

Английский текст <https://vavilov.elpub.ru/jour>

Геномная изменчивость в регуляторных районах генов, ассоциированная с заболеваниями человека: механизмы влияния на транскрипцию генов и полногеномные информационные ресурсы, обеспечивающие исследование этих механизмов

Е.В. Игнатьева^{1,2}✉, Е.А. Матросова^{1,2}

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ eignat@bionet.nsc.ru

Аннотация. Полногеномные и полноэкзомные технологии секвенирования играют важную роль в исследовании генетических аспектов патогенеза различных заболеваний. Широкое применение методов полногеномного и полноэкзомного анализа ассоциаций позволяет идентифицировать множество вариантов геномной изменчивости (ГИ), ассоциированных с заболеваниями. Эта информация накапливается в базах данных GWAS central, GWAS catalog, OMIM, ClinVar и др. Большинство вариантов, идентифицированных методикой полногеномного анализа ассоциаций, располагается в некодирующих областях генома человека. По данным проекта ENCODE, доля участков в геноме человека, потенциально задействованных в регуляции транскрипции, во много раз превышает долю кодирующих областей. Таким образом, геномная изменчивость в некодирующих областях генома может повышать предрасположенность к заболеваниям, нарушая функционирование различных регуляторных элементов (промоторов, энхансеров, участков, определяющих 3D структуру хроматина и т. д.). Однако идентификация механизмов влияния патогенных вариантов ГИ на риск развития заболеваний затруднена ввиду большого разнообразия регуляторных элементов. В обзоре рассмотрены молекулярно-генетические механизмы влияния патогенных вариантов ГИ на экспрессию генов. При этом внимание сосредоточено на транскрипционном уровне регуляции как ключевой стадии, запускающей последовательность этапов экспрессии любого гена. Пусковым событием, опосредующим влияние патогенного варианта ГИ на уровень экспрессии гена, может быть, например, изменение функциональной активности сайтов связывания транскрипционных факторов или уровня метилирования ДНК, что, в свою очередь, отражается на функциональной активности промоторов или энхансеров. Выявление регуляторных эффектов полиморфных локусов невозможно без тесной интеграции современных экспериментальных подходов с компьютерным анализом больших массивов генетических данных, получаемых на основе омиксных технологий. В обзоре кратко описаны наиболее известные открытые полногеномные информационные ресурсы, содержащие данные, полученные на основе омиксных технологий, в том числе: ресурсы, накапливающие сведения о состоянии хроматина и участках его связывания с транскрипционными факторами, выявленными с помощью технологии ChIP-seq; ресурсы по геномным локусам, для которых на основе данных ChIP-seq выявлено аллель-специфичное связывание с транскрипционными факторами; а также ресурсы, содержащие предсказанные *in silico* данные о потенциальном влиянии геномной изменчивости на сайты связывания транскрипционных факторов.

Ключевые слова: регуляция транскрипции; геномная изменчивость; патогенные геномные варианты; районы, регулирующие транскрипцию; сайты связывания транскрипционных факторов; геномные базы данных.

Для цитирования: Игнатьева Е.В., Матросова Е.А. Геномная изменчивость в регуляторных районах генов, ассоциированная с заболеваниями человека: механизмы влияния на транскрипцию генов и полногеномные информационные ресурсы, обеспечивающие исследование этих механизмов. *Вавиловский журнал генетики и селекции*. 2021;25(1): 18-29. DOI 10.18699/VJ21.003

Disease-associated genetic variants in the regulatory regions of human genes: mechanisms of action on transcription and genomic resources for dissecting these mechanisms

Е.В. Ignatieva^{1,2}✉, Е.А. Matrosova^{1,2}

¹ Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

✉ eignat@bionet.nsc.ru

Abstract. Whole genome and whole exome sequencing technologies play a very important role in the studies of the genetic aspects of the pathogenesis of various diseases. The ample use of genome-wide and exome-wide association study methodology (GWAS and EWAS) made it possible to identify a large number of genetic variants associated with diseases. This information is accumulated in the databases like GWAS central, GWAS catalog, OMIM, ClinVar, etc. Most of the vari-

ants identified by the GWAS technique are located in the noncoding regions of the human genome. According to the ENCODE project, the fraction of regions in the human genome potentially involved in transcriptional control is many times greater than the fraction of coding regions. Thus, genetic variation in noncoding regions of the genome can increase the susceptibility to diseases by disrupting various regulatory elements (promoters, enhancers, silencers, insulator regions, etc.). However, identification of the mechanisms of influence of pathogenic genetic variants on the diseases risk is difficult due to a wide variety of regulatory elements. The present review focuses on the molecular genetic mechanisms by which pathogenic genetic variants affect gene expression. At the same time, attention is concentrated on the transcriptional level of regulation as an initial step in the expression of any gene. A triggering event mediating the effect of a pathogenic genetic variant on the level of gene expression can be, for example, a change in the functional activity of transcription factor binding sites (TFBSs) or DNA methylation change, which, in turn, affects the functional activity of promoters or enhancers. Dissecting the regulatory roles of polymorphic loci have been impossible without close integration of modern experimental approaches with computer analysis of a growing wealth of genetic and biological data obtained using omics technologies. The review provides a brief description of a number of the most well-known public genomic information resources containing data obtained using omics technologies, including (1) resources that accumulate data on the chromatin states and the regions of transcription factor binding derived from ChIP-seq experiments; (2) resources containing data on genomic loci, for which allele-specific transcription factor binding was revealed based on ChIP-seq technology; (3) resources containing *in silico* predicted data on the potential impact of genetic variants on the transcription factor binding sites.

Key words: transcription regulation; genetic variability; pathogenic genetic variants; transcription regulatory regions; transcription factor binding sites (TFBSs); genomic databases.

For citation: Ignatieva E.V., Matrosova E.A. Disease-associated genetic variants in the regulatory regions of human genes: mechanisms of action on transcription and genomic resources for dissecting these mechanisms. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2021;25(1):18-29. DOI 10.18699/VJ21.003

Введение

В настоящее время, во многом благодаря развитию технологии полногеномного и полноэкзомного анализа ассоциаций (ПГАА и ПЭАА), идентифицировано большое количество полиморфизмов, ассоциированных с заболеваниями. Так, ресурс GWAS central (<https://www.gwascentral.org/>) содержит информацию о 3.2 млн вариантов генетической изменчивости, ассоциированных с заболеваниями либо фенотипическими характеристиками (Beck et al., 2020). Сравнимые по объему массивы экспериментальных данных накоплены в ряде других баз по ассоциациям генотип-фенотип (GWAS catalog, OMIM, ClinVar, HGMD, PheGenI, EGA, GAD, dbGaP).

Несмотря на наличие значительных объемов экспериментальной информации о вариантах геномной изменчивости (ГИ), ассоциированных с заболеваниями, молекулярные механизмы, лежащие в основе этих ассоциаций, изучены крайне недостаточно. Это обусловлено тем, что лишь малая доля патогенных вариантов ГИ находится в кодирующих областях генома человека, изменение нуклеотидной последовательности которых нарушает строение и функцию белков. Огромная масса полиморфных локусов, связанных с заболеваниями, располагается в некодирующих участках генома (интронах, 5'- и 3'-фланкирующих районах генов, межгенных областях). Например, из общего числа вариантов, ассоциированных с заболеваниями по данным ПГАА, ~90 % локализованы вне кодирующих районов генов (Maurano et al., 2012; Farh et al., 2015).

