

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

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

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

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

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

Detection of new regulatory SNPs associated with colorectal cancer predisposition

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

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

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

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

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

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

HOW TO CITE THIS ARTICLE:

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

ORIGINAL ARTICLE

Received 11.03.2016

Accepted for publication 26.04.2016

© AUTHORS, 2016

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

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

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

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

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

Материалы и методы

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

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

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

The data used in the work

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

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

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

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

Обработка данных была проведена в среде программирования R.

Результаты

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

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

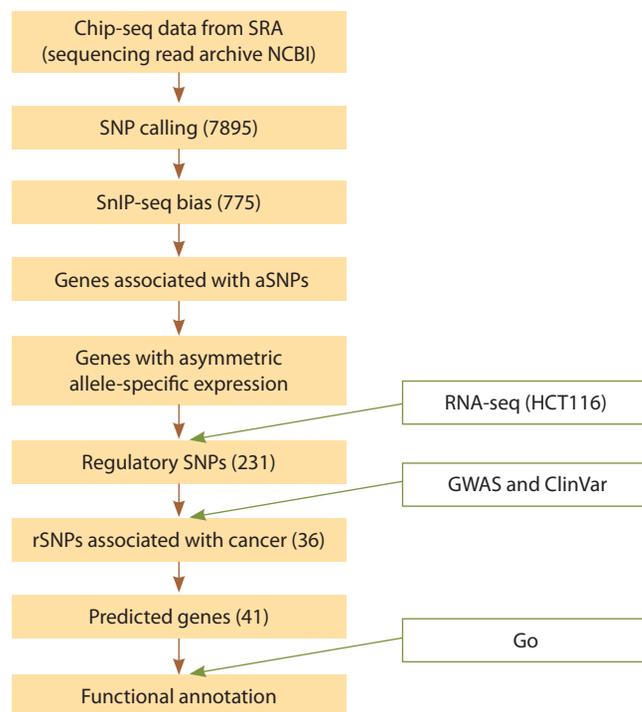


Fig. 1. Genome-wide search for regulatory SNPs.

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

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

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

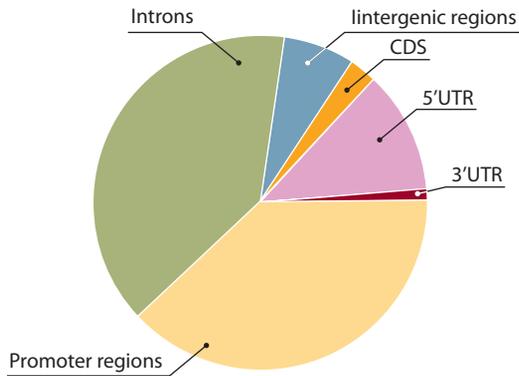


Fig. 2. Distribution of detected aSNPs over genome regions.

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

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

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

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

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

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

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

Обсуждение

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

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

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

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

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

емой в экспериментах ChIP-seq аллель-специфичности распределения гистоновых меток и негистоновых белков хроматина в местах локализации гетерозиготных SNP, которая рассматривается в качестве важного указания на регуляторную функцию таких SNP (Kasowski et al., 2010; Reddy et al., 2012; Trynka et al., 2015). Это позволило выявить 775 SNP, которые потенциально могут влиять на уровень экспрессии генов в клетках HCT-116. Для проверки этого предположения была проведена оценка аллель-специфичности экспрессии генов с использованием данных RNA-seq, полученных на той же клеточной линии. В результате была подтверждена функциональность 231 SNP (rSNP). Для выбора полиморфизмов, относящихся к развитию колоректального рака, были отобраны rSNP, находящиеся в одной группе сцепления (± 10 т.п.н.) с SNP, ассоциированными с этим заболеванием, по данным GWAS и ClinVar. С помощью этого подхода было выявлено 36 rSNP в регуляторных районах 41 гена. Функциональная аннотация этих генов подтвердила полученные ранее данные о роли генов *BALAP2L1* и *BUB3* в формировании предрасположенности к раку толстого кишечника (Cohen, 2003). Причем если ранее для гена *BUB3* были найдены связанные с колоректальным раком мутации в его кодирующей части (Cohen, 2003), с помощью предложенного подхода такие мутации были найдены в районе, соответствующем 5'-UTR. Также были получены новые данные, подтверждающие вовлеченность в развитие колоректального рака RAS/MAPK и WNT-сигнальных путей. В нашем исследовании обнаружены новые гены-кандидаты развития колоректального рака, *RRAGD* и *FZD6*, белковые продукты которых являются компонентами данных сигнальных путей. Кроме того, выявлено 14 новых потенциальных генов-кандидатов колоректального рака, перспективных для дальнейшего изучения.

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

Acknowledgments

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

Conflict of interest

The authors declare no conflict of interest.

References

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

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

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