them are several complexes responsible for chromatin remodeling and modification. It has been shown by various methods that ecdysone-dependent activation/repression of gene transcription develops with significant structural changes of chromatin on regulatory elements. The description of the molecular mechanism of this process, in particular, the role of individual proteins in it, as well as structural interactions between various regulatory elements is a matter of the future. This review is aimed to discuss the available information regarding the main regulators that interact with the ecdysone receptor. We provide a brief description of the regulator's participation in the ecdysone response and links to the corresponding study. We also discuss general aspects of the mechanism of ecdysone-dependent regulation and highlight the most promising points for further research.
Key words: transcription; chromatin; regulator; nuclear receptor; EcR; ecdysone; 20H-ecdysone; *Drosophila*.

For citation: Mazina M.Yu., Vorobyeva N.E. Mechanisms of transcriptional regulation of ecdysone response. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):212-218. DOI 10.18699/VJ19.484

Механизмы регуляции транскрипции под действием экдизона

М.Ю. Мазина, Н.Е. Воробьева ✉

Институт биологии гена Российской академии наук, группа динамики транскрипционных комплексов, Москва, Россия
✉ e-mail: vorobyeva@genebiology.ru

Механизмы экспрессии экдизон-зависимых генов исследуются на протяжении нескольких десятилетий. Исходно активация транскрипции отдельных генов под воздействием экдизона была исследована на модели политенных хромосом *Drosophila melanogaster*. Эти работы помогли изучить многочисленные аспекты развития дрозофилы и выявили ценную информацию относительно фундаментальных механизмов, управляющих работой генов. Модель, описывающая процесс активации генов экдизоном, была предложена еще много лет назад и названа по имени ее автора – Ashburner model. Данная модель до сих пор считается прекрасным описанием экдизонового каскада, который реализуется в слюнных железах во время формирования куколки дрозофилы. Однако к настоящему времени сформировалось понимание того, что ответ клеток на экдизон может развиваться разным образом в зависимости от типа клеток. Под воздействием экдизона одни и те же гены могут активироваться или репрессироваться в клетках различного происхождения. Судя по всему, за такую тканеспецифичность отвечают определенные ДНК-связывающие транскрипционные факторы, которые вовлечены в экдизон-зависимый ответ вместе с EcR/Usp гетеродимером. На сегодняшний день описано множество транскрипционных регуляторов, вовлеченных в процесс экдизонового ответа. Среди них несколько комплексов, ответственных за ремоделирование и модификацию хроматина. Различными методами было показано, что экдизон-зависимая активация/репрессия транскрипции генов протекает со значительными структурными изменениями хроматина на регуляторных элементах. Описание молекулярного механизма этого процесса, в частности роли в нем отдельных белков, а также структурных взаимодействий между различными регуляторными элементами, – дело будущего. Целью нашего обзора является обсуждение имеющейся информации относительно регуляторов транскрипции, взаимодействующих с экдизоновым рецептором. Приведено краткое описание механизма участия регулятора в экдизоновом ответе, а также

ссылки на соответствующее исследование. Обсуждаются общие аспекты механизма экдизон-зависимой регуляции транскрипции в свете последних исследований и выделены наиболее перспективные моменты, которые кажутся нам интересными для дальнейшего изучения.
Ключевые слова: транскрипция; хроматин; регулятор; ядерный рецептор; EcR; экдизон; 20Н-экдизон; дрозофила.

Introduction

For the first time, a pure hormone that controls the metamorphosis of insects, was isolated in 1954 (Butenandt, Karlson, 2014). Some time later its biological activity and role in the formation of the *Drosophila* pupa was confirmed (Chihara et al., 1972). Further, on the experimental model of the polytene chromosomes of *Drosophila melanogaster*, the role of the ecdysone in the transcriptional activation of developmental genes was discovered (Ashburner, 1971; Ashburner et al., 1974).

The studies of drosophila polytene chromosomes had led to the description of ecdysone cascade (Zhimulev et al., 2004). The concept of the cascade is that ecdysone action on the cells results in the activation of a number of "early" genes important for the process of metamorphosis. Products of these genes in turn are transcriptional regulators that activate subsequent target genes. Many years have been required for researchers to uncover the genes participating in the ecdysone cascade (Thummel, 2002). These studies have led to the fact that currently the main transcription activators of the ecdysone cascade genes (which are mainly nuclear receptors) have been studied quite well. Moreover, the functions of the products of the target genes that they induce have also been investigated. The rather poorly studied part of ecdysone-dependent transcription regulation is the molecular mechanism controlling the activation and repression of target genes, which would include a description of the transcriptional complexes involved in it, as well as the specifics of RNA polymerase II functioning during this process.

Ecdysone-binding complex

More than two decades ago, the molecular sensor of Drosophila cells – EcR ecdysone receptor – was isolated and cloned (Koelle et al., 1991). It was found that this protein is able to bind to DNA regulatory elements that determine ecdysone sensitivity (EcRE, ecdysone-response elements) and stimulate ecdysone-dependent activation of gene transcription in reporter systems (Antoniewski et al., 1993). The expression profile of the gene encoding ecdysone receptor during the development of Drosophila coincides in time with the expression of the ecdysone cascade genes. The ecdysone receptor belongs to the nuclear receptors, transcription regulator proteins that directly bind to the DNA regulatory elements of their target genes (King-Jones, Thummel, 2005). Quickly enough after the isolation of the gene, encoding EcR in Drosophila, it was found that the expression of this receptor in mammalian cells may result in the ability of these cells to support the expression of ecdysone-dependent reporter systems (Christopherson et al., 1992). The applying of an ecdysone receptor to control the inducible transcription of transgenes in mammalian cells has become quite widespread due to the high specificity of transcription activation mechanisms. The ligand-binding part of the ecdysone receptor served as the basis for the creation of numerous chimeric transcription activators, whose activity depends on the presence of ecdysone and is insensitive to other types of steroid hormones (No et al., 1996).

In *Drosophila* cells, the functional ecdysone response sensor is a heterodimer formed by an ecdysone receptor with the nuclear receptor Ultraspiracle (Usp). It has been shown that this heterodimer, but not the ecdysone receptor itself, is able to bind ecdysone hormone effectively (Yao et al., 1993). The interaction of the ecdysone receptor with DNA *in vivo* also occurs better if it is a part of the heterodimer. Mutations in the gene encoding the Usp receptor lead to disruptions in the development of *Drosophila*, namely the process of formation of the pupa controlled by the ecdysone cascade (Hall, Thummel, 1998). The *usp* gene mutants demonstrate a damage of the larval cuticle (the formation of a bilayer cuticle due to molting problems), as well as the problems with differentiation of the imaginal discs and with the launch of a cell death program for the salivary glands and the larval intestin.

Using X-ray analysis, the three-dimensional structure of the ligand-binding domains of EcR and Usp receptors was first established in complex with an ecdysone analogue – ponasterone A (ponA), and then with a 20-hydroxyecdysone (Billas et al., 2003; Browning et al., 2007). The ligand-binding pocket for ecdysone hormone is formed entirely by the amino acids of the EcR receptor. Probably, the stimulating role of Usp during the interaction of the heterodimer with ecdysone is to create the necessary stable conformation of the ligand-binding domain of EcR.

EcR and Usp heterodimer is a highly specific sensor of ecdysone and its active metabolite 20-hydroxyecdysone. It is unable to bind other ecdysteroids found in insect hemolymph (Baker et al., 2000). The ability to non-selectively bind ecdysteroids has been found for another nuclear receptor DHR38 (Baker et al., 2003). Interestingly, to form an active complex with various ecdysteroids, the DHR38 receptor interacts with Usp, a partner of the EcR receptor. The discovery of the ability of the DHR38 receptor to bind ecdysteroids explained the presence of morphological differences between mutations of the *ecr* and *usp* genes. Mutations of the *dhr38* gene, as well as of the *usp* gene, show disruption of the cuticle formation, which is not observed in mutations of the *ecr* gene.

Ecdysone-dependent genes are activated repeatedly during the development of Drosophila. Interestingly, the set of ecdysone-induced genes at different stages of development has some differences. Thus, the *e93* gene directly induced by ecdysone is activated in the process of metamorphosis only at the secondary ecdysone peak several hours after the formation of the pupa. It was shown that the activation of the transcription of the *e93* gene requires a special "competence" of the cells – the expression of *betaftz-f1* early-late gene of the ecdysone cascade. The lack of this transcription factor in the third larval stage explains the insensitivity of the *e93* gene to ecdysone in the late third larval stage (Broadus et al., 1999). The exogenous expression of the *ftz-f1* in the larvae

restores the ability of the *e93* gene to be induced by ecdysone (Woodard et al., 1994).

Reiterative expression of ecdysone-dependent genes in development suggests the existence of effective mechanisms for suppressing the transcription of previously activated genes. It is believed that the main role in this process is played by the products of the genes of early ecdysone response. Their activity as transcriptional repressors for EcR receptor gene and for some other early response genes forms negative feedback loops in the ecdysone cascade. One of these genes products, the E75A protein, is a transcriptional repressor of the *ecr* and *br-c* genes (Johnston et al., 2011). Interestingly, its activity as a repressor depends on the intracellular concentration of NO. Only heme of the E75A protein, not bound with NO, is able to interact with SMRTER and repress the transcription of target genes. The EcR receptor itself is also able to interact with the SMRTER protein, which is a repressor of ecdysone genes (ecdysone effects the EcR-SMRTER complex in the way that leads to the removal of the repressor from the complex) (Tsai et al., 1999). The *smrter* gene mutations, as well as the *ecr* gene mutation, disrupting the interaction of EcR and SMRTER, demonstrate the important role of this repressor in regulating the ecdysone and Notch signaling pathways that control the development of *Drosophila* (Heck et al., 2012).

Recently, an alternative mechanism for repression of ecdysone-dependent genes transcription has been proposed. It was shown that the dMi-2 factor can replace the Usp in the EcR-Usp heterodimer to perform repression of the genes of ecdysone response (Kreher et al., 2017). Moreover, overexpression of dMi-2 is even able to displace Usp from its complex with EcR. Interestingly, during activation of ecdysone-induced genes transcription, an increase in the level of dMi-2 binding to regions of inducible genes is observed (Kreher et al., 2017). Perhaps repressor recruitment protects ecdysone genes from transcription activation being too high and is needed to prepare the gene for ongoing transcriptional repression. Similar properties were found by us earlier for another protein of the CHD group – CHD1. We showed that CHD1 is a transcriptional repressor for *dhr3* and *hr4* ecdysone genes of early response, but it is recruited to the promoter of these genes in the first minutes after activation of their transcription by ecdysone, together with transcription coactivator complexes (Mazina et al., 2018).

General aspects of ecdysone-dependent transcription regulation

The molecular mechanisms of ecdysone-dependent activation of transcription are still badly understood because of the complexity of performing the experiments on individual *Drosophila* tissues at different stages of development. Mechanisms of gene activation are understood as the description of regulators recruitment to the promoters and activation-induced modifications of regulatory proteins and the main enzyme, RNA polymerase II. The presence of a model system of polytene chromosomes in *Drosophila* allowed researchers to study the transcription mechanisms for individual genes long before the appearance of whole-genome research tools (Belyaeva et al., 1981; Ashburner, 1990; Gonzy et al., 2002).

It was found that the puffs of ecdysone genes (in particular, *br*, *e74*, e*75*) in the active state are filled exclusively with phosphorylated form of RNA polymerase II and practically do not contain the enzyme in the non-phosphorylated state (Weeks et al., 1993). This may indicate, firstly, that the elongation of transcription at the active ecdysone genes is carried out by the fully phosphorylated form of RNA polymerase II, and, secondly, that during the active phase of the transcription, the non-phosphorylated enzyme is not recruited, but the phosphorylated form is being re-utilized. The well-known transcription timing of the ecdysone cascade genes in the salivary glands allows it to be used as a model system for studying the participation of proteins in various stages of the transcription process (preparation for it or during active transcription) (Brandt, Corces, 2008).

Unfortunately, it is difficult to investigate the detailed mechanisms of developmental genes transcription activation in tissues. Therefore, much of the research on the mechanisms of ecdysone response was carried out in the experimental system of cultured cells. An important aspect of such studies is the fact that the response of cells to treatment with ecdysone can vary significantly (depending on the presence of additional transcriptional regulators in a particular cell type) (Stoiber et al., 2016). These differences can be so significant that the same gene (for example, *CG9932* in the last study) can be activated under the influence of ecdysone in some cells and be repressed in others. This result is quite expected. The fact is that the ecdysone cascade controls the development of Drosophila, namely the transition between different stages. At the same time, the ecdysone cascade can be activated simultaneously in various tissues of the body. Obviously, the development of different organs requires specific patterns of gene activation, that is, the presence of different targets for the ecdysone receptor, depending on the type of tissue.

Several years ago, information began to appear about the possibility of ecdysone-dependent genes regulation by the RNA polymerase II pausing mechanism. It was found that the known factors of RNA polymerase II pausing, NELF and GAF, are associated with ecdysone-dependent promoters of the *ecr* gene (Lee et al., 2008; Fay et al., 2011). Later our group found that the activation of transcription of ecdysone-dependent *dhr3* and *hr4* genes in S2 *Drosophila* cells occurs by stimulating the phosphorylation of RNA polymerase II, associated with the promoters in their inactive state (Mazina et al., 2015). Later, in genome-wide experiments, we showed that such an activation mechanism is inherent to many ecdysone-dependent genes (Mazina et al., 2018).

Of course, whole-genome research methods have brought a lot of valuable information concerning the mechanisms of transcription regulation of ecdysone-dependent genes. Thus, relatively recently, the distinctive features of ecdysone-dependent enhancers were characterized by the STARR-Seq whole-genome method in S2 cells (Shlyueva et al., 2014). It turned out that the functioning of these enhancers depends not only on the binding of the heterodimer EcR-Usp, but also on the binding of other transcription factors. It was shown that the mutation of the Serpent binding site leads to a violation of the ecdysone-dependent activity of a number of regulatory elements in S2 cells, and a defect in the binding of traffic jam interferes with the activity of enhancers in the ovarian cells. Shlyueva and colleagues also showed that the vast majority of ecdysone-dependent enhancers are areas of closed chromatin

Coregulators of ecdysone-dependent transcriptional response

Name of Protein	FlyBase ID	References	Protein complex	Mechanism
TRR	FBgn0023518	Sedkov et al., 2003; Carbonell et al., 2013; Pascual-Garcia et al., 2017	TRR complex	Interacts with EcR, responsible for the incorporation of H3K4me3 modification onto ecdysone-dependent promoters
Tai	FBgn0041092	Bai et al., 2000; Xie et al., 2015; Zhang et al., 2015		Interacts with EcR and Usp, links EcR signaling to Hippo pathway
ISWI	FBgn0011604	Badenhorst et al., 2005; Ables, Drummond-Barbosa, 2010	NURF	Directly interacts with EcR in ecdysone-dependent way
Cdk8	FBgn0015618	Xie et al., 2015; Mazina et al., 2018	Mediator (cdk8 module)	Cdk8 is important for the recruitment of the EcR-Usp complex
Core mediator subunits		Homem et al., 2014; Xie et al., 2015; Mazina et al., 2018	Mediator complex	Interacts with EcR, mediates activation and repression of ecdysone-dependent genes
Crc (ATF4)	FBgn0000370	Gauthier et al., 2012		Binds to B2 isoform of EcR specifically, activates transcription of ecdysone-dependent *eth* gene
Iawc	FBgn262976	Brandt, Corces, 2008		Important for productive phase of transcription elongation
Ash2	FBgn0000139	Carbonell et al., 2013	TRR complex	Interacts with TRR, stabilizes it and facilitates H3K4me3 incorporation at promoters
dDEK	FBgn 0026533	Sawatsubashi et al., 2010	dDEK-dCK2 complex	Interacts with EcR, works as histone chaperone, stimulates incorporation of H3.3 at ecdysone dependent genes
Nup98	FBgn0039120	Pascual-Garcia et al., 2017	Nuclear pore	Contributes to transcriptional memory and enhancer looping of ecdysone-dependent genes
Mtor	FBgn0013756	Pascual-Garcia et al., 2017	Nuclear pore basket	Contributes to transcriptional memory of ecdysone-dependent genes
Putzig	FBgn0259785	Kugler et al., 2011	Associated with NURF	Interacts with EcR *in vivo*, promotes ecdysone signaling
SAYP	FBgn0087008	Vorobyeva et al., 2011, 2012	SWI/SNF	Involved in ecdysone-dependent Pol II pausing
Snr1	FBgn0011715	Zraly et al., 2006; Zraly, Dingwall, 2012	SWI/SNF	Repression of ecdysone-dependent genes, involved in Pol II elongation and splicing processes
Brm	FBgn0000212	Zraly et al., 2006; Mazina et al., 2018	SWI/SNF	Activation and repression of ecdysone-dependent genes
Neiire (CBP/p300)	FBgn0261617	Kirilly et al., 2011; Bodai et al., 2012; Mazina et al., 2018		Interacts with EcR in ecdysone-dependent way, promotes Histone H3 K27 acetylation at ecdysone-dependent genes, is important for ecdysone-dependent transcription activation
DART1	FBgn0037834	Mazina et al., 2018		Involved in ecdysone-dependent transcription activation (may be not connected to its H3R2 methylation activity)
Antimeros (Paf1)	FBgn0010750	Mazina et al., 2018	PAF complex	Important for ecdysone dependent transcription activation (stimulates Pol II CTD Ser2 phosphorylation)
Mi2	FBgn0262519	Kreher et al., 2017	NURD	Interacts with EcR (complex devoid of Usp), constrains transcription of ecdysone-dependent genes
SMRTER	FBgn0265523	Tsai et al., 1999; Johnston et al., 2011; Heck et al., 2012		Repressor of ecdysone-dependent transcription
DOR	FBgn0035542	Francis et al., 2010		Coregulator of ecdysone-dependent transcription in fat body

in an inactive state. Thus, the process of ecdysone-dependent activation of transcription probably involves the remodeling of chromatin on regulatory elements. This observation was confirmed recently by other researchers. It turned out that the remodeling and opening of chromatin on enhancer regulatory elements are the most important steps in controlling the start of transcription of at least for some ecdysone-dependent genes (Uyehara et al., 2017). Disruption of chromatin remodeling on enhancers can interfere with the process of gene activation, or lead to their untimely transcription. Interestingly, many genes involved in the ecdysone cascade are the very proteins that control chromatin density on enhancers at different stages of

development. It was shown that the *e93* gene, a well-known participant of the ecdysone cascade, can take part in both the opening and closing of chromatin on enhancers of genes expressed in wing tissue. Unfortunately, at the moment it is not clear at what stage *e93* interferes with the remodeling process: is it the initiator of the process, or does it play a role in its progression.

Various groups of researchers have demonstrated that the efficiency of ecdysone activation depends on the number of ecdysone response regulatory elements in the activated gene (Bernardo et al., 2014; Shlyueva et al., 2014). The 3C (chromatin conformation capture) analysis reveals the presence of interactions between ecdysone-dependent regulatory elements and inducible gene promoters. Probably, the ecdysone-dependent activation of transcription proceeds as a complex multi-component process involving transcription proteins associated with various regulatory elements. Experiments on individual genes demonstrate the involvement of the EcR receptor in the process of remote interactions between DNA elements (Bernardo et al., 2014). However, additional investigations are required to establish whether this statement extends to other ecdysone-dependent genes and what are the additional participants of this particular process.

Coregulators of ecdysone-dependent transcriptional response

At present, quite a lot of transcription regulator proteins participating in ecdysone-dependent transcription has been shown. Thorough screenings were even performed to find the complete set of such regulators (Davis et al., 2011). Unfortunately, there is relatively little information about the mechanisms of participation of these regulators in the ecdysone response. In a table, we collected information on the most studied ecdysone response regulators.

Perspectives

Despite the fact that the process of ecdysone-dependent transcriptional regulation has been under thorough examination for many decades, many aspects of it still remain unclear. One of them is the type of transcriptional regulation of ecdysone-dependent genes. Currently, it is known that transcription activation can be carried out by stimulating many stages of this process: recruitment of RNA polymerase II, initiating the synthesis of mRNA and stimulating polymerase withdrawal from the promoter, the phase of Pol II transition into the stage of productive elongation. It would be extremely interesting to establish which stages of regulation and in what proportion are utilized by ecdysone-dependent genes. An equally interesting question is whether the way of regulating the same genes changes at different stages of development or in different tissues.

Currently, quite a few ecdysone-dependent transcription regulators are described. Unfortunately, for the most of them the molecular mechanism describing their involvement in the process is either not investigated at all or not studied in sufficient detail. The most important aspect of the study of these regulators will be a comparison of their functions on different ecdysone-dependent genes. Already available information demonstrates that the molecular mechanism of activation may be different even for genes that are activated at the same stages of the ecdysone cascade. Probably, it is this plasticity of the ecdysone response regulation that allows it to become the leading element controlling the development of *Drosophila*.

References

Ables E.T., Drummond-Barbosa D. The steroid hormone ecdysone functions with intrinsic chromatin remodeling factors to control female germline stem cells in *Drosophila*. Cell Stem Cell. 2010;7:581-592. DOI 10.1016/j.stem.2010.10.001.

Antoniewski C., Laval M., Lepesant J.-A. Structural features critical to the activity of an ecdysone receptor binding site. Insect Biochem. Mol. Biol. 1993;23:105-114.

Ashburner M. Induction of puffs in polytene chromosomes of *in vitro* cultured salivary glands of *Drosophila melanogaster* by ecdysone and ecdysone analogues. Nature. New Biol. 1971;230:222-224.

Ashburner M. Puffs, genes, and hormones revisited. Cell. 1990;61:1-3.

Ashburner M., Chihara C., Meltzer P., Richards G. Temporal control of puffing activity in polytene chromosomes. Cold Spring Harb. Symp. Quant. Biol. 1974;38:655-662.

Badenhorst P., Xiao H., Cherbas L., Kwon S.Y., Voas M., Rebay I., Cherbas P., Wu C. The *Drosophila* nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. Genes Dev. 2005;19:2540-2545. DOI 10.1101/gad.1342605.

Bai J., Uehara Y., Montell D.J. Regulation of invasive cell behavior by Taiman, a *Drosophila* protein related to AIB1, a steroid receptor coactivator amplified in breast cancer. Cell. 2000;103:1047-1058.

Baker K.D., Shewchuk L.M., Kozlova T., Makishima M., Hassell A., Wisely B., Caravella J.A., Lambert M.H., Reinking J.L., Krause H., Thummel C.S., Willson T.M., Mangelsdorf D.J. The *Drosophila* orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. Cell. 2003;113:731-742. DOI 10.1016/S0092-8674(03)00420-3.

Baker K.D., Warren J.T., Thummel C.S., Gilbert L.I., Mangelsdorf D.J. Transcriptional activation of the *Drosophila* ecdysone receptor by insect and plant ecdysteroids. Insect Biochem. Mol. Biol. 2000;30:1037-1043. DOI 10.1016/S0965-1748(00)00075-8.

Belyaeva E.S., Vlassova I.E., Biyasheva Z.M., Kakpakov V.T., Richards G., Zhimulev I.F. Cytogenetic analysis of the 2B3-4-2B11 region of the X chromosome of *Drosophila melanogaster*. II. Changes in 20-OH ecdysone puffing caused by genetic defects of puff 2B5. Chromosoma. 1981;84:207-219.

Bernardo T.J., Dubrovskaya V.A., Xie X., Dubrovsky E.B. A view through a chromatin loop: insights into the ecdysone activation of early genes in *Drosophila*. Nucleic Acids Res. 2014;42:10409-10424. DOI 10.1093/nar/gku754.

Billas I.M.L., Iwema T., Garnier J.-M., Mitschler A., Rochel N., Moras D. Structural adaptability in the ligand-binding pocket of the ecdysone hormone receptor. Nature. 2003;426:91-96.

Bodai L., Zsindely N., Gáspár R., Kristó I., Komonyi O., Boros I.M. Ecdysone induced gene expression is associated with acetylation of histone H3 lysine 23 in *Drosophila melanogaster*. PLoS One. 2012;7:e40565. DOI 10.1371/journal.pone.0040565.

Brandt T., Corces V.G. The Lawc protein is required for proper transcription by RNA polymerase II in *Drosophila*. Mol. Genet. Genomics. 2008;280:385-396. DOI 10.1007/s00438-008-0372-z.

Broadus J., McCabe J.R., Endrizzi B., Thummel C.S., Woodard C.T. The *Drosophila* βFTZ-F1 orphan nuclear receptor provides competence for stage-specific responses to the steroid hormone ecdysone. Mol. Cell. 1999;3:143-149. DOI 10.1016/S1097-2765(00)80305-6.

Browning C., Martin E., Loch C., Wurtz J.-M., Moras D., Stote R.H., Dejaegere A.P., Billas I.M.L. Critical role of desolvation in the binding of 20-hydroxyecdysone to the ecdysone receptor. J. Biol. Chem. 2007;282:32924-32934. DOI 10.1074/jbc.M705559200.

Butenandt A., Karlson P. Über die Isolierung eines Metamorphose-Hormons der Insekten in kristallisierter Form. Z. für Naturforschung B. 2014;9(6):389-391. DOI 10.1515/znb-1954-0601.

Carbonell A., Mazo A., Serras F., Corominas M. Ash2 acts as an ecdysone receptor coactivator by stabilizing the histone methyltransferase Trr. Mol. Biol. Cell. 2013;24:361-372.

Chihara C.J., Petri W.H., Fristrom J.W., King D.S. The assay of ecdysones and juvenile hormones on *Drosophila* imaginal disks *in vitro*. J. Insect Physiol. 1972;18:1115-1123.

Christopherson K.S., Mark M.R., Bajaj V., Godowski P.J. Ecdysteroid-dependent regulation of genes in mammalian cells by a *Drosophila* ecdysone receptor and chimeric transactivators. Proc. Natl. Acad. Sci. USA. 1992;89:6314-6318.

Davis M.B., SanGil I., Berry G., Olayokun R., Neves L.H. Identification of common and cell type specific LXXLL motif EcR cofactors using a bioinformatics refined candidate RNAi screen in *Drosophila melanogaster* cell lines. BMC Dev. Biol. 2011;11:66. DOI 10.1186/1471-213X-11-66.

Fay A., Misulovin Z., Li J., Schaaf C.A., Gause M., Gilmour D.S., Dorsett D. Cohesin selectively binds and regulates genes with paused RNA polymerase. Curr. Biol. 2011;21:1624-1634. DOI 10.1016/j.cub.2011.08.036.

Francis V.A., Zorzano A., Teleman A.A. dDOR is an EcR coactivator that forms a feed-forward loop connecting insulin and ecdysone signaling. Curr. Biol. 2010;20:1799-1808. DOI 10.1016/j.cub.2010.08.055.