Известно, что некодирующие области генома содержат участки, выполняющие широкий спектр регуляторных функций. Это промоторные районы, энхансеры, негативные регуляторные элементы, районы прикрепления к ядерному матриксу, районы, определяющие структуру топологически ассоциированных доменов хроматина (topologically associating domain, TAD) и другие особенности 3D укладки генома (Mathelier et al., 2015; Meddens et al., 2019; Ibrahim, Mundlos, 2020). В геноме человека

доля участков, потенциально вовлеченных в регуляцию транскрипции, чрезвычайно высока. По данным проекта ENCODE, участки хроматина, соответствующие пикам связывания с транскрипционными факторами (ТФ), выявленным методикой ChIP-seq, занимают ~8.1 % всей геномной ДНК (ENCODE Project Consortium, 2012), что заметно больше, чем доля кодирующих областей генома человека (~1.2 %). С учетом того, что в проекте ENCODE изучались не все известные ТФ и не все линии клеток, во взаимодействии с ТФ вовлечена заведомо большая доля геномной ДНК. Суммарная протяженность участков генома человека, которые имеют характеристики хроматина, свойственные энхансерам, также существенно превышает общий размер кодирующих областей: например, только в одном исследованном типе клеток (H1-ES) энхансерные участки занимают ~3.2 % (Roadmap Epigenomics Consortium et al., 2015).

Исследования, направленные на выявление механизмов влияния патогенных вариантов ГИ на предрасположенность к заболеваниям, ведутся очень активно, что нашло отражение в целом ряде публикаций обзорного характера (Mathelier et al., 2015; Merkulov et al., 2018; Smith et al., 2018; Wang et al., 2019; Vohra et al., 2020). Наиболее обсуждаемое проявление эффекта патогенных вариантов ГИ – изменение функциональных характеристик сайтов связывания транскрипционных факторов (ССТФ) (Lewinsky et al., 2005; Chen L. et al., 2013; Claussnitzer et al., 2015; Mathelier et al., 2015; Gorbacheva et al., 2018). Показано также, что полиморфные локусы могут быть ассоциированы с особенностями паттернов метилирования ДНК (Howard et al., 2014; Kumar D. et al., 2017; Rahbar et al., 2018; Schmitz et al., 2019) и модификаций гистоновых белков (Kilpinen et al., 2013; Visser et al., 2015; Zhang et al., 2018; Cong et al., 2019), формированием петель хроматина (Visser et al., 2015; Zhang et al., 2018) и, как одним из проявлений этого процесса, – с изменениями структуры TAD (Cong et al., 2019; Mei et al., 2019). Примеры подобных эффектов будут рассмотрены ниже (табл. 1).

Таблица 1. Примеры ассоциаций полиморфных локусов с патологиями и механизмы их влияния на уровень экспрессии генов

Заболевание/ патология	Полиморфный локус	Локализация	Механизм	Литературный источник
Атопическая астма	rs928413 A→G	Промоторный район гена <i>IL33</i>	Появление сайта связывания CREB1 приводит к повышению уровня экспрессии гена <i>IL33</i>	Gorbacheva et al., 2018
Ожирение	rs1421085 T→C	Интрон гена <i>FTO</i> содержит регуляторный район генов <i>IRX3</i> и <i>IRX5</i> (удаленных от него на 517 и 1164 тыс. нуклеотидов)	Нарушение сайта связывания фактора-репрессора ARID5B приводит к повышению уровня экспрессии генов <i>IRX3</i> и <i>IRX5</i>	Claussnitzer et al., 2015
Рак поджелудочной железы	rs2001389 A→G	Участок, ограничивающий TAD, на 10-й хромосоме человека	Нарушение сайта связывания CTCF приводит к изменению 3D структуры хроматина, что вызывает снижение уровня экспрессии гена-супрессора опухолевого роста <i>MFSD13A</i>	Mei et al., 2019
Нарушения липидного метаболизма	rs174537 G→T	Энхансер генов <i>FADS1</i> и <i>FADS2</i>	Повышение уровня метилирования регуляторных районов генов <i>FADS1</i> и <i>FADS2</i> приводит к снижению уровня экспрессии генов <i>FADS1</i> и <i>FADS2</i>	Howard et al., 2014
Атопический дерматит	rs612529 T→C	Промоторный район гена <i>VSTM1</i>	Нарушение сайта связывания фактора PU.1, иницирующего деметилирование ДНК (посредством привлечения деметилаз), приводит к повышению уровня метилирования промоторного участка гена <i>VSTM1</i> , что вызывает снижение уровня экспрессии гена <i>VSTM1</i>	Kumar D. et al., 2017
Синдром ломкой X-хромосомы	Увеличение числа тринуклеотидных повторов CGG с 5–55 (норма) до 100 и более	5'-НТФ гена <i>FMR1</i>	Нарушение структуры TAD, включающего ген <i>FMR1</i> . Граница между двумя TAD смещается в 3'-направлении относительно <i>FMR1</i> , ввиду чего ген <i>FMR1</i> оказался включенным в «чужой» TAD. В этом случае наблюдается гиперметилирование CpG островков в районе промотора, что приводит к снижению уровня экспрессии гена <i>FMR1</i>	Sun et al., 2018
Ревматоидный артрит, диабет 2-го типа	rs7873784 G→C	3'-НТФ гена <i>TLR4</i>	Появление сайта связывания фактора PU.1 повышает активность энхансера, расположенного в 3'-НТФ гена <i>TLR4</i> . Повышается активность промотора гена <i>TLR4</i> , что приводит к активации его экспрессии	Korneev et al., 2020
Рак молочной железы	rs4321755 C→T	Энхансер генов <i>MRPS30</i> и <i>RP11-53O19.1</i>	Появление сайта связывания фактора GATA3 повышает активность энхансера. Усиливаются контакты энхансера с двунаправленным промотором генов <i>MRPS30</i> и <i>RP11-53O19.1</i> , что приводит к активации их экспрессии	Zhang et al., 2018

Влияние геномной изменчивости на сайты связывания транскрипционных факторов

Ключевую роль в процессе регуляции транскрипции играют транскрипционные факторы – белки, способные специфически взаимодействовать с ДНК регуляторных районов генов и иницирующие формирование транскрипционных комплексов. В геноме человека содержится более 1500 генов, кодирующих ТФ (Wingender et al., 2013). Сайты связывания ТФ, как правило, имеют протяженность 10–25 нуклеотидов (Levitsky et al., 2014; Kulakovskiy et al., 2018).

Замены нуклеотидов, а также короткие инсерции/делеции в полиморфных локусах могут приводить к повреждению (разрушению) ССТФ либо их возникновению *de novo* (см. табл. 1), и это, в свою очередь, может оказывать как негативный, так и позитивный эффект на уровень транс-

крипции генов (Chen L. et al., 2013; Gorbacheva et al., 2018). Такие варианты ГИ (и соответствующие полиморфные локусы), влияющие на транскрипционную активность генов, принято называть регуляторными (Kumar S. et al., 2017; Guo, Wang, 2018; Merkulov et al., 2018).

Патологическим (т. е. ассоциированным с заболеванием) может оказаться как аллельный вариант последовательности ДНК, содержащий нарушенный ССТФ (Lewinsky et al., 2005; Chen L. et al., 2013; Claussnitzer et al., 2015; Kumar D. et al., 2017; Mei et al., 2019), так и аллельный вариант, в котором идентифицируется возникновение сайта *de novo* (Gorbacheva et al., 2018; Zhang et al., 2018; Korneev et al., 2020) (см. табл. 1).

Патологические варианты ГИ, влияющие на функциональность ССТФ, могут располагаться не только в промоторных участках, но также в удаленных регуляторных

районах: энхансерах (Lewinsky et al., 2005; Zhang et al., 2018; Meddens et al., 2019), регуляторных районах с репрессорной функцией (Claussnitzer et al., 2015) и участках, ограничивающих TAD (Mei et al., 2019) (см. табл. 1). Так, например, ассоциированная с ожирением замена Т→С в полиморфном локусе rs1421085 нарушает функционирование негативного регуляторного района генов *IRX3* и *IRX5* (Claussnitzer et al., 2015). Локус rs1421085 расположен в интроне гена *FTO* (рис. 1) и находится на значительном расстоянии от стартов транскрипции генов *IRX3* и *IRX5* (520 и 1164 тыс. нуклеотидов). В норме участок ДНК, содержащий rs1421085 (аллель Т), связывается с транскрипционным фактором-репрессором ARID5B, что способствует снижению транскрипционной активности генов *IRX3* и *IRX5*. У носителей мутантного варианта последовательности ДНК (аллель С) сайт связывания фактора-репрессора ARID5B оказывается нарушенным, что приводит к чрезмерно высокой экспрессии генов *IRX3* и *IRX5* и ускоренному росту клеток жировой ткани (Claussnitzer et al., 2015).