Gauthier S.A., VanHaaften E., Cherbas L., Cherbas P., Hewes R.S. Cryptocephal, the *Drosophila melanogaster* ATF4, is a specific coactivator for ecdysone receptor isoform B2. PLoS Genet. 2012;8:e1002883. DOI 10.1371/journal.pgen.1002883.

Gonzy G., Pokholkova G.V., Peronnet F., Mugat B., Demakova O.V., Kotlikova I.V., Lepesant J.-A., Zhimulev I.F. Isolation and characterization of novel mutations of the Broad-Complex, a key regulatory gene of ecdysone induction in *Drosophila melanogaster*. Insect Biochem. Mol. Biol. 2002;32:121-132.

Hall B.L., Thummel C.S. The RXR homolog ultraspiracle is an essential component of the *Drosophila* ecdysone receptor. Development. 1998;125;4709-4717.

Heck B.W., Zhang B., Tong X., Pan Z., Deng W.-M., Tsai C.-C. The transcriptional corepressor SMRTER influences both Notch and ecdysone signaling during *Drosophila* development. Biol. Open. 2012;1:182-196. DOI 10.1242/bio.2012047.

Homem C.C.F., Steinmann V., Burkard T.R., Jais A., Esterbauer H., Knoblich J.A. Ecdysone and Mediator change energy metabolism to terminate proliferation in *Drosophila* neural stem cells. Cell. 2014;158:874-888. DOI 10.1016/j.cell.2014.06.024.

Johnston D.M., Sedkov Y., Petruk S., Riley K.M., Fujioka M., Jaynes J.B., Mazo A. Ecdysone- and NO-mediated gene regulation by competing EcR/Usp and E75A nuclear receptors during *Drosophila* development. Mol. Cell. 2011;44:51-61. DOI 10.1016/j.molcel. 2011.07.033.

King-Jones K., Thummel C.S. Nuclear receptors – a perspective from *Drosophila*. Nat. Rev. Genet. 2005;6:311-323.

Kirilly D., Wong J.J., Lim E.K., Wang Y., Zhang H., Wang C., Liao Q., Wang H., Liou Y.-C., Wang H., Yu F. Intrinsic epigenetic factors cooperate with the steroid hormone ecdysone to govern dendrite pruning in *Drosophila*. Neuron. 2011;72:86-100. DOI 10.1016/j.neuron.2011.08.003.

Koelle M.R., Talbot W.S., Segraves W.A., Bender M.T., Cherbas P., Hogness D.S. The drosophila *EcR* gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell. 1991;67:59-77. DOI 10.1016/0092-8674(91)90572-G.

Kreher J., Kovač K., Bouazoune K., Mačinković I., Ernst A.L., Engelen E., Pahl R., Finkernagel F., Murawska M., Ullah I., Brehm A. EcR recruits dMi-2 and increases efficiency of dMi-2-mediated remodelling to constrain transcription of hormone-regulated genes. Nat. Commun. 2017;8:14806. DOI 10.1038/ncomms14806.

Kugler S.J., Gehring E.-M., Wallkamm V., Krüger V., Nagel A.C. The Putzig-NURF nucleosome remodeling complex is required for ecdysone receptor signaling and innate immunity in *Drosophila melanogaster*. Genetics. 2011;188:127-139. DOI 10.1534/genetics.111.127795.

Lee C., Li X., Hechmer A., Eisen M., Biggin M.D., Venters B.J., Jiang C., Li J., Pugh B.F., Gilmour D.S. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. Mol. Cell. Biol. 2008;28:3290-3300. DOI 10.1128/MCB.02224-07.

Mazina M.Y., Kovalenko E.V., Derevyanko P.K., Nikolenko J.V., Krasnov A.N., Vorobyeva N.E. One signal stimulates different transcriptional activation mechanisms. Biochim. Biophys. Acta (BBA) – Gene Regul. Mech. 2018;1861:178-189. DOI 10.1016/j.bbagrm.2018.01.016.

Mazina M.Y., Nikolenko J.V., Fursova N.A., Nedil'ko P.N., Krasnov A.N., Vorobyeva N.E. Early-late genes of the ecdysone cascade as models for transcriptional studies. Cell Cycle Georget. Tex. 2015;14:3593-3601. DOI 10.1080/15384101.2015.1100772.

No D., Yao T.P., Evans R.M. Ecdysone-inducible gene expression in mammalian cells and transgenic mice. Proc. Natl. Acad. Sci. USA. 1996;93:3346-3351.

Pascual-Garcia P., Debo B., Aleman J.R., Talamas J.A., Lan Y., Nguyen N.H., Won K.J., Capelson M. Metazoan nuclear pores provide a scaffold for poised genes and mediate induced enhancer-promoter contacts. Mol. Cell. 2017;66:63-76. DOI 10.1016/j.molcel.2017.02.020.

Sawatsubashi S., Murata T., Lim J., Fujiki R., Ito S., Suzuki E., Tanabe M., Zhao Y., Kimura S., Fujiyama S., Ueda T., Umetsu D., Ito T., Takeyama K.-i., Kato S. A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. Genes Dev. 2010;24:159-170. DOI 0.1101/gad.1857410.

Sedkov Y., Cho E., Petruk S., Cherbas L., Smith S.T., Jones R.S., Cherbas P., Canaani E., Jaynes J.B., Mazo A. Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*. Nature. 2003;426:78-83.

Shlyueva D., Stelzer C., Gerlach D., Yáñez-Cuna J.O., Rath M., Boryń Ł.M., Arnold C.D., Stark A. Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. Mol. Cell. 2014;54:180-192. DOI 10.1016/j.molcel.2014.02.026.

Stoiber M., Celniker S., Cherbas L., Brown B., Cherbas P. Diverse hormone response networks in 41 independent *Drosophila* cell lines. G3: Genes Genomes Genetics. 2016;6:683-694. DOI 10.1534/g3.115.023366.

Thummel C.S. Ecdysone-regulated puff genes 2000. Insect Biochem. Mol. Biol. 2002;32:113-120. DOI 10.1016/S0965-1748(01)00112-6.

Tsai C.C., Kao H.Y., Yao T.P., McKeown M., Evans R.M. SMRTER, a *Drosophila* nuclear receptor coregulator, reveals that EcR-mediated repression is critical for development. Mol. Cell. 1999;4:175-186. DOI 10.1016/S1097-2765(00)80365-2.

Uyehara C.M., Nystrom S.L., Niederhuber M.J., Leatham-Jensen M., Ma Y., Buttitta L.A., McKay D.J. Hormone-dependent control of developmental timing through regulation of chromatin accessibility. Genes Dev. 2017;31:862-875. DOI 10.1101/gad.298182.117.

Vorobyeva N.E., Nikolenko J.V., Krasnov A.N., Kuzmina J.L., Panov V.V., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. SAYP interacts with DHR3 nuclear receptor and participates in ecdysone-dependent transcription regulation. Cell. Cycle. 2011;10:1821-1827. DOI 10.4161/cc.10.11.15727.

Vorobyeva N.E., Nikolenko J.V., Nabirochkina E.N., Krasnov A.N., Shidlovskii Y.V., Georgieva S.G. SAYP and Brahma are important for 'repressive' and 'transient' Pol II pausing. Nucleic Acids Res. 2012;40:7319-7331. DOI 10.1093/nar/gks472.

Weeks J.R., Hardin S.E., Shen J., Lee J.M., Greenleaf A.L. Locus-specific variation in phosphorylation state of RNA polymerase II *in vivo*: correlations with gene activity and transcript processing. Genes Dev. 1993;7:2329-2344.

Woodard C.T., Baehrecke E.H., Thummel C.S. A molecular mechanism for the stage specificity of the Drosophila prepupal genetic response to ecdysone. Cell. 1994;79:607-615.

Xie X.-J., Hsu F.-N., Gao X., Xu W., Ni J.-Q., Xing Y., Huang L., Hsiao H.-C., Zheng H., Wang C., Zheng Y., Xiaoli A.M., Yang F., Bondos S.E., Ji J.Y. CDK8-Cyclin C mediates nutritional regulation of developmental transitions through the ecdysone receptor in *Drosophila*. PLoS Biol. 2015;13:e1002207. DOI 10.1371/journal.pbio.1002207.

Yao T.-P., Forman B.M., Jiang Z., Cherbas L., Chen J.-D., McKeown M., Cherbas P., Evans R.M. Functional ecdysone receptor is the product of *EcR* and *Ultraspiracle* genes. Nature. 1993;366:476-479.

Zhang C., Robinson B.S., Xu W., Yang L., Yao B., Zhao H., Byun P.K., Jin P., Veraksa A., Moberg K.H. The ecdysone receptor coactivator Taiman links Yorkie to transcriptional control of germline stem cell factors in somatic tissue. Dev. Cell. 2015;34:168-180. DOI 10.1016/j.devcel.2015.05.010.

Zhimulev I.F., Belyaeva E.S., Semeshin V.F., Koryakov D.E., Demakov S.A., Demakova O.V., Pokholkova G.V., Andreyeva E.N. Polytene chromosomes: 70 years of genetic research. Int. Rev. Cytol. 2004;241:203-275.

Zraly C.B., Dingwall A.K. The chromatin remodeling and mRNA splicing functions of the Brahma (SWI/SNF) complex are mediated by the SNR1/SNF5 regulatory subunit. Nucleic Acids Res. 2012;40:5975-5987. DOI 10.1093/nar/gks288.

Zraly C.B., Middleton F.A., Dingwall A.K. Hormone-response genes are direct *in vivo* regulatory targets of Brahma (SWI/SNF) complex function. J. Biol. Chem. 2006;281:35305-35315. DOI 10.1074/jbc.M607806200.

ORCID ID

M.Yu. Mazina orcid.org/0000-0002-0480-0553
N.E. Vorobyeva orcid.org/0000-0003-4293-834X

Acknowledgements. This work was supported by Russian Science Foundation, project number 18-14-00219.

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised December 11, 2018. Accepted December 11, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):219-225
DOI 10.18699/VJ19.485

Optogenetic regulation of endogenous gene transcription in mammals

E.S. Omelina[1], A.V. Pindyurin[1, 2] ✉

[1] Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
✉ e-mail: a.pindyurin@mcb.nsc.ru

Despite the rapid development of approaches aimed to precisely control transcription of exogenous genes in time and space, design of systems providing similar tight regulation of endogenous gene expression is much more challenging. However, finding ways to control the activity of endogenous genes is absolutely necessary for further progress in safe and effective gene therapies and regenerative medicine. In addition, such systems are of particular interest for genetics, molecular and cell biology. An ideal system should ensure tunable and reversible spatio-temporal control over transcriptional activity of a gene of interest. Although there are drug-inducible systems for transcriptional regulation of endogenous genes, optogenetic approaches seem to be the most promising for the gene therapy applications, as they are noninvasive and do not exhibit toxicity in comparison with drug-inducible systems. Moreover, they are not dependent on chemical inducer diffusion rate or pharmacokinetics and exhibit fast activation-deactivation switching. Among optogenetic tools, long-wavelength light-controlled systems are more preferable for use in mammalian tissues in comparison with tools utilizing shorter wavelengths, since far-red/near-infrared light has the maximum penetration depth due to lower light scattering caused by lipids and reduced tissue autofluorescence at wavelengths above 700 nm. Here, we review such light-inducible systems, which are based on synthetic factors that can be targeted to any desired DNA sequence and provide activation or repression of a gene of interest. The factors include zinc finger proteins, transcription activator-like effectors (TALEs), and the CRISPR/Cas9 technology. We also discuss the advantages and disadvantages of these DNA targeting tools in the context of the light-inducible gene regulation systems.
Key words: zinc finger; TALE; CRISPR/Cas9; optogenetics; transcription regulation.

For citation: Omelina E.S., Pindyurin A.V. Optogenetic regulation of endogenous gene transcription in mammals. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):219-225. DOI 10.18699/VJ19.485

Оптогенетическая регуляция транскрипции эндогенных генов млекопитающих

Е.С. Омелина[1], А.В. Пиндюрин[1, 2] ✉

[1] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ e-mail: a.pindyurin@mcb.nsc.ru

Несмотря на быстрое развитие подходов, контролирующих транскрипцию экзогенных генов во времени и пространстве, разработка систем, обеспечивающих точную регуляцию экспрессии эндогенных генов, является намного более сложной. Однако для дальнейшего прогресса в области безопасной и эффективной генной терапии, а также регенеративной медицины совершенно необходимо искать подходы для контроля активности эндогенных генов. Кроме того, такие подходы представляют интерес для исследований в области генетики, молекулярной и клеточной биологии. Идеальная система для регуляции транскрипции интересующего гена должна удовлетворять следующим требованиям: быть настраиваемой, обратимой, тканеспецифичной и регулируемой во времени. Несмотря на то что в настоящее время известны системы для регуляции транскрипции эндогенных генов, индуцируемые химическими агентами, наиболее многообещающими для применения в области генной терапии представляются оптогенетические системы, поскольку они неинвазивные и нетоксичные. Также оптогенетические системы не зависят от скорости диффузии химического агента и фармакокинетики и обнаруживают высокую скорость переключения между активным и неактивным состояниями. Причем системы, контролируемые светом длинноволнового диапазона, являются наиболее предпочтительными для использования в тканях млекопитающих по сравнению с системами, использующими свет с короткой длиной волны. Это связано с тем, что дальний красный и ближний инфракрасный свет обладают наибольшей проницаемостью в ткани млекопитающих, благодаря меньшему рассеиванию света, вызванному липидами, и сниженной автофлуоресценции тканей на длинах волн более 700 нм. В данном обзоре мы рассматриваем свето-индуцируемые системы, основанные на использовании синтетических факторов, которые способны связываться с любой желаемой последователь-

ностью ДНК, обеспечивая тем самым направленную активацию или репрессию интересующих эндогенных генов. Такие синтетические факторы основаны на белках типа цинковый палец, TALE-ассоциированных белках (transcription activator-like effectors) и технологии CRISPR/Cas9. Мы обсуждаем также преимущества и недостатки этих ДНК-связывающих факторов для оптогенетической регуляции активности генов.
Ключевые слова: цинковый палец; TALE; CRISPR/Cas9; оптогенетика; регуляция транскрипции.

Introduction

The dynamic nature of gene transcription is an essential process for the cellular programming, homeostasis, environmental adaptation, development, and behavior of live organisms. Therefore, design of new approaches for the gene expression regulation is important for the development of gene therapy. A variety of drug-inducible (e. g. tetracycline and rapamycin) systems and light-sensitive tools enabling a control of exogenous reporter genes are known (Gossen, Bujard, 1992; Gossen et al., 1995; Rivera et al., 1996; Yazawa et al., 2009; Kennedy et al., 2010; Wang et al., 2012; Müller et al., 2013; Kaberniuk et al., 2016; Redchuk et al., 2017, 2018). However, the development of technologies that provide modulation of transcription in mammalian endogenous genome seems to be more challenging. Recently, some approaches allowing to control transcription of endogenous genes were designed. They are based on zinc finger proteins, transcription activator-like effectors (TALEs), and the clustered, regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated 9 (Cas9) technology. The first two strategies require engineering of proteins with unique DNA-binding specificity. However, the protein engineering protocols are laborious and time-consuming. Additionally, such proteins are not always effective and often exhibit modest activities. In contrast to zinc finger- and TALE-based synthetic proteins, Cas9-based transcription factors are targeted to DNA sequences by guide RNA molecules (Jinek et al., 2012). The ability to define the sequence specificity of transcription factors by simple exchange of short targeting sequences provides an opportunity to regulate expression of numerous genes without the need for laborious protein engineering approaches.

Here, we review the light-controlled systems for regulation of endogenous gene expression based on zinc finger proteins, TALEs, and nuclease-inactive Cas9 (dCas9). Unlike drug-inducible tetracycline and rapamycin systems, optogenetic systems are non-toxic, exhibit fast activation-deactivation switching and enable noninvasive control of animal physiology and behavior. Among all existing optogenetic systems, far-red- and near-infrared-light-controlled tools are favorable compared to tools utilizing shorter wavelengths, since long-wavelength light has the maximum penetration depth due to its low absorbance by hemoglobin, melanin, and water in mammalian tissues.

Systems based on zinc finger transcription factors

Zinc finger proteins are a major class of DNA-binding proteins that recognize specific DNA target sequences by a one-to-one non-covalent interaction between individual amino acids from the recognition helix to individual nucleotide bases (Pavletich, Pabo, 1991; Klug, 2010). Since the separate zinc fingers function as independent modules, fingers with different specificity can be linked in series to generate arrays that recognize long DNA sequences. A lot of zinc finger domains targeting different sequences were engineered by rational design or high-throughput selection (Berg, 1988; Choo, Klug, 1994a, b).

A first zinc finger-based system for transcription regulation of endogenous genes was described in 1994 (Choo et al., 1994), when a three-finger peptide binding specifically to a unique 9-bp DNA sequence was created. This peptide repressed expression of *p190 Bcr-Abl* oncogene due to a transcriptional blockage imposed by the sequence-specific binding of the peptide. Later, functional transcription factors were made by fusion of zinc finger DNA-binding modules with appropriate effector domains. For instance, synthetic transcriptional repressors were engineered by fusion of zinc fingers to a KOX1 repression domain from the KRAB zinc finger family (Papworth et al., 2003; Reynolds et al., 2003). To activate transcription of the genes of interest, fusions of the zinc finger modules to transcriptional activation domains such as NF-κB p65, herpes simplex virus VP16 and its tetrameric repeat VP64 were used (Beerli et al., 2000; Liu et al., 2001; Rebar et al., 2002). To allow the temporal regulation of

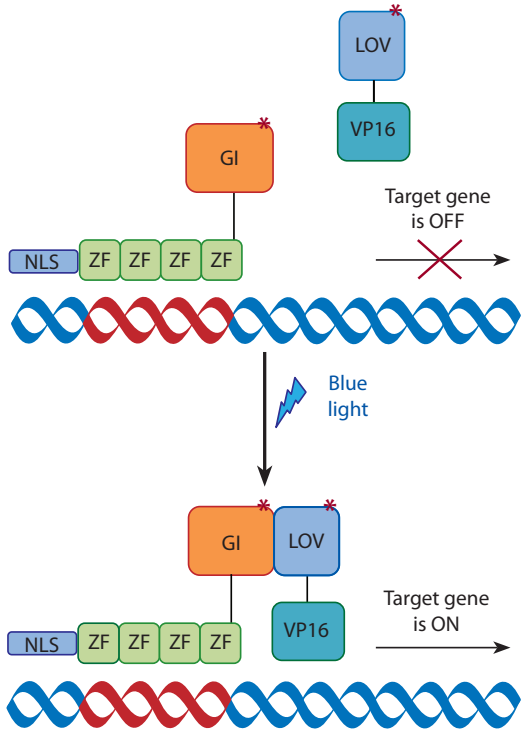


Fig. 1. The principle of the LITEZ system.

A synthetic protein consisting of the four zinc finger domains (ZF) fused with the Gigantea (GI) protein and a NLS binds its recognition site upstream of the target gene. GI and LOV form heterodimers upon blue light (450 nm) illumination, resulting in the translocation of the transcriptional activation domain VP16 to the promoter of the target gene and transcription activation. Red region in DNA helix indicates zinc finger protein binding site. Asterisks denote intrinsic NLSs in GI and LOV.

zinc finger transcription factor's activity, their drug-inducible versions were created by adding a progesterone receptor ligand-binding domain (Dent et al., 2007), estrogen receptor homodimers or retinoid X receptor-α/ecdysone receptor heterodimers (Magnenat et al., 2008).

In addition to the constitutive and chemically-induced approaches, a system of a light-inducible transcription using engineered zinc finger proteins (LITEZ) was designed (Polstein, Gersbach, 2012, 2014). This system utilizes two blue light-inducible dimerizing proteins from *Arabidopsis thaliana*, Gigantea (GI) and the LOV domain of the FKF1 protein (Fig. 1).

The LOV domain was fused to three repeats of the VP16 activation domain and GI was linked to a zinc finger protein and a nuclear localization signal (NLS). Two four-finger proteins and one six-finger protein well-characterized in the previous studies (Beerli et al., 2000; Perez et al., 2008) were used. A *GFP* reporter integrated into the genome of HEK293T cells was used as a model of endogenous target gene. Blue light illumination (450 nm) initiated heterodimerization between GI and LOV, which resulted in the translocation of the LOV-VP16 fusion to the zinc finger protein target sequence and subsequent transcription activation of the endogenous *GFP* reporter. After 30 h of pulsing blue light illumination, a 4-fold increase of GFP positive cell count in the illuminated samples (~16 % GFP positive cells) was observed compared to control samples incubated in darkness (~4 % GFP positive cells). Additionally, the authors detected 30 % increase of the mean GFP fluorescence in illuminated cells compared to the control ones. Finally, it was shown that light-induced transcriptional activation by LITEZ is reversible and repeatable by modulation of the illumination time (Polstein, Gersbach, 2012, 2014).

Systems based on transcription-activator-like effectors

The TALEs of the plant bacterial pathogen *Xanthomonas* represent another class of modular DNA-binding proteins. They recognize DNA by highly conserved tandem repeats, each 33–35 amino acids in length (Moscou, Bogdanove, 2009). These repeats specify nucleotides via unique repeat-variable diresidues (RVDs) at amino acid positions 12 and 13. There is a strong correlation between RVDs and the corresponding nucleotide in the TALE-binding site (Boch et al., 2009). The presence of this association allows to design the sequence-specific DNA-binding proteins, similarly to the construction of zinc finger transcription factors.

Since cloning of new TALE variants is challenging due to a large number of repeat domains, a hierarchical ligation-based strategy was developed to overcome this problem (Zhang et al., 2011). To generate TALE-based artificial transcription factors, VP16 and VP64 activation domains as well as SID and KRAB repression domains were fused to TALEs (Geissler et al., 2011; Miller et al., 2011; Cong et al., 2012). Small-molecule-inducible TALE transcription factors were also described. For instance, a 2–3-fold upregulation of target *icam-1* gene was observed in HeLa cells transfected with plasmids encoding TALE fusions with the ligand-binding domain of the chimeric single-chain retinoid X receptor-α/ecdysone receptor in response to ponasterone A treatment (Mercer et al., 2014).

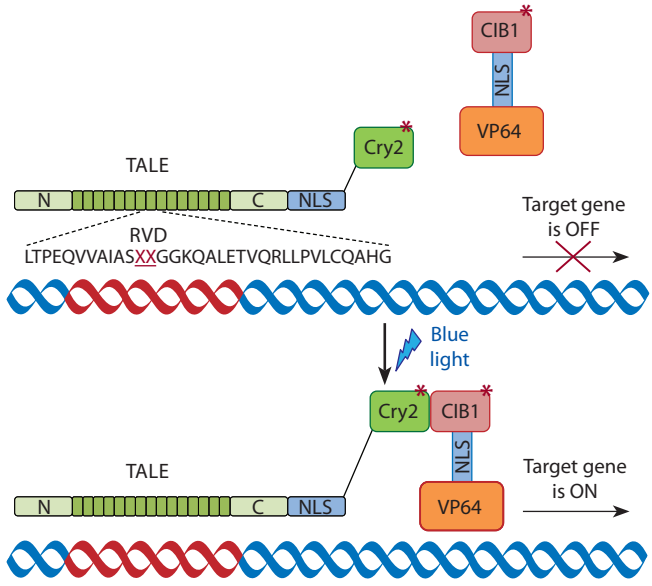


Fig. 2. The principle of the LITE system.

A TALE-Cry2 fusion protein binds to its recognition site upstream of the gene of interest. Blue light (466 nm) illumination induces heterodimerization between Cry2 and CIB1, translocating the VP64 effector domain to the target promoter. Red region in DNA helix indicates TALE binding site. Asterisks denote intrinsic NLSs in Cry2 and CIB1.

In addition to the described above TALE-based systems, an optogenetic two-hybrid system of light-inducible transcriptional effectors (LITEs) was developed (Konermann et al., 2013). The LITE system consists of two components: a TALE fused to the blue-light-sensitive cryptochrome 2 (Cry2) protein from *A. thaliana* and the interacting partner of Cry2, the CIB1 protein, fused to VP64 activation domain. Blue light (466 nm) triggers heterodimerization of Cry2 and CIB1, recruiting VP64 transcription activation domain to the promoter of the target gene (Fig. 2). A panel of different LITEs was developed and applied to upregulate expression of 28 genes (*Hat1*, *Sirt1*, *Mchr1*, *Htr1b*, etc.) in cultured mouse primary neurons. As a result, a 1.5–30-fold increase in mRNA levels was observed upon blue light (466 nm) illumination for 24 h compared to darkness (Konermann et al., 2013). Additionally, LITE system was introduced into cultured mouse primary cortical neurons to control expression of the *Grm2* gene, which resulted in an approximately 7-fold increase of its mRNA level after 24 h of blue light illumination compared to darkness. Moreover, application of the LITE system in mouse prefrontal cortex *in vivo* caused 2-fold increase in *Grm2* mRNA level after 12 h light stimulation (473 nm) compared to GFP-only controls (Konermann et al., 2013).

Modification of this system, LITE2.0, allowed a six-fold reduction of the background *Neurog2* activation compared with the original design, resulting in the 20-fold induction of the *Neurog2* transcription level under 12 h blue light (466 nm) illumination in Neuro 2a cells compared to darkness (Konermann et al., 2013). Further, TALE-based epigenetic modifier (epiTALE) was developed by fusing Cry2 with NLS and the four repeats of repressive histone effector SID, whereas TALE was linked with NLS and CIB1. As a result, a two-fold light-

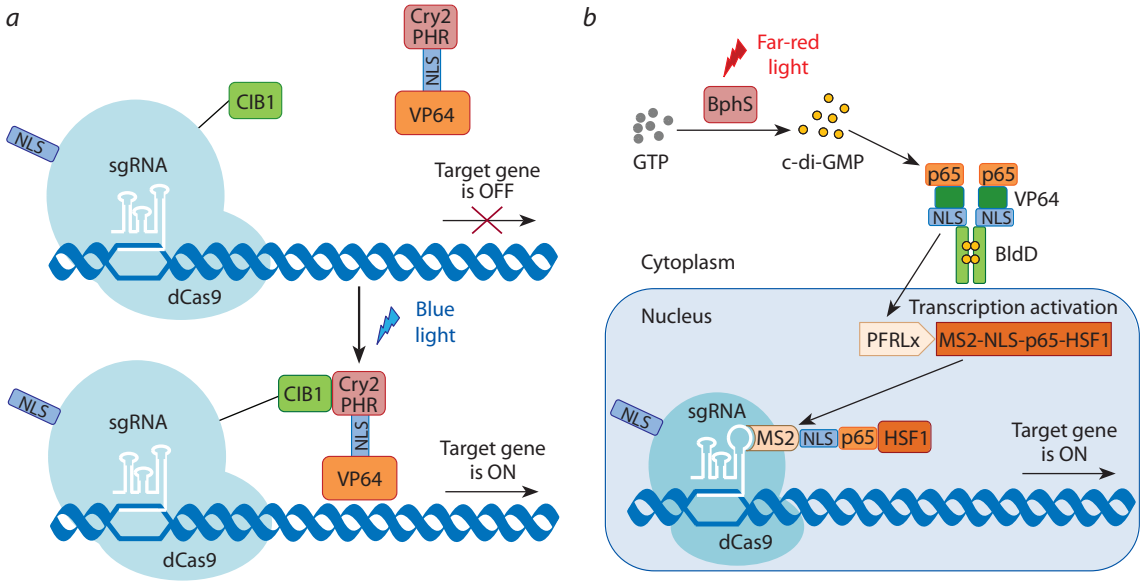


Fig. 3. The principle of the CRISPR/Cas9-based photoactivatable transcription systems.