Возможны ситуации, когда замена нуклеотида в полиморфном локусе нарушает ССТФ, и это, в свою очередь, отражается на функционировании TAD (см. табл. 1). Такой эффект выявлен в случае локуса rs2001389 (А→G), ассоциированного с риском рака поджелудочной железы (рис. 2). Локус rs2001389 расположен в участке, определяющем структуру петель хроматина в пределах TAD, содержащего 91 ген и образованного пространственно сближенными участками хроматина (Mei et al., 2019). Участок ДНК, содержащий патологический аллель G, характеризуется пониженной способностью к связыванию с транскрипционным фактором CTCF, который в данном случае выполняет функцию структурного белка хроматина. В норме связывание с фактором CTCF обеспечивает функционирование одного из участков, определяющих структуру петель хроматина в пределах рассматриваемого TAD. Нарушение связывания с фактором CTCF влечет за собой изменение 3D структуры хроматина, что нарушает экспрессию генов, содержащихся в данном TAD. При этом в наибольшей степени оказывается подавленной экспрессия гена-супрессора опухолевого роста *MFSD13A*.

Влияние геномной изменчивости на метилирование ДНК и транскрипционная активность генов

Метилирование ДНК – это модификация, которая не изменяет нуклеотидной последовательности и заключается в присоединении метильной группы к пятому атому углерода цитозина (Angeloni, Bogdanovic, 2019). Повышение уровня метилирования ДНК, как правило, приводит к долгосрочной инактивации экспрессии генов, лежащих в метилированном участке, так как, согласно общепризнанной концепции, метилирование участка ДНК способствует посадке на этот участок белковых комплексов, включающих гистон деацетилазу (HDAC) (Jones et al., 1998; Nan et al., 1998). Метилирование может также препятствовать взаимодействию транскрипционных факторов с ДНК: известно, что такой чувствительностью к метилированию обладают факторы CTCF и факторы из семейства ETS (Wang et al., 2019). Другой транскрипци-

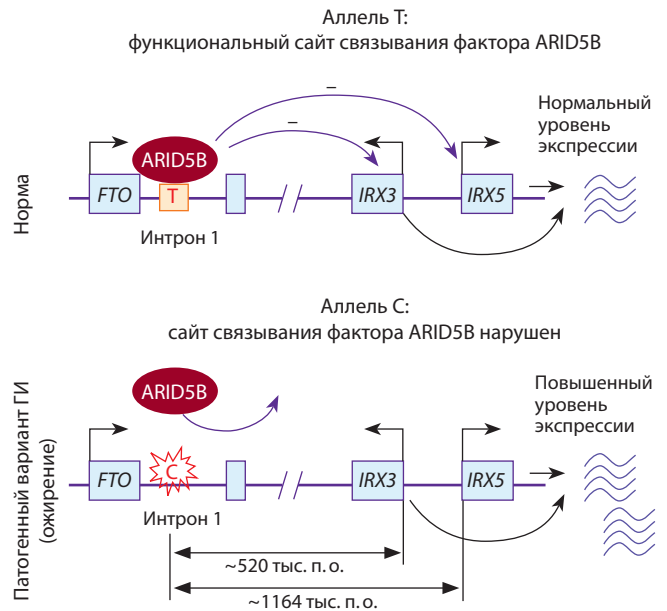


Рис. 1. Нарушение сайта связывания, вызванное заменой нуклеотида Т→С в локусе rs1421085, препятствует взаимодействию фактора-репрессора ARID5B с регуляторным районом генов *IRX3* и *IRX5*, в результате чего уровень экспрессии *IRX3* и *IRX5* повышается.

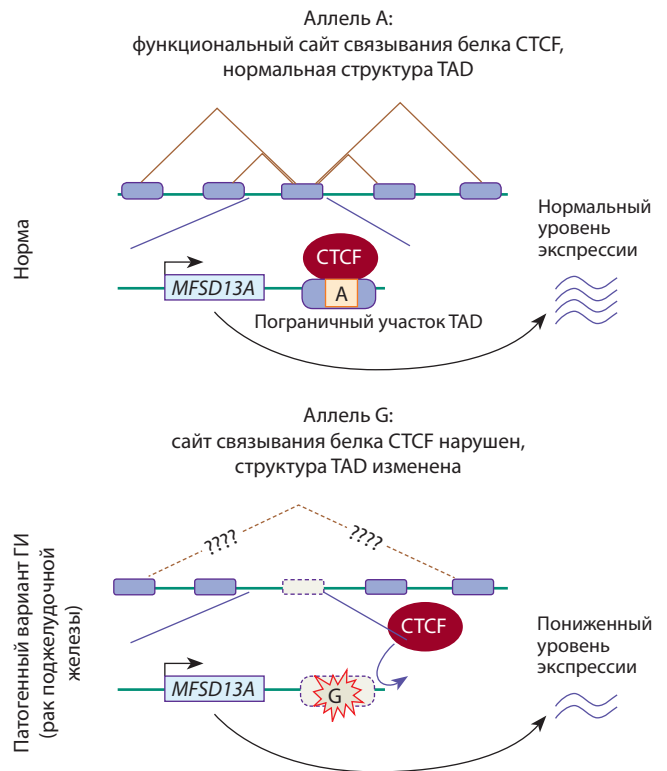


Рис. 2. Нарушение сайта связывания белка CTCF, вызванное заменой нуклеотида rs2001389, приводит к исчезновению одного из пограничных участков, определяющих структуру TAD, вследствие чего происходит снижение уровня экспрессии гена-супрессора опухолевого роста *MFSD13A*.

Контакты между участками хроматина в пределах TAD показаны коричневыми линиями. Знаки вопроса на нижнем рисунке обозначают отсутствие точных сведений о структуре TAD.

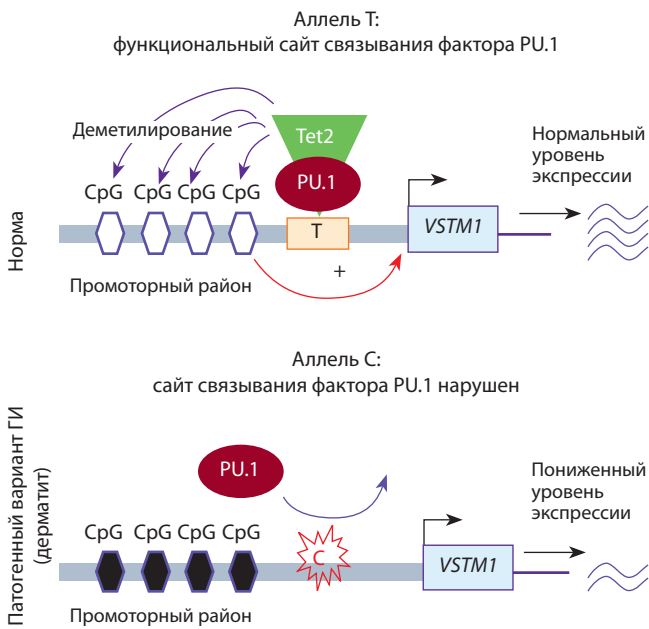


Рис. 3. Нарушение сайта связывания PU.1, вызванное заменой нуклеотида Т→С (rs612529), снижает активность деметилаз (например, Tet2), поддерживающих промоторный участок гена *VSTM1* в активном состоянии, в связи с чем экспрессия *VSTM1* снижается.

онный фактор, ZFP57, напротив, взаимодействует только с метилированной ДНК (Quenneville et al., 2011). Таким образом, метилирование цитозина может активировать различные механизмы регуляции транскрипции генов, и не всегда повышение уровня метилирования регуляторного участка ДНК сопряжено со снижением экспрессии соответствующего гена (Izzi et al., 2016; Wang et al., 2019).

Геномная изменчивость оказывает существенное влияние на метилирование участков ДНК, обладающих регуляторным потенциалом. Так, в ходе полногеномного анализа паттернов метилирования геномов 24 жителей острова Норфолк (Benton et al., 2019) был выявлен 12761 район, содержащий не менее двух CpG-динуклеотидов и имеющий аллель-специфичный уровень метилирования. В большинстве случаев (98 %) расположение районов совпадало с позициями однонуклеотидных замен, представленных в dbSNP (Benton et al., 2019).

В этой же работе (Benton et al., 2019) был проведен анализ совместного расположения районов аллель-специфичного метилирования и набора полиморфных локусов из базы данных GWAS catalogue, ассоциированных с заболеваниями человека. Оказалось, что полиморфные локусы, ассоциированные с заболеваниями, в два раза чаще перекрываются с районами аллель-специфичного метилирования, чем можно было ожидать по случайным причинам. Это означает, что изменение уровня метилирования, обусловленное ГИ, является одним из факторов, повышающих риск развития заболеваний.

В качестве примера рассмотрим полиморфный локус rs174537 (G→T), расположенный в энхансере генов *FADS1* и *FADS2*, кодирующих десатуразы жирных кислот 1 и 2. Вариант Т локуса rs174537 ассоциирован с повышенным риском патологических нарушений липидного метаболиз-

ма (см. табл. 1). Было показано, что носители аллеля Т в локусе rs174537 имели более высокий уровень метилирования регуляторных районов генов *FADS1* и *FADS2* в печени (Howard et al., 2014), что приводило к подавлению транскрипционной активности генов *FADS1* и *FADS2*.

Возможны ситуации, когда в одном из аллельных вариантов происходит деметилирование ДНК, инициированное связыванием с транскрипционным фактором (см. табл. 1). Такой механизм выявлен, например, при исследовании полиморфного локуса rs612529 Т→С. Локус расположен в промоторном районе гена *VSTM1* (рис. 3), низкая экспрессия которого в моноцитах провоцирует развитие дерматита. В этом типе клеток промоторный участок, содержащий протективный вариант Т, более активно взаимодействует с транскрипционным фактором PU.1, инициирующим деметилирование ДНК посредством привлечения деметилаз (например, Tet2). Как следствие, у носителей аллеля Т промоторный участок гена *VSTM1* оказывается полностью деметилированным, и наблюдается активная экспрессия данного гена. У носителей патогенного варианта С взаимодействие фактора PU.1 с ДНК нарушается, в результате чего промоторный участок более сильно метилирован, что сопровождается снижением экспрессии гена *VSTM1* (Kumar D. et al., 2017).

Влияние геномной изменчивости на состояние хроматина и его пространственную структуру

Присутствие патогенных вариантов ГИ отражается и на состоянии хроматина (Kilpinen et al., 2013). Известны случаи, когда наличие патогенного варианта ГИ сопровождалось изменением паттернов модификации гистонов и появлением (либо исчезновением) участков гиперчувствительности к ДНКазе I типа (McVicker et al., 2013; Visser et al., 2015; Zhang et al., 2018; Cong et al., 2019). В этих случаях были выявлены аллель-специфичные контакты между промоторами и энхансерами, количество которых коррелировало с активностью энхансерных районов.

Также известны ситуации, когда структурные варианты генома (инсерции, делеции, дупликации, инверсии, транслокации длиной более 50 нуклеотидов) приводят к изменению пространственной организации хроматина, нарушая тем самым экспрессию генов, связанных с патологическими процессами (Sun et al., 2018; Ibrahim, Mundlos, 2020). Например, ассоциированная с синдромом ломкой Х-хромосомы экспансия тринуклеотидных повторов CGG в 5'-нетранслируемом районе (НТР) гена *FMR1* нарушает структуру TAD, включающего *FMR1* (рис. 4, см. табл. 1). В норме *FMR1* находится очень близко к 5'-пограничной области TAD (на рис. 4 этот домен обозначен как TAD1), а участок ДНК, соответствующий 5'-пограничной области TAD1, гипометилирован и взаимодействует с фактором CTCF. У индивидов, имеющих повышенное количество триплетных повторов CGG (более 100), данный участок ДНК перестает выполнять барьерную функцию (он гиперметилирован и не связывается с фактором CTCF). Нарушение барьерной функции 5'-пограничной области TAD1 приводит к исчезновению TAD1 и расширению границ другого TAD (см. TAD2 на рис. 4). Ген *FMR1* оказывается включенным в чужеродный ему TAD2. В этом случае

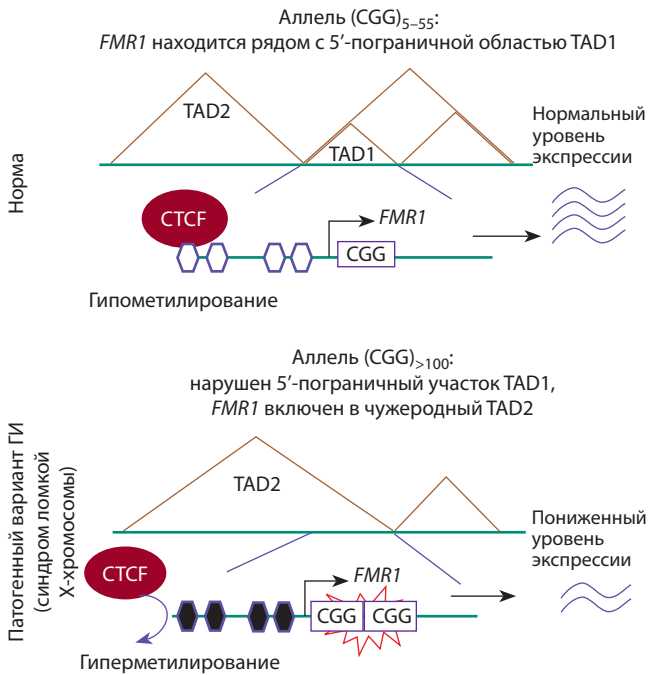


Рис. 4. При увеличении количества триплетных повторов CGG в 5'-нетранслируемом районе гена *FMR1* происходит гиперметилирование участка ДНК, соответствующего пограничной области TAD1. Это приводит к нарушению связывания факторов CTCF и нарушению барьерной функции пограничного участка.

Коричневыми линиями показаны контакты между участками хроматина в пределах TAD.

промоторный район гена *FMR1* также оказывается гиперметилированным, а экспрессия гена – резко сниженной (Park et al., 2015; Sun et al., 2018).

Для исследования молекулярно-генетических механизмов влияния геномной изменчивости на 3D структуру хроматина необходимо реконструировать пространственную укладку генома. В 3D структуре геномов выявлены такие базовые уровни организации, как: регуляторные петли ДНК со сближенными промоторами и энхансерами; топологически ассоциированные домены, внутри которых участки ДНК контактируют друг с другом чаще, чем с соседними доменами; А и В компартменты, соответствующие транскрипционно-активному и конденсированному хроматину; и наконец, хромосомные территории (Fishman et al., 2018; Hansen et al., 2018). Нарушение 3D контактов между промоторами и энхансерами в пределах TAD, вызванное, например, хромосомными перестройками, может существенно повлиять на транскрипционную активность гена, приводя к развитию патологий (Lupiáñez et al., 2015).