(*a*) A fusion protein dCas9-CIB1 directed by sgRNA binds the promoter region of the gene of interest. Under blue light (470 nm) illumination, CIB1 forms heterodimers with photolyase homology region (PHR) of Cry2, causing recruitment of VP64 transcriptional activation domain to the promoter region of the target gene; (*b*) The FACE system consists of the bacterial photoreceptor BphS, which is activated by far-red light (730 nm) and converts GTP into c-di-GMP. c-di-GMP is required for dimerization of the BldD transcription factor, resulting in its activation. The activated BldD transcription factor can turn on expression of the transgene encoding the MS2-NLS-p65-HSF1 fusion protein. The latter is recruited to the MS2 binding site located within sgRNA, inducing transcription activation of the target gene.

mediated transcriptional repression of *Grm2* accompanied by a two-fold reduction in H3K9 acetylation at the targeted *Grm2* promoter was observed in neurons after 24 h light stimulation. A set of 32 variants of epiTALE system, containing different repressive histone effector domains (e. g., histone deacetylases, methyltransferases, acetyltransferase inhibitors) was also developed. As a result, 23 variants of epiTALE system caused a 2–3-fold repression of the *Grm2* gene in primary neurons and 20 epiTALEs led to a 1.5–2-fold repression of the *Neurog2* gene in Neuro 2a cells after 24 h and 12 h of blue light (466 nm) illumination, respectively (Konermann et al., 2013).

Systems based on the CRISPR/Cas9 technology

The third type of systems for transcriptional regulation of endogenous genes in mammals is based on the CRISPR/Cas9 technology that consists of the Cas9 nuclease protein and a single guide RNA (sgRNA) allowing the nuclease to bind a specific DNA sequence through RNA-DNA base pairing (Sternberg et al., 2014). The most commonly used Cas9 from *Streptococcus pyogenes* requires a 5′-NGG protospacer-adjacent motif (PAM) immediately adjacent to a 20-nt DNA target sequence in the genome (Nishimasu et al., 2014).

A protein dCas9 lacking the nuclease activity was used for the programmable RNA-guided transcriptional regulation of diverse human genes. The first system utilizing the CRISPR/Cas9 technique for transcription regulation instead of genome editing in mammals was described in 2013 (Qi et al., 2013). After that, several studies have shown that dCas9 fused with the effector domains allows repression or activation of target genes (Gilbert et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013).

To enable precise spatiotemporal control of gene expression, a targeted photoactivation system based on dCas9 was developed (Nihongaki et al., 2015). This system consists of photolyase homology region (PHR) of the blue light-sensitive Cry2 protein fused with the VP64 transcriptional activator domain and dCas9 fused with the Cry2 binding partner, CIB1. Upon blue light (470 nm) illumination, Cry2 and CIB1 form heterodimers, resulting in recruitment of the VP64 effector domain to the target gene and, consequently, in transcriptional activation of the latter (Fig. 3, *a*).

To select the most effective light-inducible transcription CRISPR/Cas9 system, several variants of constructs, differed by type of activator domain (VP64 or p65), the quantity and location of NLSs and size of the CIB1 protein (full-length or truncated) were tested. As a result, a combination of NLS-dCas9-CIB1(Δ308–334) and NLS×3-Cry2PHR-p65 fusion proteins was selected (Nihongaki et al., 2015). This optimized system was applied to HEK293T cells for the light-induced expression of the endogenous *ascl1* gene. Expression of multiple sgRNAs in a single cell enabled synergistic light-induced activation of the gene (~50-fold in the lit state compared to darkness) in contrast to the usage of individual sgRNAs (up to 10-fold). It was shown that 3 h of blue light (470 nm) illumination was enough for an approximately ten-fold induction of the *ascl1* transcription. Light-induced activation of transcription by this system was also reversible and repeatable. Additionally, an opportunity of multiplexed photoactivation of different genes was demonstrated. To achieve that, HEK293T cells were co-transfected with the constructs NLS-dCas9-CIB1(Δ308–334), NLS×3-Cry2PHR-p65 and multiple sgRNAs, targeting the *myod1*, *nanog* and *il1rn* genes. As a result, a 3–1000-fold increase in mRNA levels of these

genes was observed, confirming that this system can be used for multiplexed photoactivation of user-defined endogenous genes (Nihongaki et al., 2015).

A similar strategy, light-activated CRISPR/Cas9 effector (LACE) system, was developed for the dynamic regulation of endogenous genes (Polstein, Gersbach, 2015). An optimized LACE system consists of two components: CIBN-dCas9-CIBN, where CIBN is the N-terminal fragment of CIB1, and Cry2-VP64. An application of this system in HEK293T cells resulted in an 11- and 400-fold upregulation of the *il1rn* gene transcription compared to darkness after 2 h and 30 h of blue light (450 nm) illumination, respectively. This system was also applied for simultaneous photoactivation of multiple human genes (*hbg1*/*2*, *il1rn* and *ascl1*) in HEK293T cells. As a result, illuminated cells had significantly greater mRNA levels of the studied genes in the lit state than in darkness (Polstein, Gersbach, 2015).

One more system utilizing dCas9 to downregulate transcription at endogenous genome loci was developed (Pathak et al., 2017). This system is based on the previously described light-inducible clustering of Cry2-tagged proteins (Ozkan-Dagliyan et al., 2013), resulting in functional loss of their activity. This clustering property of Cry2 was used to block transcription of the targeting gene with light. To achieve that, the Cry2-dCas9-VP64 fusion protein and sgRNAs, targeting the human *il1rn* promoter, were designed. In darkness, a strong induction of the *il1rn* gene was observed in HEK293T cells co-transfected with plasmids encoding Cry2-dCas9-VP64 and sgRNAs. Blue light (450 nm) exposure for three days caused a four-fold reduction of the *il1rn* transcription due to formation of clusters consisting of multiple Cry2-dCas9-VP64 fusion proteins that are unable to bind the promoter region of the *il1rn* gene (Pathak et al., 2017).

Recently, a multicomponent far-red light (FRL)-activated CRISPR/dCas9 effector (FACE) system, inducing transcription of target genes in the presence of FRL stimulation, was engineered (Shao et al., 2018). This system is based on using synthetic bacterial FRL-activated cyclic diguanylate monophosphate (c-di-GMP) synthase BphS (Ryu, Gomelsky, 2014), which converts GTP into c-di-GMP (see Fig. 3, *b*). Increased production of c-di-GMP causes dimerization of the FRL-dependent transactivator p65-VP64-NLS-BldD, where BldD is a transcription factor from *Streptomyces coelicolor* which is non-active in a monomer state. Active BldD binds to its chimeric promoter P_{FRLx}, resulting in initiation of the MS2-p65-HSF1 transactivator expression. Then, MS2 fused to transactivation domains p65 and HSF1 is recruited by sgRNAs bearing the MS2 binding site, causing an induction of endogenous gene expression. Application of the FACE system for the separate and multiplexed regulation of the *ttn*, *il1rn*, *ascl1*, and *rhoxf2* genes in HEK293 cells resulted in a high (about 100–450-fold) increase of relative mRNA levels in the lit state compared to darkness. Photoactivation of the endogenous *ascl1* gene in HEK293 cells, containing the FACE system and implanted into the dorsum of mice, resulted in a 195-fold increase of its mRNA level. Additionally, tibialis posterior muscles of mice were electroporated with plasmids encoding components of the FACE system targeting the promoter regions of the *lama1* and *fst* genes. Subsequent illumination with FRL caused a 2- and 5-fold induction of the *lama1* and *fst* transcription, respectively, compared to the control kept in darkness. Moreover, the FACE system was used to initiate functional neural differentiation of mouse induced pluripotent stem cells by FRL-induced activation of the *neurog2* gene (Shao et al., 2018).

Conclusion

A control of the expression of genes of interest requires the development of molecular tools that precisely recognize specific DNA sequences in the context of the genome. Over the past 20 years, three main methods for design of synthetic transcription factors recognizing any desired target DNA sequence were developed. Although the first two methods based on the usage of zinc finger proteins and TALEs have been widely successful for many applications, development of CRISPR/Cas9-based technology was like a breakthrough. The main advantage of this technology is the lack of a laborious cloning to obtain a site-specific DNA binding protein. Additionally, this system enables simultaneous multiplexed control of user-defined endogenous genes. Combination of the CRISPR/Cas9-based transcription system with the blue-light-induced proteins allows rapid and reversible target gene activation by blue light (Nihongaki et al., 2015). However, application of long wavelength light is more preferable due to the maximum penetration depth in mammalian tissues. In this respect, the recently described FACE system that combines the CRISPR/Cas9 technology with FRL-activated bacterial photoreceptor BphS (Shao et al., 2018) is of special interest. Despite the fact that this system is complex and multicomponent, it seems to be a good start point for the development of reversible and tunable approaches for transcription regulation of endogenous genes for the subsequent safe medical applications in humans.

References

Beerli R.R., Dreier B., Barbas C.F. 3rd. Positive and negative regulation of endogenous genes by designed transcription factors. Proc. Natl. Acad. Sci. USA. 2000;97(4):1495-1500. DOI 10.1073/pnas.040552697.

Berg J.M. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. Proc. Natl. Acad. Sci. USA. 1988;85(1):99-102.

Boch J., Scholze H., Schornack S., Landgraf A., Hahn S., Kay S., Lahaye T., Nickstadt A., Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 2009;326(5959):1509-1512. DOI 10.1126/science.1178811.

Choo Y., Klug A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proc. Natl. Acad. Sci. USA. 1994a;91(23):11168-11172.

Choo Y., Klug A. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. USA. 1994b;91(23):11163-11167.

Choo Y., Sánchez-García I., Klug A. *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature. 1994;372(6507):642-645. DOI 10.1038/372642a0.

Cong L., Zhou R., Kuo Y.C., Cunniff M., Zhang F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat. Commun. 2012;3:968. DOI 10.1038/ncomms1962.

Dent C.L., Lau G., Drake E.A., Yoon A., Case C.C., Gregory P.D. Regulation of endogenous gene expression using small molecule-controlled engineered zinc-finger protein transcription factors. Gene Ther. 2007;14(18):1362-1369. DOI 10.1038/sj.gt.3302985.

Geissler R., Scholze H., Hahn S., Streubel J., Bonas U., Behrens S.E., Boch J. Transcriptional activators of human genes with programmable DNA-specificity. PLoS One. 2011;6(5):e19509. DOI 10.1371/journal.pone.0019509.

Gilbert L.A., Larson M.H., Morsut L., Liu Z., Brar G.A., Torres S.E., Stern-Ginossar N., Brandman O., Whitehead E.H., Doudna J.A., Lim W.A., Weissman J.S., Qi L.S. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013; 154(2):442-451. DOI 10.1016/j.cell.2013.06.044.

Gossen M., Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. USA. 1992;89(12):5547-5551.

Gossen M., Freundlieb S., Bender G., Müller G., Hillen W., Bujard H. Transcriptional activation by tetracyclines in mammalian cells. Science. 1995;268(5218):1766-1769.

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.

Kaberniuk A.A., Shemetov A.A., Verkhusha V.V. A bacterial phytochrome-based optogenetic system controllable with near-infrared light. Nat. Methods. 2016;13(7):591-597. DOI 10.1038/nmeth.3864.

Kennedy M.J., Hughes R.M., Peteya L.A., Schwartz J.W., Ehlers M.D., Tucker C.L. Rapid blue-light-mediated induction of protein interactions in living cells. Nat. Methods. 2010;7(12):973-975. DOI 10.1038/nmeth.1524.

Klug A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu. Rev. Biochem. 2010; 79:213-231. DOI 10.1146/annurev-biochem-010909-095056.

Konermann S., Brigham M.D., Trevino A.E., Hsu P.D., Heidenreich M., Cong L., Platt R.J., Scott D.A., Church G.M., Zhang F. Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013;500(7463):472-476. DOI 10.1038/nature12466.

Liu P.Q., Rebar E.J., Zhang L., Liu Q., Jamieson A.C., Liang Y., Qi H., Li P.X., Chen B., Mendel M.C., Zhong X., Lee Y.L., Eisenberg S.P., Spratt S.K., Case C.C., Wolffe A.P. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. J. Biol. Chem. 2001;276(14):11323-11334. DOI 10.1074/jbc.M011172200.

Maeder M.L., Linder S.J., Cascio V.M., Fu Y., Ho Q.H., Joung J.K. CRISPR RNA-guided activation of endogenous human genes. Nat. Methods. 2013;10(10):977-979. DOI 10.1038/nmeth.2598.

Magnenat L., Schwimmer L.J., Barbas C.F. 3rd. Drug-inducible and simultaneous regulation of endogenous genes by single-chain nuclear receptor-based zinc-finger transcription factor gene switches. Gene Ther. 2008;15(17):1223-1232. DOI 10.1038/gt.2008.96.

Mercer A.C., Gaj T., Sirk S.J., Lamb B.M., Barbas C.F. 3rd. Regulation of endogenous human gene expression by ligand-inducible TALE transcription factors. ACS Synth. Biol. 2014;3(10):723-730. DOI 10.1021/sb400114p.

Miller J.C., Tan S., Qiao G., Barlow K.A., Wang J., Xia D.F., Meng X., Paschon D.E., Leung E., Hinkley S.J., Dulay G.P., Hua K.L., Ankoudinova I., Cost G.J., Urnov F.D., Zhang H.S., Holmes M.C., Zhang L., Gregory P.D., Rebar E.J. A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 2011;29(2):143-148. DOI 10.1038/nbt.1755.

Moscou M.J., Bogdanove A.J. A simple cipher governs DNA recognition by TAL effectors. Science. 2009;326(5959):1501. DOI 10.1126/science.1178817.

Müller K., Engesser R., Metzger S., Schulz S., Kämpf M.M., Busacker M., Steinberg T., Tomakidi P., Ehrbar M., Nagy F., Timmer J., Zubriggen M.D., Weber W. A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. Nucleic Acids Res. 2013;41(7):e77. DOI 10.1093/nar/gkt002.

Nihongaki Y., Yamamoto S., Kawano F., Suzuki H., Sato M. CRISPR-Cas9-based photoactivatable transcription system. Chem. Biol. 2015;22(2):169-174. DOI 10.1016/j.chembiol.2014.12.011.

Nishimasu H., Ran F.A., Hsu P.D., Konermann S., Shehata S.I., Dohmae N., Ishitani R., Zhang F., Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014;156(5):935-949. DOI 10.1016/j.cell.2014.02.001.

Ozkan-Dagliyan I., Chiou Y.Y., Ye R., Hassan B.H., Ozturk N., Sancar A. Formation of *Arabidopsis* Cryptochrome 2 photobodies in mammalian nuclei: application as an optogenetic DNA damage checkpoint switch. J. Biol. Chem. 2013;288(32):23244-23251. DOI 10.1074/jbc.M113.493361.

Papworth M., Moore M., Isalan M., Minczuk M., Choo Y., Klug A. Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors. Proc. Natl. Acad. Sci. USA. 2003; 100(4):1621-1626. DOI 10.1073/pnas.252773399.

Pathak G.P., Spiltoir J.I., Höglund C., Polstein L.R., Heine-Koskinen S., Gersbach C.A., Rossi J., Tucker C.L. Bidirectional approaches for optogenetic regulation of gene expression in mammalian cells using *Arabidopsis* cryptochrome 2. Nucleic Acids Res. 2017;45(20):e167. DOI 10.1093/nar/gkx260.

Pavletich N.P., Pabo C.O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science. 1991;252(5007): 809-817.

Perez E.E., Wang J., Miller J.C., Jouvenot Y., Kim K.A., Liu O., Wang N., Lee G., Bartsevich V.V., Lee Y.L., Guschin D.Y., Rupniewski I., Waite A.J., Carpenito C., Carroll R.G., Orange J.S., Urnov F.D., Rebar E.J., Ando D., Gregory P.D., Riley J.L., Holmes M.C., June C.H. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. 2008;26(7):808-816. DOI 10.1038/nbt1410.

Perez-Pinera P., Kocak D.D., Vockley C.M., Adler A.F., Kabadi A.M., Polstein L.R., Thakore P.I., Glass K.A., Ousterout D.G., Leong K.W., Guilak F., Crawford G.E., Reddy T.E., Gersbach C.A. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods. 2013;10(10):973-976. DOI 10.1038/nmeth.2600.

Polstein L.R., Gersbach C.A. Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. J. Am. Chem. Soc. 2012;134(40):16480-16483. DOI 10.1021/ja3065667.

Polstein L.R., Gersbach C.A. Light-inducible gene regulation with engineered zinc finger proteins. Methods Mol. Biol. 2014;1148:89-107. DOI 10.1007/978-1-4939-0470-9_7.

Polstein L.R., Gersbach C.A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nat. Chem. Biol. 2015; 11(3):198-200. DOI 10.1038/nchembio.1753.

Qi L.S., Larson M.H., Gilbert L.A., Doudna J.A., Weissman J.S., Arkin A.P., Lim W.A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013; 152(5):1173-1183. DOI 10.1016/j.cell.2013.02.022.

Rebar E.J., Huang Y., Hickey R., Nath A.K., Meoli D., Nath S., Chen B., Xu L., Liang Y., Jamieson A.C., Zhang L., Spratt S.K., Case C.C., Wolffe A., Giordano F.J. Induction of angiogenesis in a mouse model using engineered transcription factors. Nat. Med. 2002;8(12):1427-1432. DOI 10.1038/nm795.

Redchuk T.A., Karasev M.M., Omelina E.S., Verkhusha V.V. Near-infrared light-controlled gene expression and protein targeting in neurons and non-neuronal cells. Chembiochem. 2018;19(12):1334-1340. DOI 10.1002/cbic.201700642.

Redchuk T.A., Omelina E.S., Chernov K.G., Verkhusha V.V. Near-infrared optogenetic pair for protein regulation and spectral multiplexing. Nat. Chem. Biol. 2017;13(6):633-639. DOI 10.1038/nchembio.2343.

Reynolds L., Ullman C., Moore M., Isalan M., West M.J., Clapham P., Klug A., Choo Y. Repression of the HIV-1 5′ LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger tran-

scription factors. Proc. Natl. Acad. Sci. USA. 2003;100(4):1615-1620. DOI 10.1073/pnas.252770699.

Rivera V.M., Clackson T., Natesan S., Pollock R., Amara J.F., Keenan T., Magari S.R., Phillips T., Courage N.L., Cerasoli F., Jr., Holt D.A., Gilman M. A humanized system for pharmacologic control of gene expression. Nat. Med. 1996;2(9):1028-1032.

Ryu M.H., Gomelsky M. Near-infrared light responsive synthetic c-di-GMP module for optogenetic applications. ACS Synth. Biol. 2014; 3(11):802-810. DOI 10.1021/sb400182x.

Shao J., Wang M., Yu G., Zhu S., Yu Y., Heng B.C., Wu J., Ye H. Synthetic far-red light-mediated CRISPR-dCas9 device for inducing functional neuronal differentiation. Proc. Natl. Acad. Sci. USA. 2018;115(29):E6722-E6730. DOI 10.1073/pnas.1802448115.

Sternberg S.H., Redding S., Jinek M., Greene E.C., Doudna J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;507(7490):62-67. DOI 10.1038/nature13011.

Wang X., Chen X., Yang Y. Spatiotemporal control of gene expression by a light-switchable transgene system. Nat. Methods. 2012;9(3): 266-269. DOI 10.1038/nmeth.1892.

Yazawa M., Sadaghiani A.M., Hsueh B., Dolmetsch R.E. Induction of protein-protein interactions in live cells using light. Nat. Biotechnol. 2009;27(10):941-945. DOI 10.1038/nbt.1569.

Zhang F., Cong L., Lodato S., Kosuri S., Church G.M., Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat. Biotechnol. 2011;29(2):149-153. DOI 10.1038/nbt.1775.

ORCID ID

E.S. Omelina orcid.org/0000-0002-2189-5101
A.V. Pindyurin orcid.org/0000-0001-6959-0641

Acknowledgements. This work was supported by the Russian Science Foundation, grant 18-74-00050.

Conflict of interest. The authors declare no conflict of interest.

Received November 20, 2018. Revised December 11, 2018. Accepted December 11, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):226-231
DOI 10.18699/VJ19.486

The problems of genetic support of dividing the black kite (*Milvus migrans*) into subspecies

N.G. Andreyenkova[1] ✉, I.J. Starikov[2], M. Wink[2], I.V. Karyakin[3], O.V. Andreyenkov[1], I.F. Zhimulev[1, 4]

[1] Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia
[2] Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Heidelberg, Germany
[3] Sibecocenter, LLC Novosibirsk, Russia
[4] Novosibirsk State University, Novosibirsk, Russia
✉ e-mail: anata@mcb.nsc.ru

The black kite *Milvus migrans* is a common bird of prey demonstrating remarkable ecological plasticity. It inhabits a variety of habitats and is an increasingly synanthropic species. The black kite is widespread in Eurasia, Africa, Australia and adjacent islands. Palearctic kites migrate to Africa, India and China in winter, but kites of Africa and Australia are partly sedentary and partly seasonal migrants. The wide range and high mobility are the reasons of a complex population structure of the black kite. Commonly five to seven *M. migrans* subspecies are distinguished, each of which is widespread over extensive areas and has more or less an apparent phenotype. Recently, studies of genetic differences between black kite populations started to emerge. On the grounds of earlier studies of mitochondrial and nuclear genes of this species, we check whether there is a genetic support for separation of the black kite subspecies. Recent studies of some mitochondrial loci substantiate the recognition of at least the European (*M. m. migrans*), Asian (*M. m. lineatus* and *M. m. govinda*), African (*M. m. aegyptius* and *M. m. parasitus*), and Australian (*M. m. affinis*) black kite subspecies. Furthermore, the mitochondrial haplotype difference suggests that the African yellow-billed kite, including *M. m. aegyptius* and *M. m. parasitus*, should be a separate species as already proposed, or even two separate species.
Key words: black kite; *Milvus migrans*; yellow-billed kite; *M. m. aegyptius*; *M. m. parasitus*; *M. m. migrans*; *M. m. lineatus*; *M. m. govinda*; *M. m. affinis*; *M. m. formosanus*; molecular phylogeny; mitochondrial markers; subspecies.

For citation: Andreyenkova N.G., Starikov I.J., Wink M., Karyakin I.V., Andreyenkov O.V., Zhimulev I.F. The problems of genetic support of dividing the black kite (*Milvus migrans*) into subspecies. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):226-231. DOI 10.18699/VJ19.486

Проблемы генетического обоснования выделения подвидов черного коршуна (*Milvus migrans*)

Н.Г. Андреенкова[1] ✉, И.Ю. Стариков[2], М. Винк[2], И.В. Карякин[3], О.В. Андреенков[1], И.Ф. Жимулев[1, 4]

[1] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Гейдельбергский университет, Германия
[3] ООО «Сибэкоцентр», Новосибирск, Россия
[4] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ e-mail: anata@mcb.nsc.ru

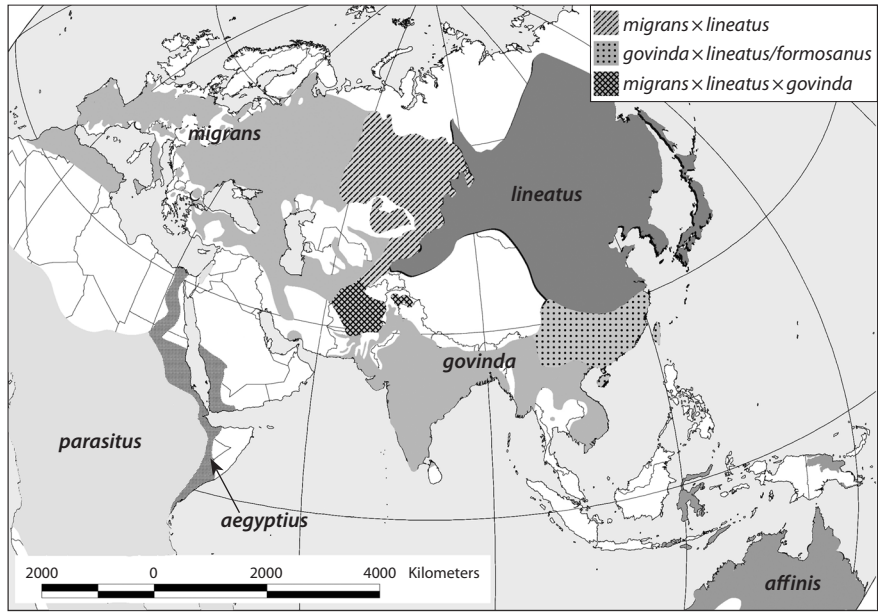
Черный коршун *Milvus migrans* – распространенный пернатый хищник, который обладает исключительной экологической пластичностью, населяет самые разные биотопы и постепенно становится все более синантропным видом. Ареал обитания черного коршуна охватывает Евразию, Африку, Австралию и прилежащие острова. Коршуны Палеарктики мигрируют на зиму в Африку, Индию и Китай, а коршуны, населяющие Африку и Австралию, частью являются оседлыми, а частью – сезонными мигрантами. Огромный ареал и высокая мобильность обусловливают сложную популяционную структуру черного коршуна. Традиционно выделяется от пяти до семи подвидов, каждый из которых населяет достаточно обширную территорию и имеет более или менее выраженные фенотипические отличия. В последние годы начала накапливаться информация о генетических различиях между популяциями черного коршуна, и эти данные не всегда подтверждают традиционные представления о филогении подвидов. В данной статье мы рассмотрим, какая информация о генетических характеристиках подвидов черного коршуна доступна на сегодняшний день.
Ключевые слова: черный коршун; *Milvus migrans*; желтоклювый коршун; *M. m. aegyptius*; *M. m. parasitus*; *M. m. migrans*; *M. m. lineatus*; *M. m. govinda*; *M. m. affinis*; молекулярная филогения; митохондриальные маркеры; подвиды.

Introduction

The black kite *Milvus migrans* is a diurnal raptor of the family Accipitridae, inhabiting a wide breeding range including continental Eurasia, Africa, Australia and several islands. This bird of prey shows remarkable ecological plasticity and inhabits a variety of habitats from the forest and forest-steppe zones to dry steppes and semi-deserts, as well as various mountain habitats. The black kite is one of the few raptors that successfully adapt to living next to humans. It intensely populates the neighborhoods of landfills, agricultural and

livestock complexes and increasingly nests even in large cities. In the 20th century, the black kite along with other raptors suffered from persecution, pesticides, and landscape changes due to human activity. Now these dangers still exist, although at a reduced scale, since poisonous chemicals and hunting are tightly controlled in European countries. Populations of many birds of prey were dangerously reduced in the second half of the 20th century, and they needed protection and restoration measures (examples are the red kite *Milvus milvus*, the bengal griffon *Gyps bengalensis*, and the peregrine falcon *Falco peregrinus*). In contrast, the black kite is quickly and independently restoring its population and rapidly spreading to new habitats, becoming a common synanthropic species.

Commonly from five to seven black kite subspecies are distinguished, each of which is widespread over an extensive range. Recently, I.V. Karyakin analyzed about 800 black kite photographs in several ornithological databases (mostly Raptors of the World in GBIF) and produced a map of subspecies distribution (Karyakin, 2017). The Palearctic is populated by two subspecies: the western part is occupied by the European black kite *M. m. migrans*, and the eastern part, by the black-eared kite *M. m. lineatus*. The kites belonging to these subspecies differ in phenotype, but their ranges overlap in Western Siberia forming a wide intergradation zone, where birds with various intermediate phenotypes occur (see the Figure). *M. m. migrans* and *M. m. lineatus* are migratory birds and spend the winter mainly in Africa and the Indian subcontinent (Ferguson-Lees, Christie, 2001). Southern and Equatorial Africa are populated by the yellow-billed kite, which is usually clearly different from the Eurasian subspecies, sometimes considered a separate species (see below) and, in turn, divided into two subspecies *M. m. parasitus* and *M. m. aegyptius*. India and the Indo-Chine countries are inhabited by the sedentary pariah kite *M. m. govinda*, and Australia, together with the adjacent islands are inhabited by the fork-tailed kite *M. m. affinis*, which migrates seasonally within the Australian continent (Marchant, Higgins, 1993). The Indian and Australian subspecies are very similar in their phenotypes, but they are considered to be strictly separated by distance. Some researchers recognize a special Taiwan kite *M. m. formosanus*, inhabiting southeastern China (see the Figure).



Breeding ranges of the black kite subspecies (Karyakin, 2017).

Other previously proposed races currently are considered as synonyms and are included in the subspecies mentioned above. In 1908, S.A. Buturlin described the subspecies *M. m. rufiventris* from Turkmenistan, later considered as a red morph of *M. m. migrans* and *M. m. ferghanensis* from Kyrgyzstan, included in *M. m. lineatus* (Dement'ev, 1936). Lectotypes and paralectotypes of both races are conserved in the Zoological Museum of Moscow University (Tomkovich, 2001). In 1923, the same author described *M. m. tianshanicus* from Kyrgyzstan mounts, which was shortly included in *M. m. lineatus*, too. Probably, it resulted from interbreeding of this taxon with nominative subspecies (Dement'ev, 1936). The next two races are currently included in *M. m. aegyptius*. *M. m. tenebrosus* was described in 1933 by Grant and Mackworth-Praed. Wintering specimens of this kite collected in 1922 (holotype) at the Ivory Coast and in 1954 in Nigeria are stored at the Natural History Museum (London) (GBIF Secretariat, 2017). *M. m. arabicus* was described by Swann in 1922, one of specimens of this proposed subspecies was collected in Lahej, Saudi Arabia, in 1899, and now it is stored at the American Museum of Natural History as a holotype (Trombone, 2013), other specimens could be found in several museums. During the examination of yellow-billed kites from Saudi Arabia and Yemen in the Natural History Museum at Tring, England, P.J. Mundy (2011) did not distinguish skins labeled *M. m. arabicus* from *M. m. aegyptius*; supposedly, the identification is not possible by their morphology. However, the nesting range of this race is situated in a semi-isolated area and the existence of this taxon can be questioned, though we should notify that the information on *M. m. arabicus* is insufficient. Data on genetic features for any of these races are not available; however, they may be of certain interest.

Population genetics of the black kite

Until now, researchers have been focused on the European black kite population. Usually the black kite is studied together with a close species, the red kite (*Milvus milvus*), which is a nearly threatened species breeding mainly in Central and Southern Europe. Roques and Negro (2005) hypothesized that the red and black kites diverged as separate species during the last glaciation in the Pleistocene, and now they are isolated reproductively, although the isolation is incomplete (Heneberg et al., 2016). These kites show sympatric speciation, the earlier archaeozoological

record of *Milvus* representative are black kite bone remains from Germany dated by 10000 ± 2000 years BP, and the remains of both species come from the same European areas since the Neolithic. In that time, coniferous boreal forests, avoided by kites, gave way to temperate-zone forests and agriculture started vigorously, resulting in wood fragmentation (Schreiber et al., 2000). In that early study, allozyme analysis of both species populations from Sachsen-Anhalt in Germany was carried out and these kites were considered very close genetically according to Nei's interspecies distance. However, the black kite is distributed much wider than the red kite and rare hybridization between these species could not apparently result in any significant changes in the black kite gene pool.

Phylogenetic studies on black kites are limited, and only a few researches analyzed relatively representative samples that allowed conclusions to be drawn about the kite population structure. Johnson et al. (2005) analyzed the *CytB* and *ND2* mitochondrial gene sequences (2146 bp in total) in 43 black and red kite samples from 27 locations, including Western Europe, Africa, South Asia, Japan, Australia, and New Guinea. Mitochondrial control region (547 bp) was also sequenced for 26 samples of the set. The authors managed to construct an informative phylogenetic tree despite a small sample size. That tree demonstrated that *CytB* and *ND2* haplotypes of the African *M. m. aegyptius* and *M. m. parasitus* differ from the haplotypes of the other *M. migrans* subspecies no less than they differ between the species *M. migrans* and *M. milvus*. According to the tree, the European *M. m. migrans* and Australian *M. m. affinis* haplotypes were found to be separate branches of the tree, but the Asian subspecies haplotypes (*M. m. lineatus* and *M. m. govinda*) were merged (Johnson et al., 2005).

The purpose of the study (Johnson et al., 2005) was to determine whether the black kites previously inhabiting the Cape Verde Islands were representatives of a particular subspecies. Therefore, almost all the samples used in this work were obtained from old museum collections and were collected mainly in the first half of the 20th century or even earlier. Since then, the black and red kite populations have experienced significant changes due to human activity. Kite populations had been reduced strongly or even disappeared completely from many areas, which was followed by repopulation (Ferguson-Lees, Christie, 2001). Thus, at least some black kite populations passed through a bottle neck and the founder effect, which certainly affected haplotype frequencies. Hence the current haplotype distribution may differ from that obtained by Johnson et al. (2005).

Scheider et al. (2004) used the mitochondrial *CytB* gene (1143 bp) to construct a phylogeny of black kites. According to this preliminary work, European *M. m. migrans* haplotypes and haplotypes of *M. m. lineatus* from Mongolia formed two separate branches, and the African *M. m. parasitus* was closer to the red kite than to the black kite. In a later research (Scheider et al., 2009) more individuals of *M. m. parasitus* were added in the analysis and most of them formed a particular clade between *M. milvus* and others *M. migrans*.

Heneberg et al. (2016) examined 184 red kite samples and 124 black kite samples from Germany and the Czech Republic. The authors studied phylogeny on the base of mitochondrial *CytB* (493 bp) and *CO1* (564 bp) gene fragments. The nuclear *Myc* (352 bp) gene was also examined, but phylogenetic trees were built for each gene separately. The authors also used sequences available in the GenBank database and concluded that the *CytB* gene is best suited for black kite subspecies differentiation, although a relatively small gene fragment was used. Ten closely related *CytB* haplotypes were found in *M. m. migrans* samples from Germany and the Czech Republic, including the most common core haplotype. *CytB* sequences of another *M. migrans* subspecies from GenBank were included to the phylogenetic tree. *M. m. lineatus* and *M. m. govinda* haplotypes were very close to each other and to *M. m. migrans* haplotypes, but Australian *M. m. affinis* haplotype formed a separate branch of the tree. Even this small *CytB* gene fragment shows that the African *M. m. parasitus* haplotype is different from the other *M. migrans* haplotypes, their genetic distances exceed those between *M. migrans* and *M. milvus* (Heneberg et al., 2016).

The same authors (Heneberg et al., 2016) found out that the nuclear *Myc* gene was not appropriate for discrimination of *M. migrans* and *M. milvus* species. All *Myc* polymorphic sites in the red kite were also polymorphic in the European black kite population. It would be impossible to determine exactly which species a particular *Myc* sequence belongs to, although the frequencies of *Myc* alleles in these two species are different. Thus, it cannot be expected that the *Myc* gene can be used to distinguish black kite subspecies. As nuclear genes have a slow evolution rate, such a result could be expected for young species.

Recently, our group published a small study of *CytB* haplotypes in the *M. m. lineatus* population from the Tyva Republic (Andreyenkova et al., 2018). We examined 20 independent black kite families and concluded that the most common haplotypes in this population coincided with the minor haplotypes found in the European *M. m. migrans* population by Heneberg et al. (2016). In addition, a *CytB* haplotype was discovered in Tuva *M. m. lineatus*, which had been found previously in India and Pakistan (*CytB-19* in terms of a study by Heneberg et al., 2016) and been assigned to the Indian *M. m. govinda*. However, since the same haplotype was found in Japan (Johnson et al., 2005), where *M. m. govinda* does not live, then the *CytB-19* haplotype unlikely belongs to *M. m. govinda*.

Genetic data on individuals of the *M. migrans* subspecies

The studies mentioned above (Scheider et al., 2004, 2009; Johnson et al., 2005; Heneberg et al., 2016) demonstrate that some mitochondrial loci allow to distinguish, at least the European (*M. m. migrans*), Asian (*M. m. lineatus* and *M. m. govinda*), African (*M. m. aegyptius* and *M. m. parasitus*), and Australian (*M. m. affinis*) black kite subspecies. Below we discuss genetic data for each subspecies in more detail.

Yellow-billed kite, *M. m. aegyptius* and *M. m. parasitus*. It is generally accepted that the yellow-billed kite includes two subspecies (*M. m. aegyptius* and *M. m. parasitus*), which breed in Northeastern and Sub-Saharan Africa, respectively (see the Figure). These kites differ from the European *M. m. migrans*, which are wintering and nesting in Africa. According to all existing data on *CytB* gene, the yellow-billed kite differs from all other black kite subspecies no less than the black kite from the red kite, and in most cases the yellow-billed and black kites form even separate clades (Wink, Sauer-Gürth, 2000, 2004;

Scheider et al., 2004, 2009; Johnson et al., 2005). In a study of West African bird phylogeny based on the *COI* mitochondrial gene, one *M. migrans* from Nigeria, probably belonging to *M. m. parasitus*, also formed a clade with *M. milvus*, and other *M. migrans* of unclear origin were more distinct (Echi et al., 2015). More comprehensive phylogenetic analysis carried by Minimal Evolution and Neighbor Joining methods with bootstrap support reveals that samples of *M. m. parasitus* from the Ivory Coast and *M. m. aegyptius* found in Kenya constitute a clade close to all other *M. migrans* (and also Cape Verde *M. milvus fasciicauda*); whereas one South African *M. m. parasitus* gets into the clade with most of *M. milvus* (Scheider et al., 2009). Nevertheless, position of yellow-billed kites with respect to other taxa is not well supported in that research (values are less than 50 %), so their exact place in kite phylogeny remains open.

Additionally, *CytB* and *ND2* genes haplotypes of *M. m. parasitus*/*aegyptius* analyzed together were clearly divided into two branches: the first one included seven samples collected in Equatorial Africa and in the south of the Arabian Peninsula, the second one included three samples from South Africa and Madagascar. These two branches were divergent from each other to the same extent as the black kite from the red kite (Johnson et al., 2005). Therefore, the mitochondrial haplotype difference suggests that the yellow-billed kite should be a separate species, as already proposed (Wink, Sauer-Gürth, 2000), or even two separate species.

Taxonomy of *M. m. parasitus*/*aegyptius* is uncertain because the yellow-billed kite is poorly investigated. Young and even adult yellow-billed kites might be sometimes confused with young black kites. In addition, the yellow-billed kite is a seasonal migrant in some areas, and it can occasionally occur outside the breeding range (Clark, Davies, 2018). All these facts lead to misidentification of kites, especially if the birds are caught during migrations. Some sources consider the yellow-billed kite as a separate species *Milvus aegyptus* and divide it into subspecies *M. a. aegyptus* and *M. a. parasitus* (Morris, Hawkins, 1998; Sinclair, Langrand, 1998; Ferguson-Lees, Christie, 2005; Gill, Donsker, 2018), but other checklists still interpret *M. m. parasitus* and *M. m. aegyptus* as black kite subspecies (Dickinson, Remsen, 2013; Clements et al., 2018; Orta et al., 2018). In any case, further research is required, including both the study of yellow-billed kite biology and molecular studies involving a large number of samples and additional mitochondrial and nuclear markers.

European black kite, *M. m. migrans.* This is the best characterized black kite subspecies in terms of both biological features and genetics. Relatively complete information about *M. m. migrans* seasonal migrations is evaluable (Panuccio et al., 2013; Literák et al., 2017). Various sequences of the mitochondrial *CytB*, *ND2*, and *CO1* genes; mitochondrial control region; nuclear *Myc* gene; and some other European black kite genes are published in the GenBank database. However, two of minor *M. m. migrans CytB* haplotypes were found as major haplotypes in the nesting *M. m. lineatus* population in Tuva Republic (Andreyenkova et al., 2018), which probably reflect a gene stream from the eastern black kite population to the western one. In any case, analysis of a sufficient sample of other subspecies is required to clarify phylogenetic relationship between *M. m. migrans* and other black kite subspecies.

Black-eared kite, *M. m. lineatus*. It has been earlier suggested to consider the black-eared kite as a separate species, *M. lineatus* (Sibley, Monroe, 1990; Zhukov, 2012; Avibase taxonomic concepts, 2017), but this proposal is inconsistent with population research. As reported by Karyakin (2017), *M. m. migrans* and *M. m. lineatus* interbreed freely and, apparently, penetrate deeper into the ranges of each other, expanding an intergradation zone. According to Scheider et al. (2004) and Andreyenkova et al. (2018), *M. m. migrans* and *M. m. lineatus* *CytB* haplotypes form two separate branches of a phylogenetic tree. A *CytB* haplotype from Japan (where *M. m. lineatus* breeds) is in the same haplogroup as samples from India and Pakistan (Johnson et al., 2005), where *M. m. lineatus*, *M. m. govinda* and, possibly, *M. m. migrans* breed (Karyakin, 2017). It is therefore unknown which subspecies this haplogroup belongs to. Recently complete mitochondrial genome of a black kite from South Korea was published (Jeon et al., 2018). This territory is included into the *M. m. lineatus* range, although the black kite is a rare species there.

We compared *CytB* and *ND2* haplotypes from this individual with the haplotypes published by Johnson et al. (2005) and found out that the Korean kite haplotype belongs to the same haplogroup as the samples from Japan, India, and Pakistan. Thus we conclude that this haplotype cannot be assigned to *M. m. govinda*, as it was done before (Heneberg et al., 2016). It is more likely that all the birds having this haplotype were representatives of *M. m. lineatus*, including individuals from India and Pakistan, especially since they were caught in winter, when *M. m. lineatus* reside in this area. On the other hand, it is still conceivable that *M. m. lineatus* and *M. m. govinda* do not differ in *CytB* sequence. There are only two studies using reliable *M. m. lineatus* samples from the Tuva Republic (Andreyenkova et al., 2018) and Mongolia (Scheider et al., 2004). Unfortunately, the latter group published only preliminary results and did not submit the *CytB* sequences to GenBank. In general, it is only known for now that *M. m. migrans* and *M. m. lineatus* differ in *CytB* sequence significantly but to a lesser degree than the species *M. milvus* and *M. migrans*. It turns out that the most widespread black kite subspecies is very poorly characterized in terms of genetics. Population biology and seasonal migrations of *M. m. lineatus* are also insufficiently understood.

Pariah kite, *M. m. govinda*. This subspecies is widespread throughout the Indian subcontinent from Pakistan in the west to Indo-China in the east (Karyakin, 2017) and is very poorly investigated, too. *M. m. govinda* is possibly intergradated with other black kite subspecies in the northwest and northeast of its range; moreover, *M. m. lineatus* spends winters in the range of *M. m. govinda* while *M. m. govinda* is considered to be sedentary. Therefore, these subspecies are difficult to distinguish despite the phenotypic differences, and in some cases the validity of subspecies identification is questionable. The only study sequencing mitochondrial genes of kites from India was performed by Johnson et al. (2005). However, all three samples from India used in that study were collected during *M. m. lineatus* wintering period. In addition, as mentioned above, their *CytB* and *ND2* haplotypes were very close to the haplotype of the sample from Japan (Johnson et al., 2005) and the sample from South Korea (Jeon et al., 2018), so we suggest that all the four samples likely belonged to

M. m. lineatus. Thereby it can be concluded that no DNA sequences of reliable *M. m. govinda* have been published so far. Phylogenetic relationships of *M. m. govinda* with other black kite subspecies remain to be established.

Fork-tailed kite, *M. m. affinis*. This subspecies is widespread in Australia, New Guinea, and Indonesia. According to current data (Karyakin, 2017) *M. m. affinis* have no contacts with other black kite subspecies; therefore, significant genetic differences between *M. m. affinis* and other subspecies could be expected, as is the case of the African yellow-billed kite. Unfortunately, very little data on the Australian black kite genetics is available. One Australian sample was used in the study of Scheider et al. (2004); based on *CytB* gene the sample was very far from the other black kite subspecies in a phylogenetic tree, both the red and black kites were ingroups. However, the authors themselves considered their result unreliable due to possible DNA degradation. This did not prevent Australian authors from the suggestion that the fork-tailed kite was the ancestor for other subspecies, and Australia was the center of black kite origin (Debus, 2005). These conclusions seem unfounded, since there is not enough information on the Australian black kite genetics. In addition, *M. m. affinis* looks very similar to Indian *M. m. govinda*. These subspecies are almost indistinguishable in phenotype, reflecting possible close relation. Beside this study, only Johnson et al. (2005) sequenced *CytB* and *ND2* genes of four *M. m. affinis* birds: (two from Australia, one from Indonesia, and one from Papua New Guinea). These sequences form a special group but belong to the same branch as the *M. m. lineatus*/*govinda* haplotypes, indicating a fairly close relationship between the Australian and Asian black kite populations (Johnson et al., 2005).

There are two short (424 and 429 bp) *CytB* sequences in Genbank database belonging to black kites (possibly the same bird, no substitutions between the sequences) from India, Andhra Pradesh (HM804897, KC439281, submitted by S. Shivaji, A.P. Reddy, and B.G. Sasikala). It is important that the samples were collected in summer when wintering *M. m. lineatus* had left India. We compared these sequences with the known black kite *CytB* sequences and found that they clearly belonged to *M. m. affinis* haplogroup from the study of Johnson et al. (2005), despite the sequences were rather short. This unexpected finding made us think that Australian black kites were not as isolated as previously thought. Another explanation is that the sequences from India belonged to *M. m. govinda* and therefore *M. m. affinis* and *M. m. govinda* are close in *CytB*.

Taiwan kite, *M. m. formosanus*. This subspecies belongs to a polytypic group of subspecies together with *M. m. lineatus* and inhabits the eastern part of China, Taiwan and Hainan. The range is located at the junction of *M. m. govinda* and *M. m. lineatus* ranges (see the Figure). Some taxonomists do not recognize *M. m. formosanus* as a separate subspecies because its phenotype is intermediate (Karyakin, 2017). It is likely that the range of this subspecies is actually an intergradation zone of *M. m. govinda* and *M. m. lineatus*. We could not find any information on genetics of this subspecies.

Conclusions

We conclude that the current knowledge of black kite phylogenetics is very far from complete, and genetic research is limited mainly to the European subspecies *M. m. migrans*. The largest amount of information in the GenBank database is related to the mitochondrial *CytB* gene. Comparison of *CytB* sequences in different black kite populations indicates that this gene allows the distinction of at least part of the subspecies (Scheider et al., 2004; Johnson et al., 2005; Heneberg et al., 2016; Andreyenkova et al., 2018). As for nuclear markers, data is insufficient: only *Myc* gene was investigated on a representative sample of red and black kites, but this sequence did not even allow to differentiate these two species (Heneberg et al., 2016).

A collection of black kite samples from the range of all subspecies and intergradation zones between subspecies is required for further study of the black kite. Mitochondrial and nuclear markers (such as microsatellites and SNPs) will clarify black kite population structure features, as well as dispersion history after the end of the Pleistocene glaciation.

References

Andreyenkova N.G., Andreyenkova O.V., Karyakin I.V., Zhimulev I.F. New haplotypes of the mitochondrial gene *CytB* in the nesting population of the Siberian black kite *Milvus migrans lineatus* Gray, 1831 in the territory of the Republic of Tyva. Doklady Biochemistry and Biophysics. 2018;482(2):242-244. DOI 10.1134/S1607672918050034.

Avibase taxonomic concepts v. 05 (Jan 2017). Black-eared kite (*Milvus lineatus*). URL: https://avibase.bsc-eoc.org/species.jsp?lang=RU&avibaseid=A691B378 on 28.11.2018.

Clark B., Davies R. African Raptors. Bloomsbury Publishing Plc, 2018.

Clements J.F., Schulenberg T.S., Iliff M.J., Roberson D., Fredericks T.A., Sullivan B.L., Wood C.L. The eBird/Clements checklist of birds of the world: v2018. Downloaded from http://www.birds.cornell.edu/clementschecklist/download/ on 16.11.2018.

Debus S. The new DNA taxonomy: implications for Australasian raptors and owls. Boobook, J. Australasian raptor assoc. 2005;23(1):4-5.

Dement'ev G.P. Order Diurnal Raptors *Falconiformes* or *Accipitres*: Full Guide of the Birds of USSR. Vol. 3. Moscow; Leningrad: KOIZ Publ., 1936;42-100. (in Russian)

Dickinson E.C., Remsen J.V. (Eds.). The Howard and Moore Complete Checklist of the Birds of the World. 4th Edn. Vol. 1. Non-Passerines. Eastbourne, U.K.: Aves Press, 2013. (Downloadable checklist). Accessed from https://www.howardandmoore.org on 16.11.2018.

Echi P.C., Suresh K.U., George S., Ratheesh R.V., Ezeonu I.M., Ejere V.C., Eyo J.E., Nwani C.D. Molecular resolution of some West African birds using DNA barcoding. Environ. Conserv. J. 2015;16:87-92.

Ferguson-Lees J., Christie D.A. Raptors of the World. Helm Identification Guides. London, UK: Christopher Helm, 2001.

Ferguson-Lees J., Christie D.A. Raptors of the World: A Field Guide. Princeton: Princeton Univ. Press, 2005.

GBIF Secretariat. 2017. *Milvus migrans* subsp. *tenebrosus* C.H.B. Grant & Praed, 1933. GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 22.11.2018.

Gill F., Donsker D. (Eds.). IOC World Bird List (v8.2). 2018. DOI 10.14344/IOC.ML.8.2.

Heneberg P., Dolinay M., Matušík H., Pfeiffer T., Nachtigall W., Bizos J., Šimčíková D., Literák I. Conservation of the red kite *Milvus milvus* (Aves: Accipitriformes) is not affected by the establishment of a broad hybrid zone with the black kite *Milvus migrans migrans* in Central Europe. PLoS One. 2016;11(7):e0159202. DOI 10.1371/journal.pone.0159202.

Jeon H.S., Myeong H., Kang S.G., Kim J.A., Lee S.H., Lee M.Y., An J. The mitochondrial genome of *Milvus migrans* (Aves, Accipi-

triformes, Accipitridae), an endangered species from South Korea. Mitochondrial DNA. Pt. B: Resources. 2018;3(2):498-499. DOI 10.1080/23802359.2018.1450678.

Johnson J.A., Watson R.T., Mindell D.P. Prioritizing species conservation: does the Cape Verde kite exist? Proc. Biol. Sci. 2005;272: 1365-1371.

Karyakin I.V. Problem of identification of Eurasian subspecies of the black kite and records of the Pariah kite in Southern Siberia, Russia. Raptors Conservation. 2017;34:49-67. DOI 10.19074/1814-8654-2017-34-49-67.

Literák I., Horal D., Alivizatos H., Matušík H. Common wintering of black kites (*Milvus migrans migrans*) in Greece, and new data on their wintering elsewhere in Europe. Slovak Raptor J. 2017;11:91-102. DOI 10.1515/srj-2017-0001.

Marchant S., Higgins P.J. *Milvus migrans* black kite. In: Handbook of Australian, New Zealand & Antarctic Birds. Vol. 2. Raptors to Lapwings. Melbourne: Oxford Univ. Press, 1993;53-62.

Morris P., Hawkins F. Birds of Madagascar: A Photographic Guide. Yale Univ. Press, 1998.

Mundy P.J. Yellow-billed kites from Arabia. Nat. Univ. of Sci. and Technol., NuSpace Instit. Repository, 2011. Available at http://ir.nust.ac.zw/xmlui/handle/123456789/562 on 24.11.2018.

Orta J., Marks J.S., Garcia E.F.J., Kirwan G.M. Black kite (*Milvus migrans*). In: del Hoyo J., Elliott A., Sargatal J., Christie D.A., de Juana E. Handbook of the Birds of the World Alive. Barcelona: Lynx Edicions, 2018. Available at https://www.hbw.com/node/52978 on 16.11.2018.

Panuccio M., Agostini N., Mellone U., Bogliani G. Circannual variation in movement patterns of the black kite (*Milvus migrans migrans*): a review. Ethol. Ecol. Evol. 2013;26(1):1-18. DOI 10.1080/03949370.2013.812147.

Roques S., Negro J.J. MtDNA genetic diversity and population history of a dwindling raptorial bird, the red kite (*Milvus milvus*). Biol. Conserv. 2005;126:41-50.