В Институте цитологии и генетики СО РАН разработан экспериментально-компьютерный подход к предсказанию физических контактов между промоторами и энхансерами в 3D структуре хроматина (Fishman et al., 2018; Belokopytova et al., 2020; Belokopytova, Fishman, 2021). Подход основан на использовании следующей информации: 1) тип клеток; 2) клеточно-специфическая локализация энхансеров в линейном геноме (из базы данных ENCODE); 3) транскрипционная активность промоторов (из экспериментов RNA-seq); 4) границы экструзии пе-

тель хроматина (из экспериментов ChIP-seq с фактором CTCF); 5) ориентация мотивов связывания CTCF-белков (из анализа геномов); 6) А либо В компартмент хроматина (по данным экспериментов Hi-C). Анализ этих данных с помощью оригинальной программы 3DPredictor (Belokopytova et al., 2020), разработанной на основе алгоритмов машинного обучения, позволяет предсказывать частоты физических контактов между промоторами и энхансерами в 3D структуре генома с точностью, превышающей точность других известных методов предсказания.

С помощью программы 3DPredictor была проанализирована 3D структура генома у мышей *DelB/DelB*, гомозиготных по делеции геномного участка длиной 1.5 Mb, содержащего ген *Epha4*. Такая делеция сопровождается появлением добавочных контактов между геном *Pax3* и энхансерным районом гена *Epha4*, что нарушает экспрессию гена *Pax3* и приводит к брахидактилии. Мыши с генотипом *DelB/DelB* являются генетической моделью патологии человека, выражающейся в нарушениях формирования конечностей (Lupiáñez et al., 2015). Тестирование 3DPredictor на этом модельном объекте продемонстрировало высокую эффективность программы: у мышей с нарушенным генотипом были предсказаны новые добавочные контакты между генами и удаленными регуляторными элементами (Belokopytova et al., 2020), и предсказания хорошо соответствовали экспериментальным данным.

Геномная изменчивость: комплексный анализ больших гетерогенных генетических данных

Как отмечено выше, многие варианты ГИ, ассоциированные с заболеваниями, находятся на значительном расстоянии от кодирующих областей генов (ENCODE Project Consortium, 2012; Maurano et al., 2012). Для идентификации молекулярно-генетических механизмов влияния таких вариантов ГИ на предрасположенность к заболеваниям необходимы дополнительные исследования. Их целью является выяснение регуляторной роли вариантов ГИ. Типичный пример – работа (Zhang et al., 2018), позволившая найти функционально значимый регуляторный вариант rs4321755, ассоциированный с риском рака молочной железы. Лocus rs4321755 располагается в удаленном энхансере, регулирующем экспрессию генов *MRPS30* и *RP11-53019.1* (см. табл. 1). Оказалось, что при наличии патогенного варианта T в локусе rs4321755 формируется новый сайт связывания фактора GATA3. Транскрипционный фактор GATA3 повышает функциональную активность энхансера, что проявляется в более интенсивных контактах энхансера с двунаправленным промотором генов *MRPS30* и *RP11-53019.1* и активации их экспрессии. Чтобы выявить этот функционально значимый регуляторный вариант ГИ, авторы разработали интегрированный экспериментально-компьютерный метод, основанный на комплексном анализе больших гетерогенных генетических данных, включая данные об аллель-специфичной экспрессии генов, полученные на основе технологии RNA-seq в сочетании с данными о гаплотипах; о локусах количественных характеристик экспрессии (eQTL); об участках генома, чувствительных к ДНКазе I; о локализации ChIP-seq пиков из баз ENCODE и GEO; о локализации

Таблица 2. Информационные ресурсы по геномным данным, полученным на основе современных высокопроизводительных экспериментальных методов

Ресурс/база данных	URL	Характеристика ресурса
Аннотация генома человека		
GENCODE*	https://www.encodegenes.org/	Аннотация генома человека, составленная на основе слияния результатов, полученных с помощью ручной аннотации, с результатами компьютерной аннотации генов
Геномная изменчивость в популяциях человека		
НарМaп (Haplotype Map)	https://www.genome.gov/10001688/international-hapmap-project ftp://ftp.ncbi.nlm.nih.gov/hapmap/	Гаплотипы и гапоблоки генома человека, а также маркирующие их репрезентативные полиморфные локусы
1000 Genomes Project (1KGP)	https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/	Генетические варианты (однонуклеотидные полиморфизмы, инсерции/делеции, структурные варианты) и генотипы, выявленные у индивидов из 26 популяций
International Genome Sample Resource (IGSR)	https://www.internationalgenome.org	Объединение данных проекта «1000 геномов» с данными, полученными методикой RNA-Seq (проект GEUVADIS), и данными проекта ENCODE, полученными на линии клеток NA12878
dbSNP	https://www.ncbi.nlm.nih.gov/snp/	Однонуклеотидные генетические варианты, микросателлиты, инсерции и делеции в геномах различных видов организмов (включая человека). Локализация полиморфных локусов на хромосомах, популяционные частоты. dbSNP накапливает как данные массового анализа, полученные в геномных проектах, так и результаты исследования отдельных локусов, представленные в публикациях
Генетические варианты, ассоциированные с заболеваниями		
GWAS central (Genome-wide association studies central)	https://www.gwascentral.org/	Частоты аллелей и генотипов человека, а также их ассоциации с фенотипами (либо патологиями) из различных источников (публикаций и исследовательских проектов)
GWAS catalog (Genome-wide association studies catalog)	https://www.ebi.ac.uk/gwas/home	Ассоциации между полиморфными локусами и фенотипическими признаками, полученные с помощью методики ПГАА
OMIM (Online Mendelian Inheritance in Man)	https://www.ncbi.nlm.nih.gov/omim	Каталог, описывающий гены человека, их генетические варианты и генетически обусловленные заболевания и синдромы человека. Данные внесены в каталог командой экспертов на основе анализа научных публикаций
ClinVar (Clinical Variations)	https://www.ncbi.nlm.nih.gov/clinvar/	Ассоциации между генами и генетическими вариантами генома человека и фенотипическими признаками
HGMD (The Human Gene Mutation Database)	http://www.hgmd.cf.ac.uk/ac/index.php	Генетические нарушения, связанные с наследственными заболеваниями человека
PheGenI (The Phenotype-Genotype Integrator)	https://www.ncbi.nlm.nih.gov/gap/phegeni	Ресурс, интегрирующий данные из GWAS catalog с данными из нескольких баз, размещенных в NCBI, включая Gene, dbGaP, OMIM, eQTL и dbSNP
EGA (The European Genome-phenome Archive)	https://ega-archive.org/	Данные о связи генотипов и фенотипов, полученные различными экспериментальными методами (ПГАА, экзомное и полногеномное секвенирование, генотипирование, секвенирование геномов отдельных клеток)
dbGaP (The database of Genotypes and Phenotypes)	https://www.ncbi.nlm.nih.gov/gap/	Данные по ассоциациям между генотипами и фенотипами человека, полученные различными методами (ПГАА, экзомное секвенирование, генотипирование когорт, исследования на близнецах и т.д.)
Данные о состоянии хроматина		
ENCODE (The Encyclopedia of DNA Elements)	http://genome.ucsc.edu/ENCODE/ https://www.encodeproject.org/	Полногеномные профили модификации гистонов, метилирования ДНК, районы связывания с ТФ (по данным ChIP-seq), области контактов между удаленными участками ДНК, участки открытого хроматина, экспрессионные данные для более 300 типов клеток
NIH Roadmap Epigenomics Mapping Consortium	http://www.roadmapepigenomics.org/	Данные, полученные с помощью методик ChIP-seq, RNA-seq, бисульфитного секвенирования. Аннотация генома человека в соответствии с классификациями состояний хроматина (15, 18 и 25 типов хроматина)