Scheider J., Wink M., Stubbe M., Hille S., Wiltschko W. Phylogeographic relationships of the black kite *Milvus migrans*. In: Chancellor R.D., Meyburg B.-U. (Eds.). Raptors Worldwide: Proc. of the VI World Conf. on Birds of Prey and Owls. World Working Group on Birds of Prey. Berlin, Germany, 2004;467-472.

Scheider J., Wink M., Wiltschko W., Stubbe M. Zur Phylogenie und Taxonomie des Schwarzmilan-Subspezies. Populationsökologie Greifvogel- und Eulenarten. 2009;6:233-241.

Schreiber A., Stubbe M., Stubbe A. Red kite (*Milvus milvus*) and black kite (*M. migrans*): minute genetic interspecies distance of two raptors breeding in a mixed community (Falconiformes: Accipitridae). Biol. J. Linn. Soc. 2000;69:351-365. DOI 10.1111/j.1095-8312.2000.tb01210.x.

Sibley G.C., Monroe B.L. Distribution and Taxonomy of Birds of the World. New Haven. Yale Univ. Press, 1990.

Sinclair I., Langrand O. Birds of the Indian Ocean islands. Cape Town: Struik Publishers, 1998.

Tomkovich P.S. Order Falconiformes. In: Types of Vertebrates in the Zoological Museum of Moscow University (Archives of the Zoological Museum of Moscow State University. Vol. XLI). Moscow: Moscow Univ. Publ., 2001;100-106.

Trombone T. AMNH Bird Collection. American Museum of Natural History. 2013. Occurrence dataset https://doi.org/10.15468/xvzdcm accessed via GBIF.org on 22.11.2018. https://www.gbif.org/occurrence/858428111 on 28.11.2018.

Wink M., Sauer-Gürth H. Advances in the molecular systematics of African raptors. In: Chancellor R.D., Meyburg B.-U. Raptors at Risk. Berlin: WWGBP, 2000;135-147.

Wink M., Sauer-Gürth H. Phylogenetic relationships in diurnal raptors based on nucleotide sequences of mitochondrial and nuclear marker genes. In: Chancellor R.D., Meyburg B.-U. (Eds.). Raptors Worldwide: Proc. of the VI World Conf. on Birds of Prey and Owls. World Working Group on Birds of Prey. Berlin and MME/BirdLife Hungary, Budapest, 2004;483-498.

Zhukov V.S. Birds of prey in the dynamic environment of the third millennium: status and prospects. In: Proc. VI Int. Conf. on Falconiformes and Owls of Northern Eurasia. Krivoi Rog, Ukraine, 2012; 18-25. (in Russian)

Acknowledgements. This work was supported by State Budgeted Project 0310-2018-0010. The authors gratefully acknowledge the resources provided by the "Molecular and Cellular Biology" core facility of the IMCB SB RAS.

Conflicts of interest. The authors declare no conflict of interest.

Received November 29, 2018. Revised December 11, 2018. Accepted December 11, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):232-238
DOI 10.18699/VJ19.487

Phylogeography and hybridization of corvid birds in the Palearctic Region

A.P. Kryukov

Federal Scientific Center of the East Asia Terrestrial Biodiversity, FEB RAS, Vladivostok, Russia
e-mail: kryukov@ibss.dvo.ru

Natural hybridization increases a lot phenotypic and genetic diversity and shapes intra-species patterns, which is a subject of phylogeography. We studied mitochondrial and complete genome variation in the bird family Corvidae, genera *Corvus*, *Pica*, *Cyanopica*, *Perisoreus* and *Nucifraga*. In the classic case of natural hybridization between carrion and hooded crows in Siberia, we found no decreased fitness of hybrids, but instead positive assortative mating which should restrict hybrid zone width. Several genetic markers were unable to discriminate between pure carrion and hooded crows. Mitochondrial DNA sequences revealed no difference between carrion and hooded crows, but instead two diverged haplogroups within the eastern part of the distribution range of the carrion crow. NGS resulted in a clear pattern of diversification of pure forms and hybrids (by using SNPs), and showed genomic regions of increased variability, the so-called "speciation islands". Comparing European and Siberian crow hybrid zones, differences in genome regions bearing genes of melanogenesis supposedly under divergent selection were found. Comparative phylogeographic analysis of 10 widely distributed Palearctic species revealed two kinds of patterns: one with a division into two haplogroups, western and eastern, and another one without such a division. These two phylogeographic patterns might be explained by different habitat preferences: mainly open fields for the first group and forests for the second one. One glacial refuge was assigned to the latter group, while west-east group species might have survived in several refuges. One of such species, the Eurasian magpie (*Pica pica*) has a gap in its range in Transbaikalia, which is currently shrinking before our eyes. The two subspecies divided by this gap differ in phenotype, mtDNA and vocalization. In their young contact zone, some hybridization occurs with small introgression limited by certain post-zygotic isolation.

Key words: phylogeography; hybrid zone; crow; magpie; speciation; genetic variation; DNA; isolation; divergence.

For citation: Kryukov A.P. Phylogeography and hybridization of corvid birds in the Palearctic Region. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):232-238. DOI 10.18699/VJ19.487

Филогеография и гибридизация врановых птиц Палеарктики

А.П. Крюков

Федеральный научный центр биоразнообразия наземной биоты Восточной Азии, Дальневосточное отделение Российской академии наук, Владивосток, Россия
e-mail: kryukov@ibss.dvo.ru

Природная гибридизация значительно увеличивает фенотипическую и генотипическую изменчивость и формирует внутривидовую структуру, изучением которой занимается филогеография. Нами исследована изменчивость митохондриального и полного геномов птиц семейства врановые Corvidae, родов *Corvus*, *Pica*, *Cyanopica*, *Perisoreus* и *Nucifraga*. Анализируя классический случай естественной гибридизации между серой и черной воронами (*Corvus corone cornix* и *C. c. corone*) в Сибири, мы не обнаружили предполагаемого уменьшения приспособленности гибридов, однако установили ассортативность скрещиваний, которая должна ограничивать ширину гибридной зоны. В результате использования нескольких подходов и генетических маркеров не найдено различий между серой и черной воронами, но в процессе секвенирования митохондриального генома установлено наличие дифференциации на два гаплотипа в пределах ареала восточной черной вороны. При полногеномном секвенировании нами впервые обнаружена явная диверсификация чистых форм и гибридов по SNP и исследованы участки генома с повышенной дифференциацией – «островки видообразования». Сопоставление геномов птиц из европейской и сибирской гибридных зон выявило частично разные участки, ответственные за меланогенез, с признаками дивергентного отбора. Сравнительная филогеография десяти видов врановых, широко распространенных в Палеарктике, позволила выделить две категории ареалов: с дифференциацией на западные и восточные группы гаплотипов и без такого разделения. Это можно объяснить различными экологическими предпочтениями: к полуоткрытым биотопам у птиц первой группы и к лесам у второй. Представители второй группы пережили плейстоценовые оледенения в предполагаемом единственном лесном рефугиуме с последующим бутылочным горлышком, что объясняет их гомогенность, а виды первой группы могли сформироваться в нескольких местах. Один из таких видов – сорока *Pica pica* – имеет разрыв ареала в Забайкалье, который в настоящее время заполняется. Принадлежащие разным гаплогруппам популяции, различающиеся по фенотипам, мтДНК и вокализации, образуют зону

контакта с ограниченной гибридизацией, но в целом репродуктивно изолированы. Таким образом, современная динамика видовых ареалов формирует филогеографическую структуру видов.
Ключевые слова: филогеография; гибридные зоны; вороны; сороки; видообразование; генетическая изменчивость; изоляция; дивергенция.

The main purpose of evolutionary biology is still investigating the processes taking place in nature. Although we generally imagine the various forces creating the incredible diversity of organisms and the mechanisms maintaining this diversity, understanding speciation in all aspects is still a challenging matter. Even if we agree that the main mechanism of divergence might be geographic isolation, other factors are disputed heavily. Among them, natural hybridization and hybrid zones were named as "natural laboratories" (Hewitt, 1988) or "windows on the evolutionary process" (Harrison, 1990). Hybrid zones can tell us a lot about the level of isolation between populations and reality of the so-called species isolating mechanisms. If hybridization between formerly isolated diverged forms occurs, this may result in increased variation and thus complicates the pattern of intraspecies' variation, which is studied by modern phylogeography. It deals with patterns and processes of distribution of genetic lineages throughout the species' ranges or particular gene genealogies (Avise, 2000) and is still in many cases based on markers of mitochondrial genomes. Phylogeography is meant to explain the biogeographic history of populations and to shed light on the formation of the species.

Within the bird family Corvidae there are several species widely distributed in the Palearctic. It includes species very variable in shape, rather large in size, numerous and observable. Thus, they are convenient subjects for evolutionary studies. Crows (*Corvus corone*) are most prominent among corvids as the hybrid zone between carrion and hooded crows is a classical text-book case of natural hybridization occured presumably by means of secondary contact in the Holocene (Mayr, 1942, 1963). This group represents the polytypic species *Corvus corone*, including the western carrion crow *C. c. corone*, eastern carrion crow *C. c. orientalis* and hooded crow *C. c. cornix* located between them (Fig. 1). The taxonomy of the forms included in this group is still a subject of hot debates and the taxa are treated as distinct species by some authors or *C. corone* is treated as a superspecies with several semispecies. The European hybrid zone is rather well studied, while the Siberian one was only discovered by starting our study.

During field trips in 1984–1991 we conducted ecological observations in the region between Ob and Yenisey rivers, mainly along the Transsiberian railway. Here the hybrid zone was ~150 km in width (Kryukov, Blinov, 1989, 1994; Blinov, Kryukov, 1992; Blinov et al., 1993). There was a clinal variation of plumage phenotypes within the zone, with the portion of intermediates (presumably hybrids) reaching 33 % at maximum in the center of the zone and decreasing to its borders. Also, phenotypic variation increased towards the center of the zone, while its edges were rather homogeneous. No ecological preferences were found for both forms, and phenotypically different pairs nested side by side. Breeding success was the same over all kinds of mating pairs. The portion of intermediates was the same in the stages of nestlings, subadults and adults. Thus, progeny of any kinds of crossings had the same fitness. In total, no selection against hybrids was observed. While checking the content of mating pairs, we noted that any kind of homogeneous and heterogeneous pairs could be met. At the same time, statistically significant positive assortative mating was discovered and proved in the center of the zone (Kryukov, Blinov, 1989; Blinov, Kryukov, 1992). It means that black-colored carrion crows mated preferentially with black-colored ones, while black-grey hooded individuals, with hooded ones. The same was confirmed for the European zone (Saino, Villa, 1992). This phenomenon was assumed to restrict the hybridization process and thus the widening of the zone.

Single phenotypic hybrids may be met even far away from the zone borders, in both allopatric initial ranges. Throughout the years we tried to use several kinds of genetic markers in order to trace the detailed structure of the zone. But no difference between *C. c. corone* and *C. c. cornix* was found at the level of karyotypes: both taxa have 80 chromosomes with identical morphology (Roslik, Kryukov, 2001). In RFLP analysis and allozymes, no difference between parental forms was found either, but several new variants of albumine, esterase and transferrine were revealed in some hybrid specimens (Kryukov et al., 1992). Genomic analyses using hypervariable DNA fingerprinting (with a minisatellite probe from bacteriophage M13) revealed no specific markers (Uphyrkina et al., 1995). Also, an RAPD-PCR analysis showed that variability in carrion and hooded crows is less than in the hybrids (Spiridonova, Kryukov, 2004).

The sequencing of the mitochondrial *cytochrome b* (*cytb*) gene revealed no difference between carrion and hooded crows (so this marker is inappropriate for further detailed study of the hybrid zone), but unexpectedly we discovered a clear difference within the eastern carrion crows in a geographic context (see Fig. 1).While homogeneous over the most part of the distribution, from West Europe to Kamchatka, haplotypes in the South Far East differed considerably (Kryukov, Suzuki, 2000). Further analysis of the mitochondrial CR revealed a more detailed pattern, with three haplogroups, the most derived being the group occupying the extreme southeastern part of the range: the Russian Far East, Korea and Japan (Haring et al., 2007). Interestingly, in East Siberia, Kamchatka, and North Sakhalin, a haplogroup slightly differentiated from the western group was found, reflecting probably additional refugia. Then, in the distinct species, *Corvus pectoralis* from China, we found the same haplotype as in southeasternmost populations of *C. corone* (Haring et al., 2012). This may imply mitochondrial capture from one species to another due to suspected ancient hybridization.

Crows became a subject for geneticists from other countries as well. Haas et al. (2009) found no difference between carrion and hooded crows in their study of *MC1R* gene, the candidate for melanogenesis, and most microsatellite loci. The same result was obtained in sequencing analyses of 25 nuclear intronic loci, but different expression profiles between carrion

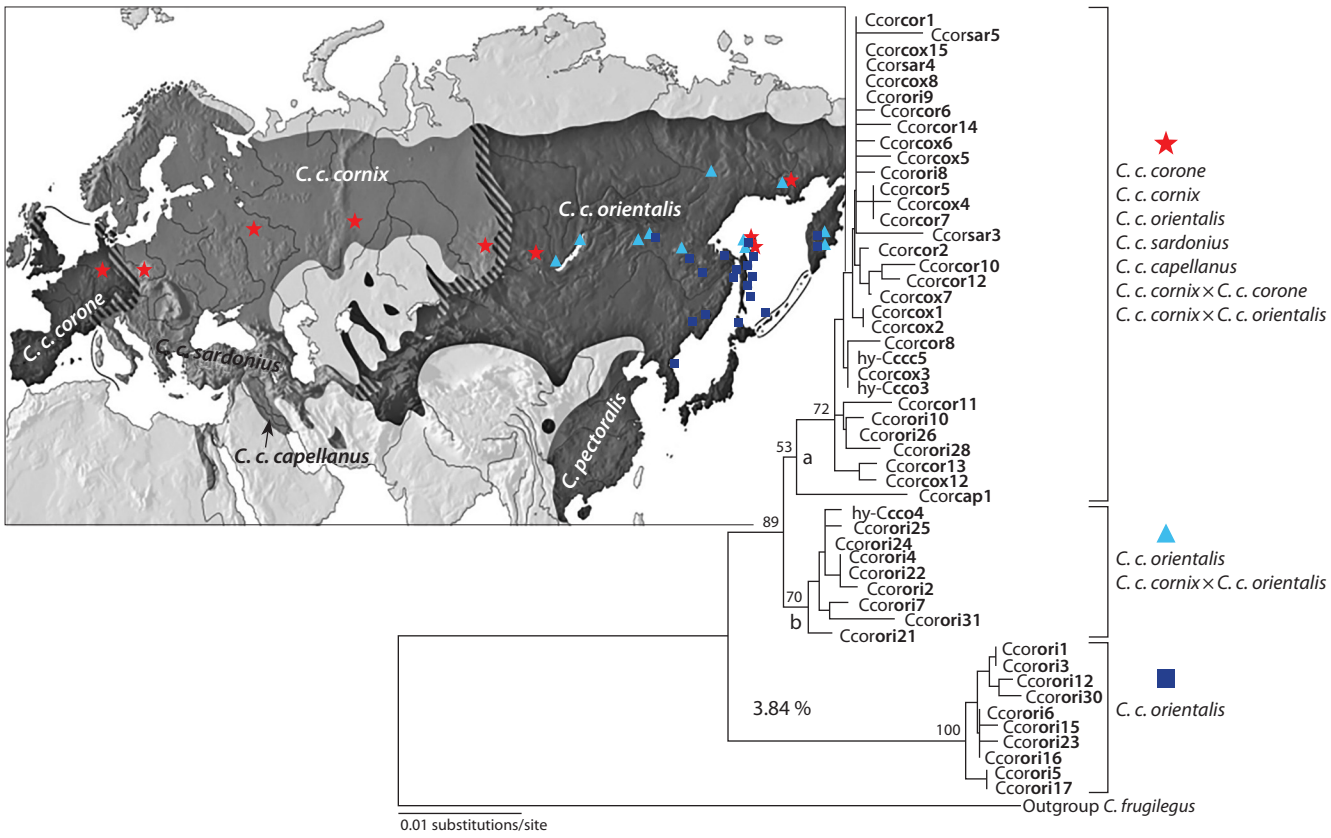


Fig. 1. Distribution of *Corvus corone* s. l. with hybrid zones shaded and collecting localities; and NJ tree based on mitochondrial Control Region sequences (modified after Haring et al., 2007).

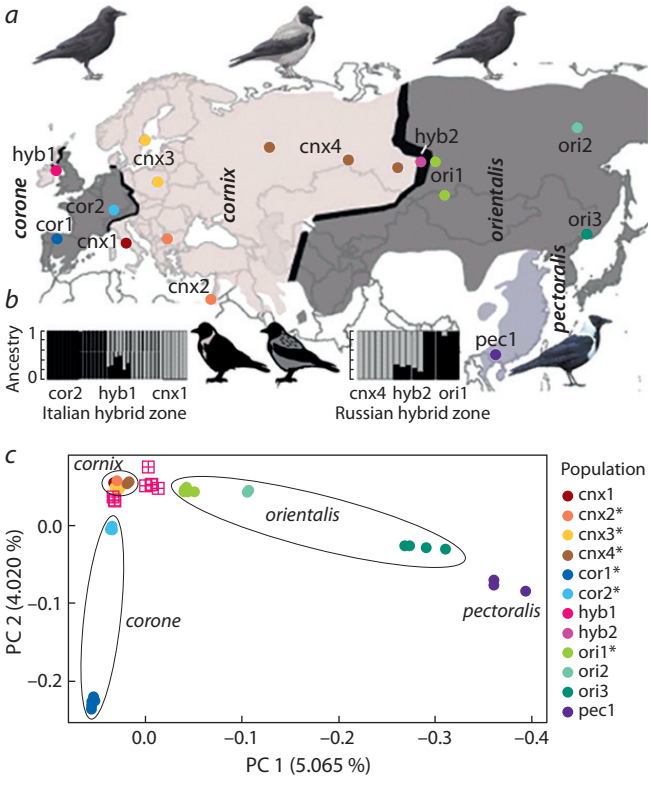


Fig. 2. *a*, The ranges of *Corvus corone* complex and collecting localities; *b*, Phenotypic content of the hybrid zones and adjacent populations; *c*, PCA-analysis by 16.6 million SNPs (modified after Vijay et al., 2016).

and hooded crows were found in some of 1301 core genes analyzed (Wolf et al., 2010). In total, all these data support the view that the three forms are very closely related and have diverged recently, only thousands of years ago. From a taxonomic view, this supports their subspecies status: *Corvus corone corone*, *C. c. orientalis* and *C. c. cornix*.

The crow hybrid zones were studied in a more general aspect as a model to understand the genetic aspects of speciation via the genome landscapes approach, specifically, to reveal a connection between gene flow across the hybrid zones and diversification visualized clearly by feather coloration (very different in carrion and hooded crows). It was found that genes associated with melanogenesis pathways and hormone regulation are expressed differently in carrion and hooded crows and are located in a very small genomic region of the so-called "speciation islands". It was suspected that in this genomic region there is an inversion. These genes were supposed to be responsible for maintaining plumage differentiation of the crossing forms in spite of ongoing, though restricted, gene flow (Poelstra et al., 2014).

Because the "speciation islands" are subject to divergent selection, they would cross the hybrid zones slower than neutral alleles. So the next genome-wide study was aimed to compare the processes in both European and Siberian hybrid zones. 124 complete genomes were analyzed and genomic regions with increased variability were found. In a PCA analysis by SNPs, we for the first time found clear differences between hooded and both carrion crow clusters, while hybrids appear in between them (Fig. 2). While comparing allopatric

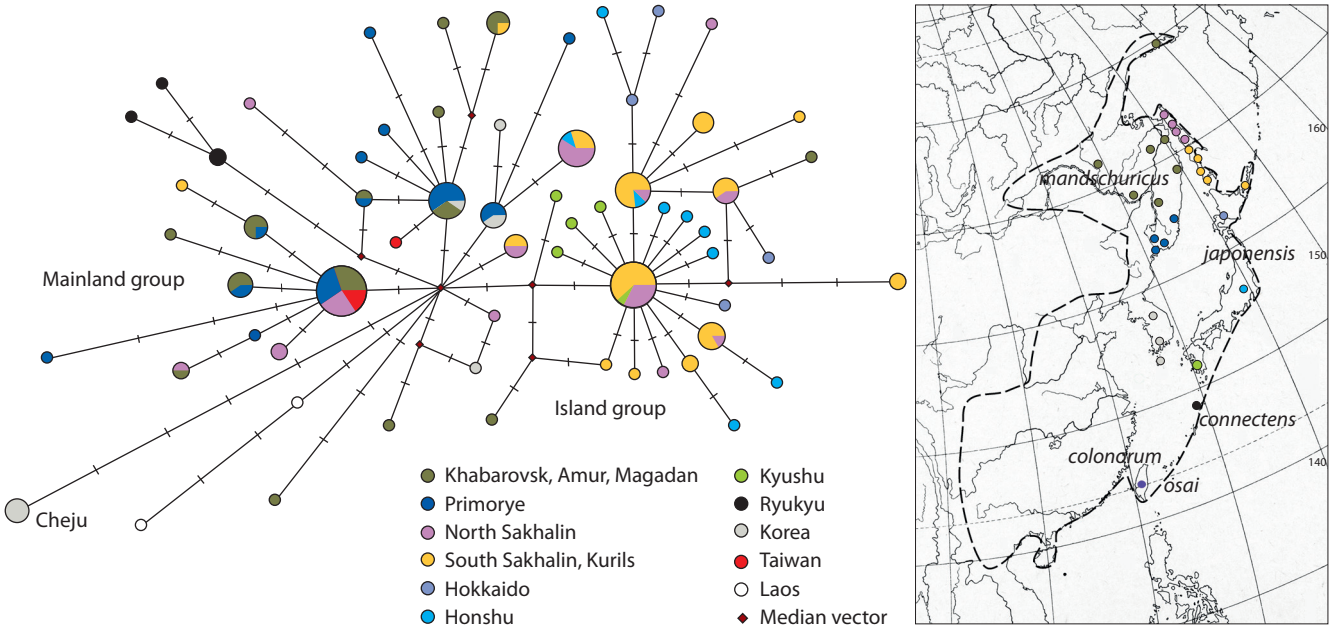


Fig. 3. Phylogenetic network based on *cytb* gene of mtDNA of *Corvus macrorchynchos*, and collecting localities (after Kryukov et al., 2012).

populations and hybrid zones by Fst in a "sliding window" method, several genomic regions of increased diversity were discovered. These local peaks of genomic differentiation in the Siberian hybrid zone were rather low compared with those of the European zone, which corresponds to their widths. The main such peaks were located at the 18th chromosome in the case of the European zone and 21st for the Siberian one. Accordingly, the candidate genes of the metabolic pathway of melanogenesis that were supposed to be involved in selection against gene flow were only partly shared among both zones, suggesting their independence (Vijay et al., 2016).

Another crow species, black-colored, with a massive bill, is the jungle crow (*Corvus macrorhynchos*), which inhabits Eastern Asia. We studied patterns of its genetic variation by using sequences of the mitochondrial *cytochrome b* gene. *C. macrorhynchos* demonstrates a high level of sequence variation with a tendency to geographic differentiation throughout its range. The haplotype network shows two haplogroups (Fig. 3). The island group comprises the populations of Sakhalin, Hokkaido, Honshu and Kyushu, while the mainland group includes the populations from Primorye, Khabarovsk, Amur, Magadan regions and Laos. Haplotypes of Taiwan and Ryukyu Islands proved to be closer to the mainland group. The population of Cheju Island (Korea) contains only one haplotype which is quite distant. This pattern allowed us to develop a phylogeographic hypothesis regarding the two ways of settling of the Japanese island from the mainland: one from Sakhalin and Hokkaido until Kyushu, and the other from the South via Taiwan or, less likely, from the Korean Peninsula. Interestingly, the lowest level of nucleotide diversity was found within the marginal north-west population in the western part of Amur region. *C. macrorhynchos* experienced population growth in the recent past and is still expanding its range, at least in the north and north-west of the range (Kryukov et al., 2012).

The azure-winged magpie (*Cyanopica cyanus*) represents the well-known case of a widely distinct, disjunctive range. It is distributed on the Iberian Peninsula in the west and on the Eastern Asia. Both groups of subspecies, western and eastern, differ from each other by color and size. Usually these isolates were treated as Tertiary relics, assuming the subdivision of a former continuous range during the Pleistocene. However there was a hypothesis that Portuguese or Spanish sailors might have brought these nice birds from Asia (China or less probably Japan) in the Medieval and released them in Europe. In such a case, the genetic affinity between the European and Asian populations would be lower than that between the Chinese and Japanese populations.

We performed analyses of the mitochondrial CR and the *cytb* gene from seven subspecies throughout the *C. cyanus* range (Kryukov et al., 2004). The results of the phylogenetic analyses reveal a clear pattern concerning the grouping of haplotypes derived from the eastern and western isolates of the azure-winged magpie. All kinds of trees, NJ, ML and MP, represent the similar subdivision for two clades, with the p-distance between them about 5.3 % for CR and 6.3 % for *cytb*. Within the clades, differentiation is very low (0.2–0.5 %). Applying the widely used value of 2 % divergence per myr (Klicka, Zink, 1997) to our *cytb* data set results in considerably high divergence times, e. g. >3 mya between Asian *C. c. cyanus* and European *C. c. cooki*. According to that rate, the phenotypic and genotypic differentiation of the azure-winged magpie had already started in the Pliocene, whereas repeated glacial expansions and restrictions of distribution ranges during the Pleistocene might have completed the process and enforced the differences.

Our multivariate morphometric analysis showed a clear differentiation between the Iberian azure-winged magpie and the Asian ones, which proved to be different in several measurements and in the space of two main canonical roots. The morphological data are in accordance with the genetic results (Fok et al., 2002; Kryukov et al., 2004), where sequence divergence between the western and eastern isolates was found to be

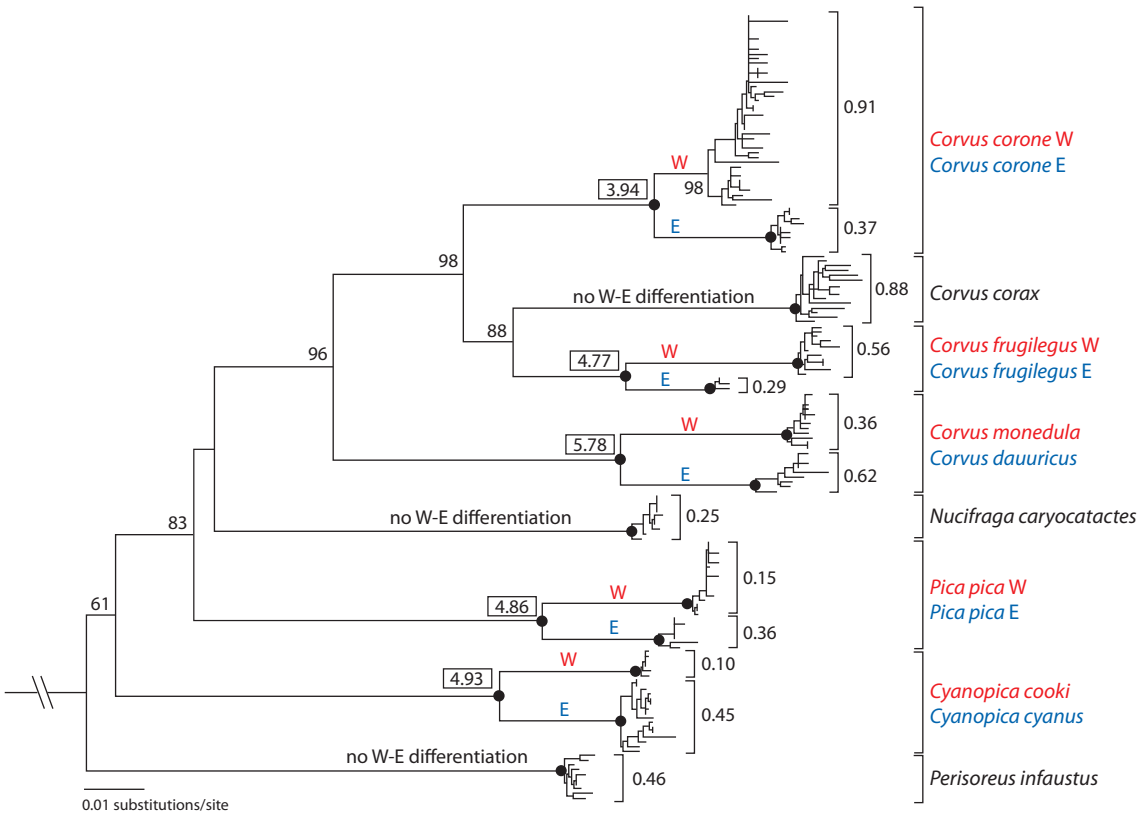


Fig. 4. Comparative phylogeography of 10 corvid species: ML tree based on CR of mtDNA (after Haring et al., 2007).

much higher than that among the Asian populations. Both data sets clearly reject the hypothesis of a recent introduction of azure-winged magpies into Europe. Since the morphological characters mainly represent the nuclear genome, this analysis can be regarded as an important complement to the genetic investigations, which altogether corroborate the suggested subdivision into the two species: *Cyanopica cyanu*s Pallas, 1776 and *Cyanopica cooki* Bonaparte, 1850.

In addition to the above results, we obtained some more data from several corvid bird species by sequencing mitochondrial CR. Besides *Corvus corone* and *Cyanopica cyanus*,we found deep splits into two lineages within each of the following species: *Corvus frugilegus*, *Pica pica* and between the species pair *Corvus monedula – Corvus dauuricus* (formerly treated as conspecific). Each of them is divided into two haplogroups corresponding to the western and eastern parts of their range. In contrast to the taxa with the west-east pattern, no differentiation into clearly divergent lineages was detected in three other taxa: the raven (*Corvus corax*), the Siberian jay (*Perisoreus infaustus*) and the nutcracker (*Nucifraga caryocatactes*) (Fig. 4).

It is notable that in each of these species distances between the western and eastern groups are similar, 4–5.8 %, which implies nearly the same timing of diversification events. Sequence diversity is rather low within each subclade and in the three single-group species. To explain these two types of phylogeographic pattern, one has to take into account ecological conditions which may have accompanied the climatic changes during the Pleistocene. We noted that all species of the first group preferably nest in semi-open habitats and forest edges, while the second group (single-group species) live mainly in forests. The raven (*C. corax*) is an ubiquist. We arrived at the conclusion that a prominent factor influencing the pattern of genetic differentiation seems to be the preference for either open to semi-open habitats (the west-east pattern) or forest dominated habitats (the single group pattern). Separated refuge areas (western and eastern) during cold periods led to accumulation of diversity. There could be several such open field refuges, but only one, namely, Altai-Sayan main forest refuge is known for LGM (Nazarenko, 1982; Anijalg et al., 2018) in which *P. infaustus* and *N. caryocatactes* might have survived (and probably suffered severe bottlenecks), followed by rapid spreading after the end of the Pleistocene. It seems that the two genetic patterns mentioned are mainly associated with different habitat requirements and to some extent with social and breeding behaviour. The comparatively high genetic diversity found in *C. corax* suggests that this species did not pass through severe bottlenecks as others. As a generalist, it could even have remained more widely distributed. Thus, comparative phylogeography helps get a general view on the interaction between macroecological features and patterns of the genetic variation.

One of the most common corvid species, the Eurasian magpie (*Pica pica* Linnaeus, 1758), is widely distributed throughout most Holarctic, with several isolated occurrences in Kamchatka, South Africa, North America and Saudi (Fig. 5). In the centre of the range, in Transbaikalia, there is a gap. Analyses of sequences of the mtDNA revealed deep genetic splits into four main lineages: (1) group West (European-Siberian), (2) group East (southern Far East), (3) *P. p. mauritanica* (North Africa) and (4) *P. p. hudsonia*

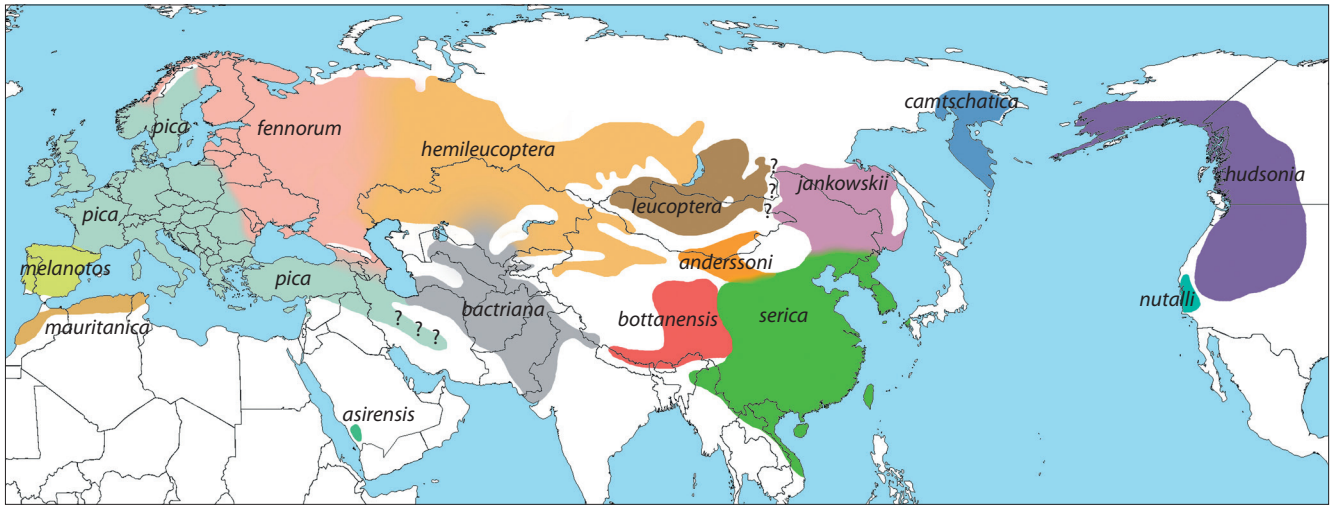


Fig. 5. Distribution of magpie *Pica pica* and its subspecies (after Kryukov et al., 2017).

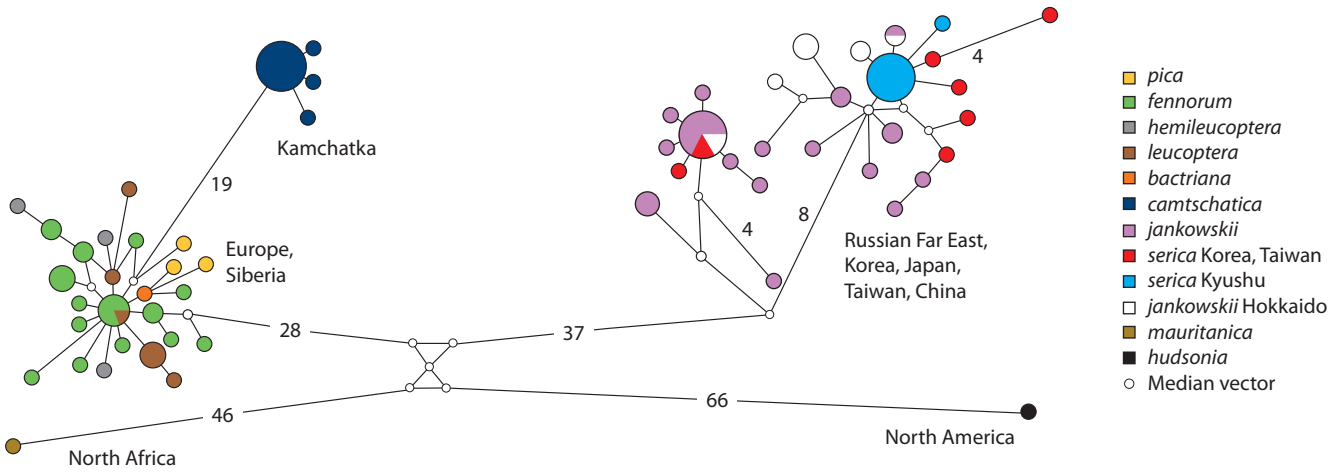


Fig. 6. Phylogenetic median-joining network for *Pica pica* based on CR of mtDNA (after Kryukov et al., 2017).

(North America) (Kryukov et al., 2017). These lineages show a geographic pattern corresponding to the subspecies or groups of subspecies with distinct phenotypic features.

In the phylogenetic network, the western subspecies group differs from the eastern one by more than 66 substitutions (Fig. 6). Differentiation within the widely-distributed group West is rather small, implying recent expansion. We detected a separate lineage in Kamchatka with clear affinity to this Euro-Siberian group. The Kamchatka lineage is homogeneous too, implying a recent bottleneck. Group East consists of two subclades without clear geographic pattern, presumably due to admixing of populations which diverged in Pleistocene refuges. The homogeneity of the Kyushu population supports historical reports on the species' introduction from Korea. The high variation in the recently established Hokkaido population can be explained by an ongoing ship-assisted invasion from the variable populations of the Far Eastern mainland. Bioacoustic data reflect phylogeographic patterns, i. e. the split into mt lineages, and differentiate groups of subspecies. Overall, our data support a scenario of divergence in geographic isolation. Furthermore, we report for the first time the fast spreading of *P. p. jankowskii* towards the west along the upper Amur River, and a slower shifting of *P. p. leucoptera* in the opposite direction, thus yielding a new contact zone in Transbaikalia. Here the density of the Argun' river population, bearing admixture of the both mitochondrial haplogroups, is very low, but we found a few mixed pairs and phenotypical intermediates. Breeding success was dramatically decreased and in many nests only unfertilized eggs were found. This may imply sterility of F1 hybrids. This case may serve as a sample of a reproductive isolation in the zone of the secondary contact of the forms diverged by phenotypes, DNA and vocalization when they were isolated geographically.

Here I shortly overlooked several cases of intra-species genetic variation of corvid birds widely distributed in the Palearctic. The degree of their divergence depends on historical and biogeographic factors. Most species demonstrate an allopatric mode of speciation. Pleistocene cyclic glaciations and interglacial periods have greatly influenced the species distribution ranges and interaction of forms, some of them were shaped due to repeated isolation in Pleistocene refugia. In the case when distribution ranges came to the secondary contacts after spreading, successful introgression within the narrow

hybrid zones occurs (*Corvus c. cornix* with *C. c. corone* and *C. c. orientalis*), or semisympatry with reproductive isolation (*Pica p. leucoptera* and *P. p. jankowskii*). Timing of the initial divergence and secondary contacts is hard to determine and is always disputable, mainly due to the lack of reliable calibration by fossil records in birds. Intra-species taxonomy, traditionally based on morphology, often does not correspond with DNA sequence data. However, mtDNA sequences still represent a useful tool for phylogeographic aims, even when using short DNA regions, provided that sampling covers the main part of the species distribution range.

References

Anijalg P., Ho S.Y.W., Davison J., Keis M., Tammeleht E., Bobowik K., Tumanov I.L., Saveljev A.P., Lyapunova E.A., Vorobiev A.A., Markov N.A., Kryukov A.P., Kojola I., Swenson J.E., Hagen S.B., Eiken H.G., Paule L., Saarma U. Large-scale migrations of brown bears in Eurasia and to North America during the Late Pleistocene. J. Biogeogr. 2018;45:394-405. DOI 10.1111/jbi.13126.

Avise J.C. Phylogeography: The History and Formation of Species. Harward Univ. Press, 2000.

Blinov V.N., Blinova T.K., Kryukov A.P. Interactions of Hooded and Carrion Crows (*Corvus cornix* L., *C. corone* L.) in their zone of sympatry and hybridization: the zone structure and possible isolating factors. In: Rossolimo O.L. (Ed.). Hybridization and Species Problem in Vertebrates. Moscow: Moscow State University Publ., 1993; 97-117. (in Russian)

Blinov V.N., Kryukov A.P. Evolutionary stability of hybrid zones: assortative mating instead of elimination of hybrids in Carrion and Hooded Crows. Doklady Akademii Nauk = Proceedings of the Russian Academy of Sciences. 1992;325(5):1085-1087. (in Russian)

Fok K.W., Wade C.M., Parkin D.T. Inferring the phylogeny of disjunct populations of the azure-winged magpie *Cyanopica cyanus* from mitochondrial control region sequences. Proc. R. Soc. Lond. B. 2002; 269:1671-1679. DOI 10.1098/rspb.2002.2057.

Haas F., Pointer M., Saino N., Brodin A. An analysis of population genetic differentiation and genotype-phenotype association across the hybrid zone of carrion and hooded crows using microsatellites and *MC1R*. Mol. Ecol. 2009;18:294-305. DOI 10.1111/j.1365-294X.2008.04017.x.

Haring E., Däubl B., Pinsker W., Kryukov A., Gamauf A. Genetic divergences and intraspecific variation in corvids of the genus *Corvus* (Aves: Passeriformes: Corvidae) – a first survey based on museum specimens. J. Zool. Syst. Evol. Res. 2012;50(3):230-246. DOI 10.1111/j.1439-0469.2012.00664.x.

Haring E., Gamauf A., Kryukov A. Phylogeographic patterns in widespread corvid birds. Mol. Phylogenet. Evol. 2007;45(3):840-862. DOI 10.1016/j.ympev.2007.06.016.

Harrison R.G. Hybrid zones: windows on evolutionary process. Oxford Surv. Evol. Biol. 1990;7:69-128.

Hewitt G.M. Hybrid zones: natural laboratories for evolutionary studies. Trends Ecol. Evol. 1988;3:158-167.

Klicka J., Zink R.M. The importance of recent ice ages in speciation: a failed paradigm. Science. 1997;277:1666-1669.

Kryukov A.P., Blinov V.N. Interaction of Hooded and Carrion crows (*Corvus cornix* L., *C. corone* L.) in the zone of sympatry and hybridization: is there selection against hybrids? Zhurnal Obshchey Biologii = Journal of General Biology. 1989;50(1):128-135. (in Russian)

Kryukov A.P., Blinov V.N. Hybrid zone of Hooded and Carrion crows in Siberia. J. Ornithologie. 1994;135:47.

Kryukov A., Iwasa M.A., Kakizawa R., Suzuki H., Pinsker W., Haring E. Synchronic east-west divergence in azure-winged magpies (*Cyanopica cyanus*) and magpies (*Pica pica*). J. Zool. Syst. Evol. Res. 2004;42:342-351.

Kryukov A., Spiridonova L., Mori S., Lobkov E., Redkin Ya., Goroshko O., Arkhipov V., Haring E. Deep phylogeographic breaks in magpie *Pica pica* across the Holarctic: concordance with bioacoustics and phenotypes. Zool. Sci. 2017;34:185-200. DOI 10.2108/zs160119.

Kryukov A., Spiridonova L., Nakamura S., Haring E., Suzuki H. Comparative phylogeography of the two crow species, Jungle crow *Corvus macrorhynchos* and Carrion crow *Corvus corone*. Zool. Sci. 2012;29(8):484-492. DOI 10.2108/zsj.29.484.

Kryukov A.P., Suzuki H. Phylogeography of carrion, hooded and jungle crows (Aves, Corvidae) inferred from partial sequencing of the mitochondrial *cytochrome b* gene. Genetika = Genetics. 2000;36:922-929. (in Russian)

Kryukov A.P., Uphyrkina O.V., Chelomina G.N. Analysis of the Crow genomes (Corvidae, Passeriformes) from the zone of overlapping areas and hybridization. Genetika = Genetics. 1992;28(6):136-140. (in Russian)

Mayr E. Systematics and The Origin of Species. New York: Columbia Univ. Press, 1942.

Mayr E. Animal Species and Evolution. Cambridge, Mass: The Belknap Press; Harward Univ. Press, 1963.

Nazarenko A.A. On faunistic cycles (extinction – expansion – extinction…) with special reference to the East Palearctic dendrophilous avifauna. Zhurnal Obshchey Biologii = Journal of General Biology. 1982;43(6):823-835. (in Russian)

Poelstra J.W., Vijay N., Bossu C.M., Lantz H., Ryll B., Müller I., Baglione V., Unneberg P., Wikelski M., Grabherr M.G., Wolf J.B.W. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science. 2014;344:1410-1414. DOI 10.1126/science.1253226.

Roslik G.V., Kryukov A.P. A karyological study of some Corvine birds (Corvidae, Aves). Genetika = Genetics. 2001;37(7):796-806. (in Russian)

Saino N., Villa S. Pair composition and reproductive success across a hybrid zone of Carrion Crows and Hooded Crows. Auk. 1992;109: 543-555.

Spiridonova L.N., Kryukov A.P. Genetic variability of carrion and hooded crows and their hybrids according to RAPD-PCR data. Tsitologiya i Genetika = Cytology and Genetics. 2004;38(2):31-39. (in Russian)

Uphyrkina O.V., Vasiliev V.A., Kryukov A.P., Ryskov A.P. Genomic fingerprints in crows: a study of the genetic structure of populations of the hybrid zone. Genetika = Genetics. 1995;31(7):753-757. (in Russian)

Vijay N., Bossu C.M., Poelstra J.W., Weissensteiner M.H., Suh A., Kryukov A.P., Wolf J.B.W. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat. Commun. 2016;7:13195. DOI 10.1038/ncomms13195.

Wolf J.B.W., Bayer T., Haubold B., Schilhabel M., Rosenstiel P., Tautz D. Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. Mol. Ecol. 2010;19:162-175. DOI 10.1111/j.1365-294X.2009.04471.x.

ORCID ID

A.P. Kryukov orcid.org/0000-0001-7010-7338

Acknowledgements. This review was based on collaboration with many colleagues who contributed to field work as well as to laboratory work. Among them, I am much grateful for Drs. V. Blinov, O. Goroshko, L. Spiridonova, E. Haring, H. Suzuki, J. Wolf, and many others who collected samples for analyses. I am obliged to Dr. E. Haring for editing the manuscript. The work was supported by the Russian Foundation for Basic Research, project #16-04-01304.

Conflict of interest. The author declares no conflict of interest.

Received November 22, 2018. Revised December 12, 2018. Accepted December 13, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):239-243
DOI 10.18699/VJ19.488

A meiotic mystery in experimental hybrids of the eastern mole vole (*Ellobius tancrei*, Mammalia, Rodentia)

V.G. Tambovtseva[1] ✉, S.N. Matveevsky[2], A.A. Kashintsova[2], A.V. Tretiakov[2], O.L. Kolomiets[2], I.Yu. Bakloushinskaya[1] ✉

[1] Koltzov Institute of Developmental Biology, RAS, Moscow, Russia
[2] Vavilov Institute of General Genetics, RAS, Moscow, Russia
✉ e-mail: lynx1994@gmail.com; i.bakloushinskaya@idbras.ru

Chromosomal rearrangements can lead to the formation of new stable karyotypes, nevertheless changing the architectonics of the nucleus. The differences in locations might promote Robertsonian (Rb) translocations and encourage meiotic drive in favour of changed chromosomes or against them. We hypothesized that hybridization and meiotic drive may produce new chromosomal forms in *Ellobius tancrei*. We crossed two forms with 2*n* = 50, and two pairs of different Rb metacentrics with partial (monobrachial) homology. In 10 years of inbred crossings (sister–brother), we got 9 generations of hybrids (262 litters, 578 animals). In the first hybrid generation, two trivalents, a tetravalent and 20 bivalents were revealed at meiotic prophase I. Hybrids of the first generation had lower fertility, fertility increased starting from the third generation. Instead of returning to parental karyotypes, starting from the second generation, hybrids obtained new chromosome sets, with different 2*n* (48, 49, 51, 52) and combinations of Rb metacentrics. Analysis of F4, F7 and F9 hybrids revealed that synapsis of homologous parts take place despite the presence of heterozygotes and monobrachial homology of Rb metacentrics. The most common meiotic disturbance was delayed synapsis, which resumed later compared to the homologous crossings. The late synaptic adjustments nevertheless provide a proper segregation of chromosomes and normal sets in the gametes. Therefore, some cells pass through meiosis successfully and promote viable gametes. We proved the hypothesis that origin of monobrachially homologous Rb translocations may lead to divergence in several generations, due to meiotic drive.
Key words: Robertsonian translocations; synaptonemal complex; hybridization; meiotic drive.

For citation: Tambovtseva V.G., Matveevsky S.N., Kashintsova A.A., Tretiakov A.V., Kolomiets O.L., Bakloushinskaya I.Yu. A meiotic mystery in experimental hybrids of the eastern mole vole (*Ellobius tancrei*, Mammalia, Rodentia). Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):239-243. DOI 10.18699/VJ19.488

Особенности мейоза у экспериментальных гибридов слепушонок (*Ellobius tancrei*, Mammalia, Rodentia)

В.Г. Тамбовцева[1] ✉, С.Н. Матвеевский[2], А.А. Кашинцова[2], А.В. Третьяков[2], О.Л. Коломиец[2], И.Ю. Баклушинская[1] ✉

[1] Институт биологии развития им. Н.К. Кольцова Российской академии наук, Москва, Россия
[2] Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия
✉ e-mail: lynx1994@gmail.com; i.bakloushinskaya@idbras.ru

Хромосомные перестройки могут приводить к формированию новых устойчивых кариотипов, изменяя при этом архитектонику ядра. Различия в локализации хромосом, вероятно, обусловливают неслучайность робертсоновских (Rb) транслокаций и влияют на мейотический драйв как в пользу измененных хромосом, так и против них. Мы предположили, что благодаря гибридизации и мейотическому драйву могут возникать новые хромосомные формы у слепушонок *Ellobius tancrei*. Был поставлен эксперимент по гибридизации двух форм с одинаковым 2*n* = 50, но с разными робертсоновскими метацентриками. За десять лет инбредных скрещиваний (сестра–брат) получено девять поколений гибридов (262 помета, 578 животных). У гибридов первого поколения сохранялось диплоидное число 2*n* = 50, но в силу неполной (монобрахиальной) гомологии в профазе мейоза I помимо 20 бивалентов были выявлены два тривалента и один тетравалент. Гибриды первого поколения имели более низкую плодовитость по сравнению с родительскими формами. Начиная с третьего поколения плодовитость повышалась. Вместо возвращения к родительским кариотипам гибриды второго и последующих поколений получили новые наборы хромосом, отличающиеся диплоидным числом (48, 49, 51, 52) и комбинациями робертсоновских метацентриков. Анализ мейоза у гибридов F4, F7 и F9 показал, что гомологичный синапсис хромосом может быть затруднен из-за наличия гетерозигот и монобрахиальной гомологии, однако некоторые клетки успешно проходят мейоз, что приводит к формированию жизнеспособных гамет и рождению гибридов. Проведенный нами эксперимент воспроизводит процессы, происходящие в естественных условиях,

и дает основания считать мейотический драйв основным механизмом, обеспечивающим диверсификацию и быструю фиксацию возникающих хромосомных форм в природе.
Ключевые слова: робертсоновские транслокации; синаптонемный комплекс; гибридизация; мейотический драйв; видообразование.

Introduction

The role of chromosomal rearrangements in speciation is still a matter of considerable debate (Faria, Navarro, 2010; Dobigny et al., 2017). It is well illustrated in a large number of studies that chromosomal rearrangements can lead to the formation of new balanced karyotypes, nevertheless changing the architectonics of the nucleus and its functioning (Qumsiyeh, 1999; Shapiro, 2002; Graphodatsky et al., 2011; Romanenko et al., 2018). However, direct mechanisms of new karyotype fixation are still in the scale. In general terms, a transformation of chromatin structures may alter the genetic system of the species, due to the modulation accessibility of transcription factors to DNA binding sites, thus regulating gene expression. A concept of chromosome territories proposes a non-random distribution of chromosomes in nuclei; the nuclear architecture constitutes the basis for gene expression regulation (Cremer T., Cremer C., 2001).

Robertsonian (Rb) translocations join two acrocentrics into one metacentric chromosome. This is the most common type of chromosomal rearrangements in mammals (King, 1993). Translocations restructure the organization of the nuclei, especially when chromosomal territories of fused acrocentrics are located far from each other (Berríos et al., 2017). These changes unavoidably influence the hybrids' fertility because of different Rb fusions inherited from parents. In the case of partial, or monobrachial, homology, which originated by the combination of different acrocentrics in Rb biarmed chromosomes, meiosis in hybrids should become more complicated (Baker, Bickham, 1986). These hybrids' gametes reveal the formation of complex chains in meiotic prophase I. As a result, we observe the reduction of fertility in hybrids, and this may be treated as a starting point of full or partial reproductive isolation. That is why chromosomal rearrangements, including Rb translocations, are considered to be some of the mechanisms of speciation (King, 1993).

Fixation of new karyotype variations might occur by non-random chromosomal segregation, for example, when a new Rb metacentric fixes in generations instead of homologous acrocentric chromosomes. This may be a perfect illustration for a conception of meiotic drive (Sandler, Novitski, 1957; de Villena, Sapienza, 2001; Lindholm et al., 2016). Previously it was believed that monobrachial homology causes full reproductive isolation and leads to complete sterility of hybrids. The reason was that chromosome disjunction after multivalent formation cannot be balanced and resulted in aneuploid gametes and inviable zygotes production (Baker, Bickham, 1986). However, recently it was shown for different groups of animals that species and forms with monobrachially homologous chromosomes are naturally occurring (Nunes et al., 2011; Potter et al., 2017).