Окончание табл. 2

Ресурс/база данных	URL	Характеристика ресурса
Локусы количественных характеристик экспрессии eQTL		
Genotype-Tissue Expression (GTEx) project	https://www.gtexportal.org/home/	Данные по экспрессии и eQTL в 54 типах клеток человека, имеющих здоровый фенотип
eQTL databases	https://www.hsph.harvard.edu/liming-liang/software/eqtl/	eQTL в лимфобластоидных линиях клеток
exSNP	http://www.exsnp.org/	eQTL и их связи с заболеваниями человека
eQTL Catalogue	https://www.ebi.ac.uk/eqtl/	Cis-eQTL и QTL, выявленные на основе анализа данных публикаций, а также из проекта GTEx
eQTL Browser	http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/	eQTL, выявленные на основе анализа данных из научных публикаций
Коллекции экспериментов ChIP-seq, направленных на идентификацию ССТФ		
Cistrome Data Browser	http://cistrome.org/db/#/	Данные экспериментов ChIP-seq, DNase-seq и ATAC-seq, выявляющих в геномах человека и мыши (1) участки хроматина, взаимодействующие с ТФ; (2) участки хроматина, доступные для действия эндонуклеазы; (3) участки, содержащие посттрансляционные модификации гистонов. Данным присвоены статусы согласно шести критериям качества
Gene Transcription Regulation Database (GTRD)	https://gtrd.biouml.org/#!	Коллекция экспериментов ChIP-seq, направленных на поиск сайтов связывания ТФ в геноме человека и мыши
ReMap (Global map of regulatory elements)	http://remap.univ-amu.fr/	Коллекция экспериментов ChIP-seq, ChIP-exo, DAP-seq из публичных ресурсов (GEO, ENCODE, ENA). Участки хроматина, контактирующие с ТФ, транскрипционными коактиваторами, хроматин-ремоделлирующими факторами
Аллель-специфичное связывание ТФ, выявленное на основе анализа данных экспериментов ChIP-seq в комбинации с данными о генотипах исследованных клеток		
AlleleDB	http://alleledb.gersteinlab.org/	Данные по аллель-специфичному связыванию ТФ, полученные на основе анализа экспериментов ChIP-seq для образцов от 383 индивидов, чьи геномы были секвенированы в ходе выполнения проекта «1000 геномов»
AlleleSeq	http://alleleseq.gersteinlab.org/	Данные по аллель-специфичному связыванию для шести ТФ (cFos, cMyc, JunD, Max, NfkB, CTCF), полученные при анализе данных ChIP-seq в лимфобластоидной клеточной линии GM12878
Потенциальные эффекты вариантов геномной изменчивости на ССТФ		
HaploReg	https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php	Аннотация вариантов ГИ. Приведены данные о состоянии хроматина, сцеплении, консервативности, перекрытии с регуляторными мотивами, eQTL
SNP2TFBS	http://ccg.vital-it.ch/snp2tfbs/	Варианты ГИ, идентифицированные проектом «1000 геномов», изменяющие сходство ССТФ с весовыми матрицами
rSNPBase	http://rsnp3.psych.ac.cn/index.do	Эффекты ОНП на ССТФ, регулируемые гены, регуляторные сети
rVarBase	http://rv.psych.ac.cn/	Эффекты вариантов ГИ (включая вариации числа копий) на потенциальные ССТФ, данные о состоянии хроматина и регулируемых генах
Информационные ресурсы широкого профиля		
UCSC Genome Browser	https://genome.ucsc.edu/	Интеграция на основе графического интерфейса данных о первичных последовательностях и аннотации геномов, а также характеристиках геномных районов (нуклеотидном составе, геномной изменчивости, состояниях хроматина, экспрессии, контактах между участками хроматина и т.д.). Программное средство UCSC table browser позволяет извлекать данные в текстовом виде
Ensembl Genome Browser	https://www.ensembl.org/index.html	Интеграция на основе графического интерфейса данных о первичных последовательностях и аннотации геномов, а также характеристиках геномных районов (нуклеотидном составе, геномной изменчивости, состояниях хроматина и т.д.). Программное средство BioMart data mining tool позволяет извлекать данные в текстовом виде
GEO (Gene Expression Omnibus)	https://www.ncbi.nlm.nih.gov/gds	Крупнейший репозиторий данных по функциональной геномике человека и других видов организмов, полученных на основе омиксных технологий (экспрессия, профили состояния хроматина, генотипирование и др.)

* Аннотация генома человека из базы GENCODE доступна также через UCSC Genome Browser (<https://genome.ucsc.edu/>) и Ensembl genome browser (<https://www.ensembl.org/index.html>).

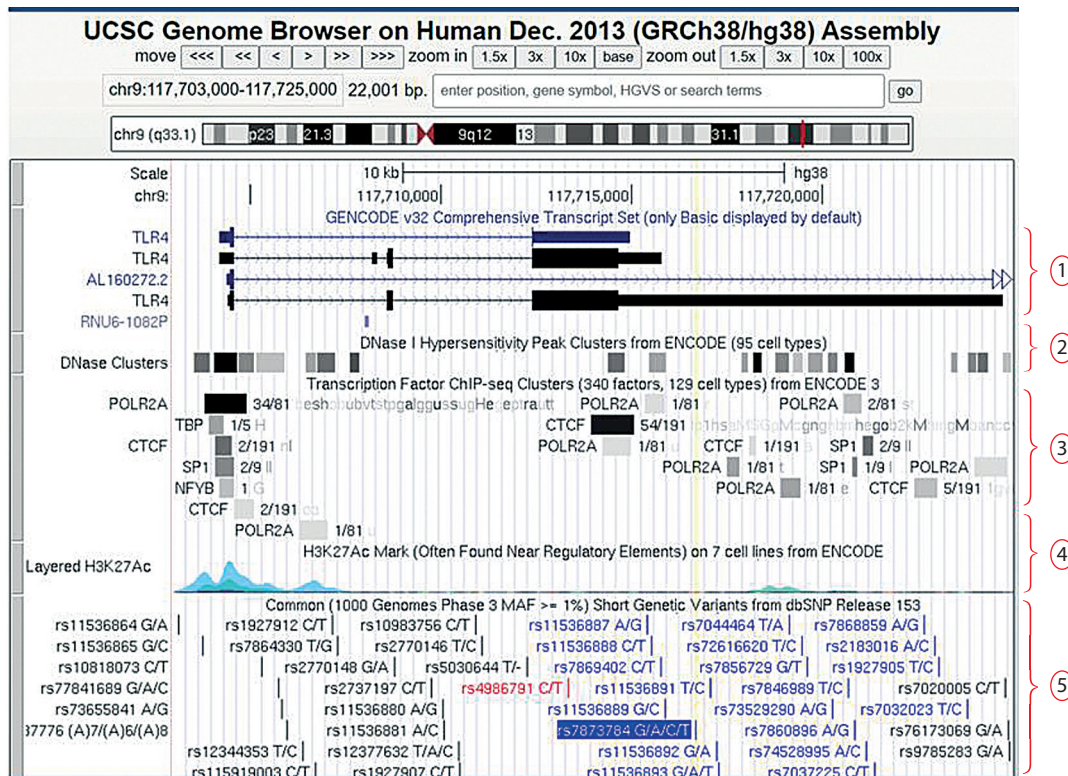


Рис. 5. Пример графического представления информации об участке 9-й хромосомы человека (хромосомные координаты chr9:117,703,000–117,725,000) в геномном браузере Университета г. Санта-Круз, США (UCSC Genome Browser, <https://genome.ucsc.edu/>).

1 – транскрипты гена *TLR4*, позиции которых приведены согласно данным базы GENCODE; 2 – участки гиперчувствительности к ДНКазе I типа; 3 – участки хроматина, для которых с помощью методики ChIP-seq (данные проекта ENCODE) показано взаимодействие с ТФ; 4 – участки хроматина, содержащие модифицированный гистоновый белок H3 (модификация H3K27Ac часто присутствует в участках, имеющих регуляторные функции); 5 – позиции вариантов геномной изменчивости. Желтой вертикальной линией отмечено расположение варианта ГИ rs7873784, локализованного в 3'-HTP гена *TLR4* и ассоциированного с ревматоидным артритом и диабетом 2-го типа (см. табл. 1). По данным (Korneev et al., 2020), замена G→C в локусе rs7873784 приводит к возникновению сайта связывания транскрипционного фактора PU.1, что повышает активность энхансера, расположенного в 3'-HTP гена *TLR4*.