Our research was focused on the eastern mole vole (*Ellobius tancrei* Blasius, 1884, Rodentia, Mammalia) noted for wide karyotypic variability caused by Rb translocations (Vorontsov et al., 1980; Lyapunova et al., 1984, 2010), unique sex chromosomes (XX♂/XX♀) and an enigmatic sex determination system (Bakloushinskaya, Matveevsky, 2018). The ancestral karyotype of this species contains 52 acrocentric chromosomes and a pair of submetacentrics, but the diploid number can be reduced to $2n = 30$ by Rb translocations and fixation of new Rb metacentrics (Bakloushinskaya et al., 2013). Our study of natural variability in *E. tancrei* enables us to hypothesize that hybridization and meiotic drive may produce new chromosomal forms, and the aim of the present study was an evaluation of this hypothesis experimentally.

Material and methods

For the experimental crossing, we chose two individuals of *E. tancrei* from a natural habitat, and their progeny, a line of strictly inbred hybrids F1–F9. We crossed two forms with $2n = 50$, and two pairs of Rb metacentrics: 2Rb(4.12) and 2Rb(9.13), nicknamed 'Khodza Obi-Garm', according to the closest settlement in the Varzob River Valley, and 2Rb(2.18) and 2Rb(5.9), in the form which was named 'Voidara', according to the closest settlement in the Surkhob River Valley (Tajikistan). The homology of Rb chromosomes had been previously verified by chromosome painting (Bakloushinskaya et al., 2010; Matveevsky et al., 2015). During 10 years of inbred crossings (sister–brother), despite reduced fertility, we got, in total, 9 generations of hybrids (262 litters, 578 animals). For the presented research, we used data on F1–F9 hybrids, 50 animals in total.

Analysis of fertility and the experimental design were carried out using a database accumulating data from breeding and field logs. Animals were treated according to established international protocols, such as the Guidelines for Humane Endpoints for Animals Used in Biomedical Research, and Regulations for Laboratory Practice in Russian Federation, and under the supervision of the Ethics Committee for Animal Research of the Koltzov Institute of Developmental Biology, RAS.

For all the specimens under analysis, we prepared slides, using fixed cells from bone marrow for mitotic metaphases and suspensions of spermatocytes to study meiosis (Ford, Hamerton, 1956; Graphodatsky, Radjabli, 1988). G-banding was carried out for all preparations of metaphase chromosomes, using trypsin treatments, in order to identify Rb metacentrics, according to Seabright (1971). The suspensions and spreads of spermatocytes were made as described by Kolomiets et al. (2010) or Peters et al. (1997). Immunostaining was designed as in our previous studies (Kolomiets et al., 2010; Matveevsky et al., 2016). Synaptonemal complexes (SC) and centromeres in pachytene spermatocytes were detected using antibodies to axial SC elements – SYCP3 (Abcam, UK) – and the kinetochores (ACA, Antibody Incorporated, USA). The slides were analyzed with an Axioimager D1 microscope (Carl Zeiss, Jena, Germany). Images were processed using Adobe Photoshop CS3 Extended.

Results

We had previously demonstrated (Matveevsky et al., 2015) that parental forms possessed two types of $2n = 50$ karyo-

types with two different pairs of Rbs: 2Rb(4.12) and 2Rb(9.13) in 'Khodza Obi-Garm', and 2Rb(2.18) and 2Rb(5.9), in 'Voidara'. F1 hybrids had $2n = 50$, NF = 56, but all the Rb metacentrics were different: Rb(9.13), Rb(2.18), Rb(5.9) and Rb(4.12). In the first hybrid generation, monobrachial homology leads to formation of two trivalents [(2/2.18/18) and (4/4.12/12)], and a tetravalent (5/5.9/9.13/13) in meiotic prophase I (Matveevsky et al., 2015, 2017) (Fig. 1, *a*, Fig. 2, *a*). Hybrids of the first generation had lower fertility, but did not exhibit any health problems, and their longevity was the same as the parental ones' (up to seven years).

Hybrid fertility began to increase in the third generation. We have specially focused on the most intrigued result, which was the emergence in hybrids of new chromosome sets, with different $2n$ (48, 49, 51, 52) and combinations of Rb metacentrics (see Fig. 1, *b–d*). For example, all the progeny of F2, No. 26432, ♂, $2n = 48$, 2Rb(4.12), 2Rb(2.18), 1Rb(5.9), 1Rb(9.13) (see Fig. 1, *c*, *d*) and No. 26433, ♀, $2n = 52$, 2Rb(5.9) got karyotypes identical to F1 hybrids, $2n = 50$, 1Rb(9.13), 1Rb(5.9), 1Rb(2.18), 1Rb(4.12). In meiosis, they demonstrated the same meiotic disturbances, such as the presence of a tetravalent and two trivalents in the pachytene and diakinesis, anaphase bridges, etc. (see Fig. 1, *a*, *h* and Fig. 2, *a*). When we crossed these F3 mole voles, whose karyotype was identical to F1, we appeared to return to the initial crossings. That was the reason for a long-term experiment, because we had to override a high level of heterozygosity, which appeared continuously. Anyway, the substitution of F2 chromosomal sets determined descent generations' variety and appeared to be the first step in the emergency of new balanced chromosomal forms.

G-banding analysis revealed that 4 Rb metacentrics described for these crossings seem to have different evolutionary destinies. Three of them – Rb(9.13), Rb(2.18) and Rb(4.12) – have a tendency to fix in generations, while Rb(5.9) reveals negative dynamics. The smallest Rb(9.13) has an advantage

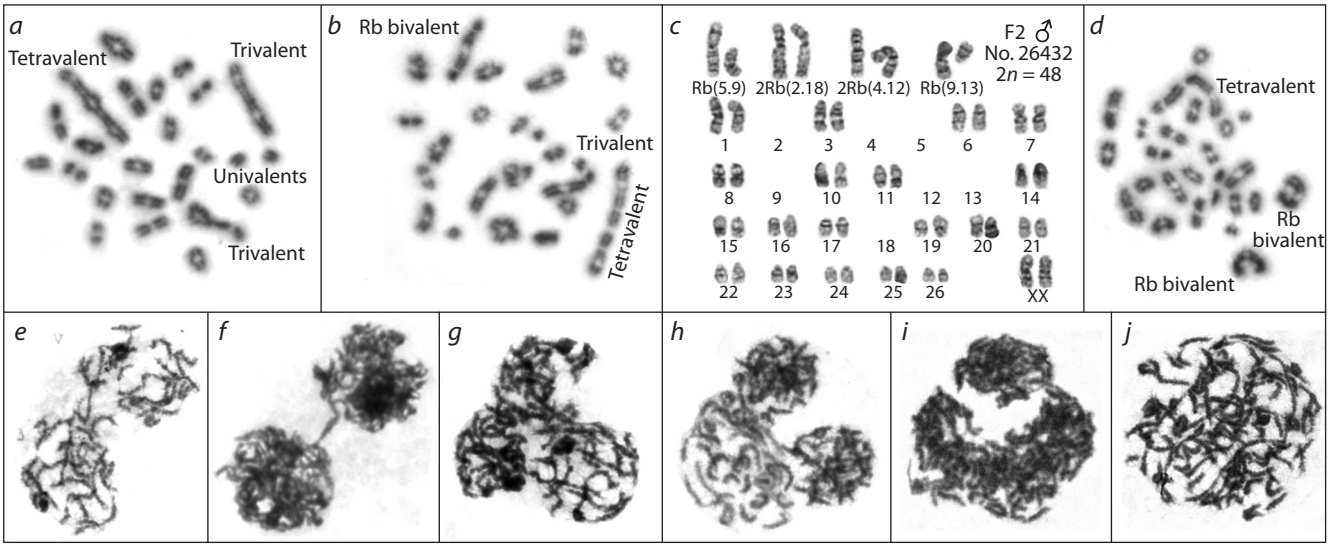


Fig. 1. Meiotic and mitotic chromosomes of *E. tancrei* hybrids (F1–F9).

a, F1, No. 26182, $2n = 50$: 1Rb(5.9), 1Rb(2.18), 1Rb(4.12), 1Rb(9.13), diakinesis; *b*, F2, No. 26453, $2n = 49$: 2Rb(4.12), 1Rb(5.9), 1Rb(2.18), 1Rb(9.13), diakinesis; *c*, F2, No. 26432, $2n = 48$: 2Rb(2.18), 2Rb(4.12), 1Rb(5.9), 1Rb(9.13), karyotype and diakinesis; *d*, abnormalities in meiosis; *e*, F2, No. 26335, $2n = 49$: 2Rb(2.18), 2Rb(9.13), 1Rb(4.12), anaphase I bridge; *f*, F3, No. 26140, $2n = 51$: 2Rb(9.13), 1Rb(2.18), anaphase I bridge; *g*, F2, No. 26139, $2n = 50$: 2Rb(9.13), 1Rb(4.12), 1Rb(2.18); *h*, F3, No. 26436, $2n = 50$: 1Rb(4.12), 1Rb(5.9), 1Rb(9.13), 1Rb(2.18); *i*, F9, No. 26995, $2n = 50$: 2Rb(9.13); *j*, F9, No. 26995, $2n = 50$: 2Rb(9.13), a polyploid cell. Magnification (*a*, *b*, *d–j*): 1000×.

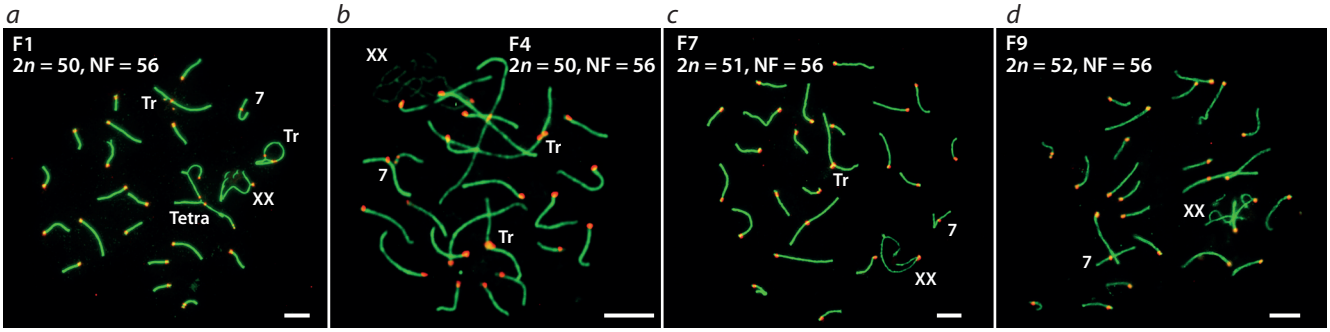


Fig. 2. Chromosome synapsis in pachytene spermatocytes of *E. tancrei* hybrids of different generations.

Axial SC elements were identified using anti-SYCP3 antibodies (green), anti-ACA for kinetochores (red). Tetra: tetravalent; Tr: trivalent; XX: male sex chromosomes. Bivalent No. 7 is an exclusive non-Rb submetacentric in all forms and hybrids of *E. tancrei*. NF: number of chromosome arms.
a, F1, No. 26990, $2n = 50$, NF = 56: 19 bivalents, 1 tetravalent, 2 trivalent, sex (XX) bivalent; *b*, F4, No. 26572. $2n = 50$: 21 bivalents, 2 trivalents, sex (XX) bivalent; *c*, F7, No. 26986, $2n = 51$: 23 bivalents, 1 trivalent, sex (XX) bivalent; *d*, F9, No. 26995. $2n = 52$: 25 bivalents, sex (XX) bivalent. Bar (*a–d*) = 5 μm.

compared to others. In F7–F9, it seems to be in a stable homozygotic state. As the Rb metacentrics started to fix and produce new karyotypes, we observed reduction of the number of complicated figures (trivalents, tetravalents) in meiosis (see Fig. 1, *e–g* and Fig. 2, *b–d*), despite the presence of other deviation (see Fig. 1, *f*, *i*, *h*). In F2 we first noted specimens, homozygotic by one or two pairs of Rb metacentrics. These hybrids also showed minimum abnormalities in the pachytene (see Fig. 2, *d*) and diakinesis, but there were other abnormalities, such as polyploid cells. The ploidy of such cells was rather difficult to count, most cells were probably tetraploid ones (see Fig. 1, *j*). Some animals were somatic mosaic, for example, F7 No. 26986 had $2n = 50–51$ 2Rb(9.13), 1Rb(2.18), 0–1 Rb(4.12) in bone marrow cells. In all spermatocytes studied we distinguished $2n = 51$ only (see Fig. 2, *c*).

Analysis of SCs in F1, F4, F7 and F9 hybrids proved that synapsis and recombination of homologous parts take place, despite the presence of heterozygotes and monobrachial homology of Rb metacentrics (see Fig. 2). Such synapsis is a prerequisite for the proper chromosome disjunction. Most probably, the base for a balanced chromosome segregation is a cis configuration for Rb trivalents, which we observed regularly. Therefore, some cells passed through meiosis successfully and promoted the efficient number of viable gametes.

Discussion

The hypothesis of meiotic drive as a prominent mechanism of quick fixation of chromosomal rearrangements expects preferential inheritance of Rb metacentrics. What is more, some previous Rb inheritance research conducted on shrews (Searle, Wójcik, 1998) suggests that large Rb metacentrics have an advantage compared to smaller ones, but other studies on domestic mice did not prove that (Castiglia, Capanna, 2000). However, those studies did not observe the cases of hybrids with Rb chromosomes having monobrachial (partial or single arm) homology. Our data are mainly consistent with the hypothesis of meiotic drive, but the thesis concerning the dynamics of fixation of large and small Rb metacentrics was not proved. Rb metacentrics of the line under investigation can be split into two groups: small (Rb(9.13) and Rb(4.12)) and large (Rb(5.9) and Rb(2.18)). We observed clear positive dynamics of inheritance and fixation only for Rb(9.13), the smallest of represented Rb metacentrics. Moreover, large Rb(5.9), also containing acrocentric 9, is the only one which revealed negative dynamics. Such a "competition" for the acrocentrics and an advantage of the smaller Rb metacentric made it clear that our results are at variance with some of the previous studies. So, this matter needs follow-up investigations. Preferential inheritance of the smaller chromosomes may be explained by a specific architectonics of the interphase nucleus, which is still unknown for mole voles. Another possible reason is a less complicated picture of meiosis, due to the compactness of chromosomes. For two other Rb metacentrics – Rb(4.12) and Rb(2.18) – the observed segregation was almost Mendelian. But it still confirms the hypothesis of meiotic drive: despite the fact that the acrocentric karyotype is prone to Rb translocations, chromosomal rearrangements are still serious mutations. They change the architectonics and functioning of the nucleus in unexpected ways, so the probability of their fixation in generations is not high. That is why the absence of negative selection may be interpreted as the effect of meiotic drive.

We have observed all possible prophase I SC combinations in *E. tancrei* heterozygous spermatocytes previously (Bogdanov, Kolomiets, 2007; Matveevsky, Kolomiets, 2016). Some complicated figures – trivalents and a tetravalent – were observed in meiosis of mole voles of all studied generations, F1–F9. But we revealed a clear tendency of reduction of their number. All F1 individuals have the same chromosomal chains in prophase I. In F2 homozygotization started, so only a few animals showed the same level of meiotic abnormalities. In F3, some specimens demonstrated normal process of meiosis, not disturbed by monobrachial homology and multivalent formation. In all generations under analysis, abnormalities of late stages of meiosis seem to be the same. They can be summed up as disturbance of synchronism and spatial distribution of chromosomes.

A particularly interesting output of this experiment was a meiotic solution for heterozygotes, carrying Rb metacentric and homologous acrocentrics. The most common meiotic disturbance was delayed synapsis, which resumed later comparing to the homologous crossings. The late synaptic adjustments nevertheless provide a proper segregation of chromosomes and normal sets in the gametes. The question of a lower rate of recombination due to delayed synapses is still open; apparently, it may lead to negative consequences in an evolutionary perspective. It should be noted that the impact of different SC combination on the meiotic progression is significantly variable (Ratomponirina et al., 1988; de la Fuente et al., 2007; Berríos et al., 2017). Some aspects about it were discussed early (Bakloushinskaya et al., 2010; Matveevsky et al., 2015). Another interesting phenomenon was an occurrence of numerous polyploid pachytenes in hybrid meiosis. Possibly, the consequences of chromosome abnormalities, which we observed in the prophase I, might be opposite ones. Longevities and fissions at the zygotene are most probably precursors for new chromosome changes, otherwise abnormalities at the anaphase lead to cell elimination and disappear under natural selection.

It was possible to get a pure line in three generations, if cross animals with identical karyotypes, but we decided to make a 'black box' with a sister–brother breeding system to avoid artificial selection. As a result, we got heterozygous progeny for many generations, but the endpoint appeared to be more similar to natural selection. The main result was an emergence of new balanced chromosomal set and lack of meiotic disturbances.

Conclusion

To conclude, we want to highlight the fact of formation and fixation of new karyotypes in inbred line of *Ellobius tancrei* with monobrachially homologous Rb metacentrics. This model shows the way for origin of new stable chromosomal forms which may occur in wild as well. In several generations, it may lead to divergence and can be treated as early stages of speciation.

References

Baker R.J., Bickham J.W. Speciation by monobrachial centric fusions. Proc. Natl. Acad. Sci. USA. 1986;83:8245-8248. DOI 10.1073/pnas.83.21.8245.

Bakloushinskaya I., Matveevsky S. Unusual ways to lost Y chromosome and survive with changed autosomes: a story of mole voles

Ellobius (Mammalia, Rodentia). OBM Genetics. 2018;2(3). DOI 10.21926/obm.genet.1803023.

Bakloushinskaya I.Yu., Romanenko S.A., Graphodatsky A.S., Matveevsky S.N., Lyapunova E.A., Kolomiets O.L. The role of chromosome rearrangements in the evolution of mole voles of the genus *Ellobius* (Rodentia, Mammalia). Russ. J. Genetics. 2010;46:1143-1145. DOI 10.1134/S1022795410090346.

Bakloushinskaya I., Romanenko S.A., Serdukova N.A., Graphodatsky A.S., Lyapunova E.A. A new form of the mole vole *Ellobius tancrei* Blasius, 1884 (Mammalia, Rodentia) with the lowest chromosome number. Comp. Cytogenet. 2013;7:163-169. DOI 10.3897/CompCytogen.v7i2.5350.

Berríos S., Fernández-Donoso R., Ayarza E. Synaptic configuration of quadrivalents and their association with the XY bivalent in spermatocytes of Robertsonian heterozygotes of *Mus domesticus*. Biol. Res. 2017;50(1):38. DOI 10.1186/s40659-017-0143-6.

Bogdanov Yu.F., Kolomiets O.L. Synaptonemal Complex – Indicator of the Dynamics of Meiosis and Chromosome Variation. Moscow: KMK Scientific Press, 2007. (in Russian)

Castiglia R., Capanna E. Contact zones between chromosomal races of *Mus musculus domesticus*. 2. Fertility and segregation in laboratory-reared and wild mice multiple heterozygous for Robertsonian rearrangements. Heredity. 2000;85:147-157. DOI 10.1046/j.1365-2540.2000.00743.x.

Cremer T., Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. 2001;2(4):292-301. DOI10.1038/35066075.

de la Fuente R., Parra M.T., Viera A., Calvente A., Gómez R., Suja J.Á., Rufas J.S., Page J. Meiotic pairing and segregation of achiasmate sex chromosomes in eutherian mammals: the role of SYCP3 protein. PLoS Genet. 2007;3(11):e198. DOI 10.1371/journal.pgen.0030198.

de Villena F.P., Sapienza C. Female meiosis drives karyotypic evolution in mammals. Genetics. 2001;159:1179-1189.

Dobigny G., Britton-Davidian J., Robinson T.J. Chromosomal polymorphism in mammals: an evolutionary perspective. Biol. Rev. 2017;92(1):1-21. DOI 10.1111/brv.12213.

Faria R., Navarro A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol. Evol. 2010;25:660-669. DOI 10.1016/j.tree.2010.07.008.

Ford C.E., Hamerton J.L. A colchicine, hypotonic citrate, squash sequence for mammalian chromosomes. Stain Techn. 1956;31:247-251. DOI 10.3109/10520295609113814.

Graphodatsky A.S., Radjabli S.I. Chromosomes of Farm and Laboratory Mammals. Atlas. Novosibirsk: Nauka Publ., 1988. (in Russian)

Graphodatsky A.S., Stanyon R., Trifonov V.A. The genome diversity and karyotype evolution of mammals. Mol. Cytogenet. 2011;4:22. DOI 10.1186/1755-8166-4-22.

King M. Species Evolution. The Role of Chromosome Change. Cambridge: Univ. Press, 1993.

Kolomiets O.L., Matveevsky S.N., Bakloushinskaya I.Yu. Sexual dimorphism in prophase I of meiosis in mole vole (*Ellobius talpinus* Pallas) with isomorphic (XX) chromosomes in males and females. Comp. Cytogenet. 2010;4:55-66. DOI 10.3897/compcytogen.v4i1.25.

Lindholm A.K., Dyer K.A., Firman R.C., Fishman L., Forstmeier W., Holman L., Johannesson H., Knief U., Kokko H., Larracuente A.M., Manser A., Montchamp-Moreau C., Petrosyan V.G., Pomiankowski A., Presgraves D.C., Safronova L.D., Sutter A., Unckless R.L., Verspoor R.L., Wedell N., Wilkinson G.S., Price T.A.R. The ecology and evolutionary dynamics of meiotic drive. Trends Ecol. Evol. 2016;31:315-326. DOI 10.1016/j.tree.2016.02.001.

Lyapunova E.A., Bakloushinskaya I.Y., Saidov A.S., Saidov K.K. Dynamics of chromosome variation in mole voles *Ellobius tancrei* (Mammalia, Rodentia) in Pamiro-Alay in the period from 1982 to 2008. Russ J. Genetics. 2010;46:566-571. DOI 10.1134/S1022795410050091.

Lyapunova E.A., Ivnitskii S.B., Korablev V.P., Yanina I.Yu. Complete Robertsonian fan of the chromosomal forms in the mole-vole superspecies *Ellobius talpinus*. Doklady Akademii Nauk SSSR = Proceedings of the Academy of Sciences of the USSR. 1984;274:1209-1213. (in Russian)

Matveevsky S., Bakloushinskaya I., Kolomiets O. Unique sex chromosome systems in *Ellobius*: How do male XX chromosomes recombine and undergo pachytene chromatin inactivation? Sci. Rep. 2016;6,29949. DOI 10.1038/srep29949.

Matveevsky S., Bakloushinskaya I., Tambovtseva V., Romanenko S., Kolomiets O. Analysis of meiotic chromosome structure and behavior in Robertsonian heterozygotes of *Ellobius tancrei*: a case of monobrachial homology. Comp. Cytogenet. 2015;9(4):691-706. DOI 10.3897/CompCytogen.v9i4.5674.

Matveevsky S.N., Kolomiets O.L. Synaptonemal complex configurations in Robertsonian heterozygotes. Tsitologia. 2016;58(4):309-314.

Matveevsky S., Kolomiets O., Bogdanov A., Hakhverdyan M., Bakloushinskaya I. Chromosomal evolution in mole voles *Ellobius* (Cricetidae, Rodentia): Bizarre sex chromosomes, variable autosomes and meiosis. Genes. 2017;8(11):306. DOI 10.3390/genes8110306.

Nunes A.C., Catalan J., Lopez J., da Graça Ramalhinho M., da Luz Mathias M., Britton-Davidian J. Fertility assessment in hybrids between monobrachially homologous Rb races of the house mouse from the island of Madeira: implications for modes of chromosomal evolution. Heredity. 2011;106:348-356. DOI 10.1038/hdy.2010.74.

Peters A.H., Plug A.W., van Vugt M.J., De Boer P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. Chromosome Res. 1997;5(1):66-68.

Potter S., Bragg J.G., Blom M.P., Deakin J.E., Kirkpatrick M., Eldridge M.D., Moritz C. Chromosomal speciation in the genomics era: disentangling phylogenetic evolution of Rock-wallabies. Front. Genet. 2017;8:10. DOI 10.3389/fgene.2017.00010.

Qumsiyeh M.B. Structure and function of the nucleus: anatomy and physiology of chromatin. Cell. Mol. Life Sci. 1999;55:1129-1140.

Ratomponirina C., Brun B., Rumpler Y. Synaptonemal complexes in Robertsonian translocation heterozygous in lemurs. In: Brandham P.E. (Ed). Kew Chromosome Conference 3. London: HMSO, 1988;65-73.

Romanenko S.A., Serdyukova N.A., Perelman P.L., Trifonov V.A., Golenishchev F.N., Bulatova N.S., Stanyon R., Graphodatsky A.S. Multiple intrasyntenic rearrangements and rapid speciation in voles. Sci. Rep. 2018;8:14980. DOI 10.1038/s41598-018-33300-6.

Sandler L., Novitski E. Meiotic drive as an evolutionary force. Am. Nat. 1957;91:105-110.

Seabright M. A rapid banding technique for human chromosomes. Lancet. 1971;2:971-972. DOI 10.1016/S0140-6736(71)90287-X.

Searle J.B., Wójcik J.M. Chromosomal evolution: the case of *Sorex araneus* In: Wójcik J.M., Wolsan M. (Eds.). Evolution of Shrews. Białowieża: Mammal Research Institute PAS, 1998;219-268.

Shapiro J.A. Genome organization and reorganization in evolution: formatting for computation and function. Ann. N. Y. Acad. Sci. 2002;981:111-134. DOI 10.1111/j.1749-6632.2002.tb04915.x.

Vorontsov N.N., Lyapunova E.A., Borissov Y.M., Dovgal V.E. Variability of sex chromosomes in mammals. Genetica. 1980;52/53:361-372.

ORCID ID

V.G. Tambovtseva orcid.org/0000-0002-2607-4169
S.N. Matveevsky orcid.org/0000-0002-5808-7685
I.Yu. Bakloushinskaya orcid.org/0000-0002-5808-7685

Acknowledgements. This study was funded in part by the IDB RAS Government basic research program No. 0108-2019-0007, VIGG RAS State Assignment Contract No. 0112-2019-0002, Program of the Russian Academy of Sciences' Presidium No. 41 'Biodiversity of Natural Systems and Biological Resources of Russia', and the Russian Foundation for Basic Research No. 16-04-01447.

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised December 17, 2018. Accepted December 17, 2018.

Вавиловский журнал генетики и селекции. 2019;23(2):244-249
DOI 10.18699/VJ19.489

Runs of homozygosity in spontaneous abortions from families with recurrent pregnancy loss

N.A. Skryabin ✉, S.A. Vasilyev, T.V. Nikitina, D.I. Zhigalina, R.R. Savchenko, N.P. Babushkina, M.E. Lopatkina, A.A. Kashevarova, I.N. Lebedev

Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia
✉ e-mail: nikolay.skryabin@medgenetics.ru

Recurrent pregnancy loss (RPL) is a severe reproductive pathology with a significant component of unexplained etiology. Extended homozygous regions as a possible etiological factor for RPL were sought in the genomes of embryos. Twenty-two paired first-trimester spontaneously aborted embryos from eleven women with recurrent miscarriage were analyzed. All embryos had normal karyotypes according to metaphase karyotyping and conventional comparative genomic hybridization. SurePrint G3 Human CGH + SNP 4 × 180K microarrays (Agilent Technologies) were used to search for homozygous regions. As a result, 39 runs of homozygosity (ROH) were identified in extraembryonic tissues of 15 abortuses. Verification of recurrent homozygous regions was performed by Sanger sequencing. The presence of occasional heterozygous SNPs was shown in 25 extended ROHs, which may indicate that they did not arise *de novo* but were inherited from parents. In the course of inheritance in a series of generations, they may accumulate mutations, leading to heterozygosity for several sites in the initially homozygous population-specific regions. Homozygotization of recessive mutations is one of the putative mechanisms of the influence of such inherited ROHs on RPL development. The high frequency of extended ROHs detected in the present study may point to a role of inbreeding in RPL etiology. Homozygous regions may also occur due to uniparental disomy, and abnormalities of genomic imprinting may be another mechanism responsible for the pathological manifestation of ROHs in embryogenesis. Indeed, five predicted imprinted genes were identified within ROHs according to the Geneimprint database: *OBSCN*, *HIST3H2BB*, *LMX1B*, *CELF4*, and *FAM59A*. This work reports the first finding of a high frequency of extended ROHs in spontaneously aborted embryos with normal karyotypes from families with RPL.
Key words: recurrent pregnancy loss; array-based comparative genomic hybridization; runs of homozygosity; spontaneous abortions; genomic imprinting.

For citation: Skryabin N.A., Vasilyev S.A., Nikitina T.V., Zhigalina D.I., Savchenko R.R., Babushkina N.P., Lopatkina M.E., Kashevarova A.A., Lebedev I.N. Runs of homozygosity in spontaneous abortions from families with recurrent pregnancy loss. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(2):244-249. DOI 10.18699/VJ19.489

Районы гомозиготности в тканях абортусов из семей с привычным невынашиванием беременности

Н.А. Скрябин ✉, С.А. Васильев, Т.В. Никитина, Д.И. Жигалина, Р.Р. Савченко, Н.П. Бабушкина, М.Е. Лопаткина, А.А. Кашеварова, И.Н. Лебедев

Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия
✉ e-mail: nikolay.skryabin@medgenetics.ru

Привычное невынашивание беременности является тяжелой репродуктивной патологией, при этом почти в половине случаев его этиология остается невыясненной. В настоящей работе проведено исследование протяженных гомозиготных районов (runs of homozygosity – ROH) в геноме как возможного этиологического фактора в развитии привычного невынашивания беременности. Всего проанализировано 22 парных абортуса первого триместра беременности от 11 женщин с привычным невынашиванием беременности. У всех абортусов был нормальный кариотип по результатам метафазного кариотипирования и сравнительной геномной гибридизации на метафазных пластинках. Для поиска гомозиготных регионов были использованы микрочипы SurePrint G3 Human CGH + SNP 4 × 180K (Agilent Technologies). В результате в экстраэмбриональных тканях 15 абортусов выявлены 39 гомозиготных областей. Для образцов с повторяющимися ROH была проведена верификация с помощью секвенирования по Сэнгеру. Для 25 протяженных ROH было показано наличие единичных гетерозиготных SNP, что может свидетельствовать в пользу того, что они образуются не *de novo*, а наследуются от родителей. При этом в ходе наследования в ряду поколений в них могут накапливаться мутации, приводящие к гетерозиготизации изначально гомозиготных участков генома. Один из возможных механизмов влияния таких унаследованных ROH на патогенез привычного невынашивания беременности – гомозиготизация рецессивных мутаций. Высокая частота длинных ROH, обнаруженная в настоящем исследовании, указывает на роль инбридинга в этиологии привычного невынашивания беременности. Гомозиготные области

могут возникать также из-за однородительской дисомии, следовательно, аномалии геномного импринтинга могут быть другим механизмом, ответственным за патологическое проявление гомозиготных районов в эмбриогенезе. В составе обнаруженных регионов гомозиготности в соответствии с базой данных Geneimprint было выявлено пять предположительно импринтированных генов: *OBSCN*, *HIST3H2BB*, *LMX1B*, *CELF4* и *FAM59A*. В результате проведенного исследования впервые показана высокая частота длинных ROH в тканях спонтанных абортусов с нормальным кариотипом из семей с привычным невынашиванием беременности.
Ключевые слова: привычное невынашивание беременности; сравнительная геномная гибридизация на микрочипах; районы гомозиготности; ROH; спонтанные аборты; геномный импринтинг.

Introduction

Spontaneous abortions occur in approximately 15 % of pregnancies in different populations; however, in some families, pregnancy loss occurs more than once (Nikitina et al., 2016). According to the guidelines of the European Society of Human Reproduction and Embryology, a diagnosis of recurrent pregnancy loss (RPL) should be considered after the loss of two or more pregnancies (Goddijn, 2017). The prevalence of RPL varies from 2 to 5 % of the total number of pregnancies (El Hachem et al., 2017). Recurrent pregnancy loss may be caused by uterine abnormalities, parental chromosomal aberrations, antiphospholipid antibodies, polycystic ovarian syndrome, diabetes mellitus, and hyperthyroidism. In about half of all cases, the RPL cause remains unexplained (Nikitina et al., 2016).

Homozygotization of recessive mutations due to runs of homozygosity (ROH) or uniparental disomy (UPD) may be one of the causes (Robberecht et al., 2012; Niida et al., 2018). In a consanguine marriage with manifestations of recessive diseases, the homozygous risk locus is likely to fall into an extended homozygous region. Thus, pathogenic recessive variants can be identified in affected individuals from inbred population by identifying ROHs. This approach to the search for pathological mutations made it possible to map genes associated with many recessive Mendelian diseases (Botstein, Risch, 2003).

The emergence of regions with ROHs may result from inbreeding, but ROHs are found in human genomes even among outbred populations (Ceballos et al., 2018). Population history and cultural factors may influence homozygosity levels, and therefore ROHs are characterized by population specificity, while data on their prevalence in Russian populations are unavailable yet.

The aim of this paper was to study the role of homozygotization of various genome regions in the pathogenesis of RPL.

Materials and methods

Patients supervised by the Laboratory of Cytogenetics of the Research Institute of Medical Genetics (Tomsk, Russia) from 1987 to 2018 were analyzed. Material from first-trimester spontaneously aborted embryos was obtained from gynecologic and obstetric clinics in Tomsk and Seversk (Russia). Information was recorded about the maternal and paternal age, gynecological anamnesis of the women, the number and outcomes of previous pregnancies, and features of the present gestation.

Tissue samples obtained by curettage were transferred to the cytogenetic laboratory in sterile saline. The products of conception were examined, and extraembryonic tissues were separated from decidua and blood clots. All abortuses had normal karyotypes according to the results of metaphase karyotyping and conventional comparative genomic hybridization of extraembryonic tissues. For conventional karyotyping, metaphase chromosomes were obtained after long-term culture of fibroblasts of extraembryonic mesoderm in DMEM/F12 (1:1) medium (Sigma, United States) supplemented with 20 % fetal bovine serum (HyClone, United States). Colchicine (Sigma, United States) was added 4 h before chromosome harvesting, and the samples were processed by standard techniques. All specimens were G-banded with trypsin-Giemsa (Sigma, United States) to identify chromosomes.

For conventional comparative genomic hybridization (cCGH), test and reference DNA were labeled by nick translation (Rooney, Czepulkowski, 1992). Hybridization of the DNA libraries on metaphase plates derived from peripheral blood lymphocytes of a healthy male was carried out with suppression with 50-fold excess C_0t-1 DNA (Agilent Technologies, United States) for 72 h at 37 °C in a Thermo-Brite hybridization chamber (Abbott Molecular, United States). Metaphase chromosomes were stained with DAPI. Hybridization signals were detected using Axio Imager Z2 fluorescence microscope (Carl Zeiss, Germany) with a set of corresponding filters. The cCGH results were processed with Isis CGH software (Metasystems, Germany).

A search for homozygous regions was performed with SurePrint G3 Human CGH + SNP 4 × 180K microarrays (Agilent Technologies) in the extraembryonic tissues from 22 paired spontaneous abortions with the normal karyotype from 11 women with RPL. Results were visualized with Cytogenomics software (Agilent Technologies). Default Analysis Method (CGH+SNP v2) was used with a LOH threshold of 6.0. The resolution of this platform for LOH detection is about 2 Mb.

Some recurrent homozygous regions were verified by Sanger sequencing. To confirm the status of heterozygosity, 18 single nucleotide polymorphism (SNP) markers located in the analyzed chromosomal regions were chosen. Criteria for the choice of markers: the level of heterozygosity in Caucasians close to 0.5 and uniform distribution within assumed homozygosity regions. DNA sequences flanking the analyzed SNPs (155 to 350 bp for different markers) were analyzed for each marker. In the presence of other polymorphic variants in the analyzed fragments, their heterozygosity status was also taken into account. Sequencing was performed on an Applied Biosystems 3730 automated analyzer with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, United States) according to the manufacturer's protocol. Data analysis was performed with Sequencing Analysis v5.4 software.

The analysis of the gene composition was carried out with the "WEB-based GEne SeT AnaLysis Toolkit" program based on overrepresentation enrichment analysis (ORA) (http://www.webgestalt.org). Gene categories were classified ac-

Table 1. List of ROHs

Family	Abortions	Coordinates, GRCh37/hg19	Chromosomal regions	Size, Kb	Total length, Kb	# SNP Probes	# Genes	Heterozygous SNP
1	1a	–	–	–	–	–	–	–
	1b	chr10:59128631-61154870	10q21.1	2026	7993	44	8	0
		chr18:29703102-35670326	18q12.1-q12.2	5967		95	28	7
2	2a	chr5:31798252-34639248	5p13.3-p13.2	2841	16003	50	15	0
		chr7:123895098-127020636	7q31.33	3126		79	6	1
		chr8:50166541-52720312	8q11.21-q11.23	2554		73	2	1
		chr12:39044654-41735719	12q12	2691		56	8	0
		chr14:41387006-46177851	14q21.1-q21.2	4791		73	10	1
	2b	chr7:114786932-120722820	7q31.2-q31.31	5936	11849	75	21	2
		chr17:57263-393-63176847	17q22-q24.1	5913		53	79	0
3	3a	chr1:222474942-228854703	1q41-q42.13	6380	14534	63	65	0
		chr7:85117732-90826127	7q21.11-q21.13	5708		104	23	3
		chr10:57626235-60072490	10q21.1	2446		59	4	0
	3b	chr1:69035118-74558526	1p31.2-p31.1	5523	17172	83	12	3
		chr1:222474942-228854703	1q41-q42.13	6380		65	65	1
		chr7:86143354-90826127	7q21.11-q21.13	4683		88	23	3
		chr10:19811839-25672136	10p12.31-p12.1	5860		94	29	3
4	4a	–	–	–	–	–	–	–
	4b	chr2:153190223-160121961	2q23.3-q24.2	6932	60758	95	19	2
		chr3:175630053-181557730	3q26.31-q26.33	5928		103	20	1
		chr4:38169660-47837883	4p14-p12	9668		146	48	1
		chr4:68962758-77834868	4q13.2-q21.1	8872		108	90	2
		chr9:123453281-132044983	9q33.2-q34.11	8592		81	140	3
		chr11:37140377-39818049	11p12	2678		61	0	0
		chr11:11982521-17593273	11p15.3-p15.1	5611		88	29	1
		chr14:88986214-92648712	14q31.3-q32.12	3662		62	29	1
		chr22:37357169-46172372	22q12.3-q13.31	8815		81	164	0
5	5a	chr11:37902019-40336165	11p12	2434	7327	53	1	0
		chr18:29334151-34227604	18q12.1-q12.2	4893		90	27	4
	5b	–	–	–	–	–	–	–
6	6a	chr3:67122886-70783491	3p14.1-p13	3661	6862	58	12	4
		chr18:32132169-35332795	18q12.1-q12.2	3201		49	20	6
	6b	chr6:80423690-82979725	6q14.1	2556	2556	54	6	0
7	7a	–	–	–	–	–	–	–
	7b	chr14:38541336-41364951	14q21.1	2824	2824	52	9	0
8	8a	–	–	–	–	–	–	–
	8b	chr2:180975918-184732178	2q31.3-q32.1	3756	3756	70	12	2
9	9a	chr6:144072401-147738560	6q24.2-24.3	3666	3666	66	17	1
	9b	chr6:83789798-86115244	6q14.1-q14.3	2325	2325	46	11	2
10	10a	chr2:199760671-205164757	2q33.1-q33.3	5404	13103	87	48	4
		chr11:37115764-39546879	11p12	2431		46	0	0
		chr21:28139416-33407446	21q21.3-q22.11	5268		89	56	4
	10b	chr2:82226742-85375676	2p12-p11.2	3149	5580	49	9	0
		chr11:37115764-39546879	11p12	2431		48	0	0
11	11a	–	–	–	–	–	–	–
	11b	–	–	–	–	–	–	–

cording to "Gene Ontology (GO)" database (http://www.geneontology.org).

Results and discussion

Thirty-nine homozygous regions were identified in extraembryonic tissues of 15 embryos (Table 1). The average size of the identified regions was 4.6 Mb (from 2.0 to 9.6 Mb). Due to the limitation of the resolution of the used platform, adequate comparison of the results with population-specific data available in the literature is difficult. However, comparisons were made with studies in which identified ROHs were ranked by size (McQuillan et al., 2008; Pemberton et al., 2012). It was shown that extended ROHs (>2.5 Mb) are not typical for unrelated individuals and for populations with low levels of inbreeding. (About 20 % of samples had extended ROHs.) In contrast, isolated endogamous populations had a high frequency of extended ROHs (~80 %) (McQuillan et al., 2008). The frequency of extended ROHs in our samples was 68.1 % (15/22). In addition, multiple extended ROHs were detected in 40 % of the samples (9/22). It is possible that the high frequency of extended ROHs is associated with a possible role of inbreeding in the etiology of RPL.

In two embryos from one couple (family # 3), two identical homozygous regions (1q41-q42.13 and 7q21.11-q21.13) and several individual ROHs (3a – 10q21.1, 3b – 1p31.2-p31.1 and 10p12.31-p12.1) were identified. In family # 10, a common ROH (11p12) and additional individual regions (10a – 2q33.1-q33.3 and 21q21.3-q22.11, 10b – 2p12-p11.2) were found. This finding is likely to point to inbreeding, but its verification is impossible at present, as DNA samples from the parents are not available. Several ROHs were detected in more than one embryo but in different families: two ROHs were found in two embryos (7q21.11-q21.13, 10q21.1) and one in three embryos (18q12.1-q12.2).

When comparing the repeated homozygous regions found in this study with the most frequent ROHs in other populations, no matches were found (Nothnagel et al., 2010; Pemberton et al., 2012). This may be indicative of a specific prevalence pattern of ROHs in the Siberian population or result from a strong sample bias in comparison with general populations. For more detailed analysis, characterization of Russian populations by ROH structure, location and frequency of occurrence is required.

Samples with recurrent ROHs were selected for verification by Sanger sequencing (Table 2). As a result, occasional heterozygous SNPs were identified in some regions. Detailed analysis of the results of the microarray study also revealed occasional heterozygous SNPs within the homozygous regions. Out of 39 regions of homozygosity, occasional heterozygous SNPs were present in 25 cases (64 %), whereas only homozygous SNPs were identified in 14 spontaneous abortions (36 %). The presence of heterozygous sites in extended homozygous regions suggests that these chromosome regions are not the result of *de novo* loss of heterozygosity. Such homozygous regions may occur when identical haplotypes are inherited from each parent (Peripolli et al., 2017). Mutations may accumulate in them in a series of generations, leading to heterozygosity in some sites of initially homozygous regions.

The information on clinical effects of ROHs or their associations with phenotypic traits is scarce. One of the possible

Table 2. Genotypes for highly polymorphic SNPs in recurrent ROHs analyzed by Sanger sequencing

Family	Embryo	Coordinates, GRCh37/hg19	Chromosomal regions	Size, Kb	SNP	Genotype
3	3a	chr14:41387006-46177851	14q21.1-q21.2	4791	rs10144740	A
					rs7154765	T
					rs961720	T
4	4a	chr1:222474942-228854703	1q41-q42.13	638	rs2003114	A
	4a	chr7:85117732-90826127	7q21.11-q21.13	5708	rs2519713	T
	4b	chr1:222474942-228854703	1q41-q42.13	638	rs3856155	T
					rs849904	G
	4b	chr7:86143354-90826127	7q21.11-q21.13	4683	rs17301014	A
					rs10281928	AG
					rs4728956	A
5	5b	chr11:37140377-39818049	11p12	2678	rs7942537	C
					rs10768276	A
6	6a	chr11:37902019-40336165	11p12	2434	rs10837260	C
					rs11035463	AG
8	8b	chr14:38541336-41364951	14q21.1	2824	rs2250334	GA
					rs956109	G
					rs9989165	C
					rs7154829	G

mechanisms of the influence of ROH on the pathogenesis of RPL may be homozygotization of recessive mutations significant for early embryo development. Some studies demonstrate that ROHs are significantly associated with many common diseases, such as Alzheimer's disease, rheumatoid arthritis, autism, and coronary artery disease (Yang et al., 2012; Gamsiz et al., 2013; Christofidou et al., 2015; Ghani et al., 2015). Most of the ROHs in our study contained genes; the average number of genes per ROH was about 30. The region 22q12.3-q13.31 of 8 Mb in size contained the largest number of genes, 164 (sample 4b). There were no genes in three spontaneous abortions with corresponding homozygous regions (samples 4b, 10a, and 10b), no known genes were identified. All regions in these samples were mapped to 11p12.

To study the overrepresentation of groups of genes by biological functions among all genes localized in ROHs, enrichment analysis was carried out. As a result, a group of genes involved in the metabolism of flavonoids was identified (GO: 0009812, FDR = 0.0001). This functional group includes mainly genes from the UDP-glucuronosyltransferase family (*UGT2B11*, *UGT2A1*, *UGT2B28*, *UGT2B4*, *UGT2B7*, *UGT2B10*, *UGT2B15*, *UGT2B17*, *UGT2A3*, *SULT1B1*). However, a more detailed analysis showed that all genes from this group mapped to the same region 4q13.2-q21.1 in only one sample, 4b. Consequently, the overrepresentation of this group of genes is due to their nonrandom occurrence within one ROH region but not systemic overrepresentation in different samples. The absence of common pathogenetic pathways for different samples may be indicative of the role of homozygosity for different genes, which is quite possible in the light of the extended size of ROHs and the number of affected genes.

As mentioned above, in addition to regions with occasional heterozygous SNPs, regions containing only homozygous SNPs were identified. Perhaps such sites arise from *de novo* mitotic recombination, that is, UPD occurs. Papenhausen et al. (2011) showed that about 31 % of the samples (29/92) with homozygous regions had appeared from UPD. In our study, 36 % of regions (14/39) without heterozygous SNPs were identified, and they might have been generated by *de novo* mitotic recombination. Imprinting abnormalities can be one of the mechanisms by which such UPD regions contribute to the etiology of miscarriage. When analyzing the gene composition according to the Geneimprint database, five predicted imprinted genes were found: *OBSCN*, *HIST3H2BB*, *LMX1B*, *CELF4*, and *FAM59A*.

Conclusion

This is the first study to demonstrate a high frequency of extended ROHs in spontaneous abortions with a normal karyotype from families with RPL. Future studies should be dedicated to further analysis of the prevalence of ROHs in Russian populations and consideration of this phenomenon from the viewpoint of association with common diseases.

References

Botstein D., Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 2003;33:228-237. DOI 10.1038/ng1090.

Ceballos F.C., Hazelhurst S., Ramsay M. Assessing runs of homozygosity: a comparison of SNP array and whole genome sequence low coverage data. BMC Genomics. 2018;19(1):106. DOI 10.1186/s12864-018-4489-0.

Christofidou P., Nelson C.P., Nikpay M., Qu L., Li M., Loley C., Debiec R., Braund P.S., Denniff M., Charchar F.J., Arjo A.R., Trégouët D.A., Goodall A.H., Cambien F., Ouwehand W.H., Roberts R., Schunkert H., Hengstenberg C., Reilly M.P., Erdmann J., McPherson R., König I.R., Thompson J.R., Samani N.J., Tomaszewski M. Runs of homozygosity: association with coronary artery disease and gene expression in monocytes and macrophages. Am. J. Hum. Genet. 2015;97(2):228-237. DOI 10.1016/j.ajhg.2015.06.001.

El Hachem H., Crepaux V., May-Panloup P. Recurrent pregnancy loss: current perspectives. Int. J. Womens Health. 2017;9:331-345. DOI 10.2147/IJWH.S100817.

Gamsiz E.D., Viscidi E.W., Frederick A.M., Nagpal S., Sanders S.J., Murtha M.T., Schmidt M., Triche E.W., Geschwind D.H., State M.W., Istrail S., Cook E.H., Jr., Devlin B., Morrow E.M. Intellectual disability is associated with increased runs of homozygosity in simplex autism. Am. J. Hum. Genet. 2013;93(1):103-109. DOI 10.1016/j.ajhg.2013.06.004.

Ghani M., Reitz C., Cheng R., Vardarajan B.N., Jun G., Sato C., Naj A., Rajbhandary R., Wang L.S., Valladares O., Lin C.F., Larson E.B., Graff-Radford N.R., Evans D., De Jager P.L., Crane P.K., Buxbaum J.D., Murrell J.R., Raj T., Ertekin-Taner N., Logue M., Baldwin C.T., Green R.C., Barnes L.L., Cantwell L.B., Fallin M.D., Go R.C., Griffith P.A., Obisesan T.O., Manly J.J., Lunetta K.L., Kamboh M.I., Lopez O.L., Bennett D.A., Hendrie H., Hall K.S., Goate A.M., Byrd G.S., Kukull W.A., Foroud T.M., Haines J.L., Farrer L.A., Pericak-Vance M.A., Lee J.H., Schellenberg G.D., George-Hyslop P.St., Mayeux R., Rogaeva E. Association of long runs of homozygosity with Alzheimer disease among African American individuals. JAMA Neurol. 2015;72(11):1313-1323. DOI 10.1001/jamaneurol.2015.1700.

Goddijn M. ESHRE. Recurrent Pregnancy Loss. Guideline of the European Society of Human Reproduction and Embryology, 2017.

McQuillan R., Leutenegger A.L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A.K., Farrington S.M., Rudan P., Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Campbell H., Wilson J.F. Runs of homozygosity in European populations. Am. J. Hum. Genet. 2008;83(3):359-372. DOI 10.1016/j.ajhg.2008.08.007.

Niida Y., Ozaki M., Shimizu M., Ueno K., Tanaka T. Classification of uniparental isodisomy patterns that cause autosomal recessive disorders: proposed mechanisms of different proportions and parental origin in each pattern. Cytogenet. Genome Res. 2018;154(3):137-146. DOI 10.1159/000488572.

Nikitina T.V., Sazhenova E.A., Tolmacheva E.N., Sukhanova N.N., Kashevarova A.A., Skryabin N.A., Vasilyev S.A., Nemtseva T.N., Yuriev S.Y., Lebedev I.N. Comparative cytogenetic analysis of spontaneous abortions in recurrent and sporadic pregnancy losses. Biomed. Hub. 2016;1:446099. DOI 10.1159/000446099.

Nothnagel M., Lu T.T., Kayser M., Krawczak M. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum. Mol. Genet. 2010;19(15):2927-2935. DOI 10.1093/hmg/ddq198.

Papenhausen P., Schwartz S., Risheg H., Keitges E., Gadi I., Burnside R.D., Jaswaney V., Pappas J., Pasion R., Friedman K., Tepperberg J. UPD detection using homozygosity profiling with a SNP genotyping microarray. Am. J. Med. Genet. Pt. A. 2011;155(4):757-768. DOI 10.1002/ajmg.a.33939.

Pemberton T.J., Absher D., Feldman M.W., Myers R.M., Rosenberg N.A., Li J.Z. Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. 2012;91(2):275-292. DOI 10.1016/j.ajhg.2012.06.014.

Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R., Baldi F. Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 2017;48(3):255-271. DOI 10.1111/age.12526.

Robberecht C., Pexsters A., Deprest J., Fryns J.P., D'Hooghe T., Vermeesch J.R. Cytogenetic and morphological analysis of early products of conception following hystero-embryoscopy from couples with recurrent pregnancy loss. Prenat. Diagn. 2012;32(10):933-942. DOI 10.1002/pd.3936.

Rooney D.E., Czepulkowski B.H. Human Cytogenetics. A Practical Approach. New York: Oxford Univ. Press, 1992.

Yang H.C., Chang L.C., Liang Y.J., Lin C.H., Wang P.L. A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. PLoS One. 2012;7(4):e34840. DOI 10.1371/journal.pone.0034840.

ORCID ID

N.A. Skryabin orcid.org/0000-0002-2491-3141
S.A. Vasilyev orcid.org/0000-0002-5301-070X
T.V. Nikitina orcid.org/0000-0002-4230-6855
D.I. Zhigalina orcid.org/0000-0003-2800-8875
R.R. Savchenko orcid.org/0000-0003-3839-3543
N.P. Babushkina orcid.org/0000-0001-6133-8986
M.E. Lopatkina orcid.org/0000-0002-7051-3472
A.A. Kashevarova orcid.org/0000-0002-0716-4302
I.N. Lebedev orcid.org/0000-0002-0482-8046

Acknowledgements. This study was supported by the Russian Science Foundation, project 16-15-10229.

Conflict of interest. The authors declare no conflict of interest.

Received December 4, 2018. Revised February 8, 2019. Accepted February 11, 2019.

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, ВИНИТИ, базы данных Emerging Sources Citation Index (Web of Science), Zoological Record (Web of Science), Scopus, 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».

✉ e-mail: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: 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 22.03.2019. Выход в свет 29.03.2019. Формат 60 × 84 $^1/_8$. Усл. печ. л. 15.11.
Уч.-изд. л. 18.2. Тираж 150 экз. Заказ № 44. Цена свободная.

Отпечатано в типографии ФГУП «Издательство СО РАН», Морской проспект, 2, Новосибирск, 630090.