регуляторных мотивов, предсказанных компьютерными программами. Сходные сценарии интегрированных экспериментально-компьютерных исследований были реализованы и в других работах (Chen C.-Y. et al., 2014; Claussnitzer et al., 2015; Zhao et al., 2019; Li et al., 2020).

Проведение исследований подобного рода стало возможным благодаря развитию современных высокопроизводительных экспериментальных подходов, позволяющих получать различные типы данных в масштабе всего генома (параллельное высокопроизводительное секвенирование, методики ChIP-seq, 3C, Hi-C, ChIA-PET, футпринтинг ДНК с использованием ДНКазы I типа, бисульфитное секвенирование и т. д.), и наличию открытых информационных ресурсов, накапливающих подобные экспериментальные данные. В табл. 2 приведена краткая характеристика информационных ресурсов, содержащих геномные данные, полученные на основе омиксных технологий, и используемых для изучения механизмов повреждающего влияния ГИ на транскрипцию генов. Это данные по аннотации генома человека (GENCODE); о геномной изменчивости в популяциях человека (MapMap, 1000 Genomes Project, IGSR, dbSNP); данные о вариантах ГИ, ассоциированных с заболеваниями (GWAS central, GWAS catalog, ClinVar,

HGMD, OMIM и др.); о состоянии хроматина (ENCODE, NIH Roadmap Epigenomics Mapping Consortium); о локусах количественных характеристик экспрессии (GTEx project, eQTL databases, exSNP и др.); об экспериментах ChIP-seq, направленных на идентификацию ССТФ (Cistrome Data Browser, GTRD, ReMap); об аллель-специфичном связывании ТФ, выявленном на основе анализа данных экспериментов ChIP-seq в комбинации с данными о генотипах исследованных клеток (AlleleDB, AlleleSeq); об эффектах вариантов ГИ на ССТФ, предсказанных на основе компьютерного анализа (HaploReg, SNP2TFBS, rSNPBase, rVarBase).

К отдельной категории информационных ресурсов относятся геномный браузер Университета г. Санта-Круз, США (UCSC Genome Browser, <https://genome.ucsc.edu/>) и геномный браузер базы Ensembl совместного научного проекта Европейского института биоинформатики и Института Сенгера (Ensembl Genome Browser, <https://www.ensembl.org/index.html>). Они используются как средства интеграции данных о характеристиках геномных районов, полученных разными экспериментальными методами и из разных информационных источников (Lee et al., 2020; Yates et al., 2020). Веб-сайты этих браузеров обеспечивают

доступ к первичным последовательностям и аннотациям геномов многих видов организмов, включая позвоночных животных и ряд других модельных видов. Графические интерфейсы браузеров позволяют в интерактивном режиме получать масштабируемые карты геномных районов и отображать на картах различные характеристики (например, локализацию транскриптов, вариантов геномной изменчивости, участков хроматина, взаимодействующих с ТФ по данным методики ChIP-seq, участков гиперчувствительности к ДНКазе I типа и т. д.) (рис. 5).

Веб-сайты геномных браузеров UCSC Genome Browser и Ensembl Genome Browser оснащены программными средствами доступа к данным в текстовом виде: UCSC table browser (<https://genome.ucsc.edu/cgi-bin/hgTables>) и BioMart data mining tool (<https://www.ensembl.org/info/data/biomart/index.html>).

Информационные ресурсы по аллель-специфичному связыванию транскрипционных факторов и предсказанным *in silico* эффектам вариантов геномной изменчивости на ССТФ

Как отмечалось выше, достаточно часто влияние патогенных вариантов ГИ на экспрессию генов реализуется через изменение функциональной активности сайтов связывания транскрипционных факторов. В связи с этим чрезвычайно полезными могут оказаться информационные ресурсы, включающие полногеномные данные об аллель-специфичном связывании ТФ, идентифицированном с помощью методики ChIP-seq. Разработано несколько подходов к выявлению аллель-специфичного связывания ТФ (Rozowsky et al., 2011; Reddy et al., 2012; Waszak et al., 2014; Younesy et al., 2014). Эти подходы основаны на анализе данных экспериментов ChIP-seq в комбинации с данными секвенирования, позволяющими выявлять гетерозиготные позиции в геноме и генотип исследуемых клеток. Таким образом, для каждого обследованного типа клеток может быть выявлен свой набор геномных локусов, взаимодействующих с конкретным транскрипционным фактором аллель-специфичным образом. Например, в работе (Cavalli et al., 2016a) были проанализированы данные экспериментов ChIP-seq для 55 ТФ в клеточной линии HepG2 и 57 ТФ в линии HeLa-S3. В клетках HepG2 был найден 3001 локус ГИ, имеющий аллель-специфичные сигналы, а в клетках HeLa-S3 обнаружено 712 таких локусов. Авторы отмечают выраженный тканеспецифичный характер аллель-специфичного связывания ТФ: из всего набора выявленных локусов только 34 были обнаружены в обеих клеточных линиях (Cavalli et al., 2016a).

Данные об аллель-специфичном связывании ТФ представлены в информационных ресурсах: AlleleDB (<http://alleledb.gersteinlab.org/>) (Chen J. et al., 2016), AlleleSeq (<http://alleleseq.gersteinlab.org/>) (Rozowsky et al., 2011) (см. табл. 2), а также в виде дополнительных материалов к публикациям (Cavalli et al., 2016a, b, 2019; Shi et al., 2016).

Исследования, направленные на идентификацию аллель-специфичного связывания ТФ, позволили оценить количество генетических вариантов, влияющих на связывание конкретного транскрипционного фактора на ДНК в конкретном типе клеток. Среднее количество таких событий, зарегистрированных для отдельного транскрип-

ционного фактора, может составлять от 19 до 37 для клеток с нормальным кариотипом (GM12878, H1-hESC) и от 12 до 55 для клеток с раковым кариотипом (SK-N-SH, K562) (Cavalli et al., 2016a, b).

При построении гипотез о механизмах влияния генетических вариантов на риск развития патологий также могут быть использованы данные об эффектах вариантов ГИ на ССТФ, предсказанных *in silico* на основе компьютерных программ распознавания ССТФ. Подобные сведения содержатся в специализированных базах данных: HaploReg (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>) (Ward, Kellis, 2012), SNP2TFBS (<http://ccg.vital-it.ch/snp2tfbs/>) (Kumar S. et al., 2017), rSNPBase (<http://rsnp3.psych.ac.cn/index.do>) (Guo, Wang, 2018), rVarBase (<http://rv.psych.ac.cn>) (см. табл. 2).

Закключение

Существенная доля патогенных генетических вариантов, ассоциированных с заболеваниями, локализована в некодирующих областях генома. Такие генетические варианты могут с большой долей вероятности нарушать функционирование регуляторных районов, контролирующих транскрипционную активность генов. Наглядным подтверждением этой возможности являются рассмотренные в нашем обзоре примеры механизмов влияния патогенных генетических вариантов на экспрессию генов. Исследования, позволившие идентифицировать такие механизмы, носят комплексный характер и основаны на анализе больших гетерогенных генетических данных. Имеющийся в настоящее время арсенал информационных ресурсов, содержащих омиксные данные, обеспечивает широкие возможности для подобных исследований. В будущем, с развитием экспериментальных технологий и биоинформатических методов анализа полученных с их помощью данных, а также с расширением спектра исследуемых типов клеток, эти возможности еще более возрастут.

Список литературы / References

- Angeloni A., Bogdanovic O. Enhancer DNA methylation: implications for gene regulation. *Essays Biochem.* 2019;63(6):707-715. DOI 10.1042/EBC20190030.
- Beck T., Shorter T., Brookes A.J. GWAS Central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. *Nucleic Acids Res.* 2020;48(D1):D933-D940. DOI 10.1093/nar/gkz895.
- Belokopytova P., Fishman V. Predicting genome architecture: challenges and solutions. *Front. Genet.* 2021. DOI 10.3389/fgene.2020.617202.
- Belokopytova P.S., Nuriddinov M.A., Mozheiko E.A., Fishman D., Fishman V. Quantitative prediction of enhancer-promoter interactions. *Genome Res.* 2020;30(1):72-84. DOI 10.1101/gr.249367.119.
- Benton M.C., Lea R.A., Macartney-Coxson D., Sutherland H.G., White N., Kennedy D., Mengersen K., Haupt L.M., Griffiths L.R. Genome-wide allele-specific methylation is enriched at gene regulatory regions in a multi-generation pedigree from the Norfolk Island isolate. *Epigenetics Chromatin.* 2019;12(1):60. DOI 10.1186/s13072-019-0304-7.
- Cavalli M., Baltzer N., Umer H.M., Grau J., Lemnian I., Pan G., Wallerlman O., Spalinskas R., Sahlén P., Grosse I., Komorowski J., Wadelius C. Allele specific chromatin signals, 3D interactions, and motif predictions for immune and B cell related diseases. *Sci. Rep.* 2019;9(1):2695. DOI 10.1038/s41598-019-39633-0.

- Cavalli M., Pan G., Nord H., Wallén Artzt E., Wallerman O., Wadelius C. Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals. *Genomics*. 2016a;107(6):248-254. DOI 10.1016/j.ygeno.2016.04.006.
- Cavalli M., Pan G., Nord H., Wallerman O., Wallén Artzt E., Berggren O., Elvers I., Eloranta M.L., Rönnblom L., Lindblad Toh K., Wadelius C. Allele-specific transcription factor binding to common and rare variants associated with disease and gene expression. *Hum. Genet.* 2016b;135(5):485-497. DOI 10.1007/s00439-016-1654-x.
- Chen C.-Y., Chang I.-S., Hsiung C.A., Wasserman W.W. On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC Med. Genomics*. 2014;7:34. DOI 10.1186/1755-8794-7-34.
- Chen J., Rozowsky J., Galeev T.R., Harmanci A., Kitchen R., Bedford J., Abyzov A., Kong Y., Regan L., Gerstein M. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat. Commun.* 2016;18(7):11101. DOI 10.1038/ncomms11101.
- Chen L., Liang Y., Qiu J., Zhang L., Chen X., Luo X., Jiang J. Significance of rs1271572 in the estrogen receptor beta gene promoter and its correlation with breast cancer in a southwestern Chinese population. *J. Biomed. Sci.* 2013;20:32. DOI 10.1186/1423-0127-20-32.
- Claussnitzer M., Dankel S.N., Kim K.-H., Quon G., Meuleman W., Haugen C., Glunk V., Sousa I.S., Beaudry J.L., Puviondrand V., Abdennur N.A., Liu J., Svensson P.-A., Hsu Y.-H., Drucker D.J., Mellgren G., Hui C.-Ch., Hauner H., Kellis M. *FTO* obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 2015; 373:895-907. DOI 10.1056/NEJMoa1502214.
- Cong Z., Li Q., Yang Y., Guo X., Cui L., You T. The SNP of rs6854845 suppresses transcription via the DNA looping structure alteration of super-enhancer in colon cells. *Biochem. Biophys. Res.* 2019;514: 734-741. DOI 10.1016/j.bbrc.2019.04.190.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. DOI 10.1038/nature11247.
- Farh K.K.-H., Marson A., Zhu J., Kleinewietfeld M., Housley W.J., Beik S., Shores N., Whitton H., Ryan R.J.H., Shishkin A.A., Hatan M., Carrasco-Alfonso M.J., Mayer D., Luckey C.J., Patsoopoulos N.A., De Jager P.L., Kuchroo V.K., Epstein C.B., Daly M.J., Hafler D.A., Bernstein B.E. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337-343. DOI 10.1038/nature13835.
- Fishman V.S., Salnikov P.A., Battulin N.R. Interpreting chromosomal rearrangements in the context of 3-dimensional genome organization: a practical guide for medical genetics. *Biochemistry*. 2018; 83(4):393-401. DOI 10.1134/S0006297918040107.
- Gorbacheva A.M., Korneev K.V., Kuprash D.V., Mitkin N.A. The risk G allele of the single-nucleotide polymorphism rs928413 creates a CREB1-binding site that activates *IL33* promoter in lung epithelial cells. *Int. J. Mol. Sci.* 2018;19(10):2911. DOI 10.3390/ijms19102911.
- Guo L., Wang J. rSNPBase 3.0: an updated database of SNP-related regulatory elements, element-gene pairs and SNP-based gene regulatory networks. *Nucleic Acids Res.* 2018;46(D1):D1111-D1116. DOI 10.1093/nar/gkx1101.
- Hansen A.S., Cattoglio C., Darzacq X., Tjian R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*. 2018; 9(1):20-32. DOI 10.1080/19491034.2017.1389365.
- Howard T.D., Mathias R.A., Seeds M.C., Herrington D.M., Hixson J.E., Shimmin L.C., Hawkins G.A., Sellers M., Ainsworth H.C., Sergeant S., Miller L.R., Chilton F.H. DNA methylation in an enhancer region of the *FADS* cluster is associated with *FADS* activity in human liver. *PLoS One*. 2014;9(5):e97510. DOI 10.1371/journal.pone.0097510.
- Ibrahim D.M., Mundlos S. Three-dimensional chromatin in disease: what holds us together and what drives us apart? *Curr. Opin. Cell Biol.* 2020;64:1-9. DOI 10.1016/j.cell.2020.01.003.
- Izzi B., Pistoni M., Cludts K., Akkor P., Lambrechts D., Verfaillie C., Verhamme P., Freson K., Hoylaerts M.F. Allele-specific DNA methylation reinforces *PEAR1* enhancer activity. *Blood*. 2016;128: 1003-1012. DOI 10.1182/blood-2015-11-682153.
- Jones P.L., Veenstra G.J., Wade P.A., Vermaak D., Kass S.U., Landsberger N., Strouboulis J., Wolffe A.P. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* 1998;19:187-191. DOI 10.1038/561.
- Kilpinen H., Waszak S.M., Gschwind A.R., Raghav S.K., Witwicki R.M., Orioli A., Migliaiavacca E., Wiederkehr M., Gutierrez-Arcelus M., Panousis N., Yurovsky A., Lappalainen T., Romano-Palumbo L., Planchon A., Bielser D., Bryois J., Padiouleau I., Udin G., Thurnheer S., Hacker D., Core L.J., Lis J.T., Hernandez N., Raymond A., Deplancke B., Dermitzakis E.T. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science*. 2013;342:744-747. DOI 10.1126/science.1242463.
- Korneev K.V., Sviriaeva E.N., Mitkin N.A., Gorbacheva A.M., Uvarova A.N., Ustiugova A.S., Polanovsky O.L., Kulakovskiy I.V., Afanasyeva M.A., Schwartz A.M., Kuprash D.V. Minor C allele of the SNP rs7873784 associated with rheumatoid arthritis and type-2 diabetes mellitus binds PU.1 and enhances TLR4 expression. *Biochim. Biophys. Acta Mol. Basis Dis.* 2020;1866(3):165626. DOI 10.1016/j.bbdis.2019.165626.
- Kulakovskiy I.V., Vorontsov I.E., Yevshin I.S., Sharipov R.N., Fedorova A.D., Rumynskiy E.I., Medvedeva Y.A., Magana-Mora A., Bajic V.B., Papatsenko D.A., Kolpakov F.A., Makeev V.J. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* 2018;46(D1):D252-D259. DOI 10.1093/nar/gkx1106.
- Kumar D., Puan K.J., Andiappan A.K., Lee B., Westerlaken G.H., Haase D., Melchioti R., Li Z., Yusof N., Lum J., Koh G., Foo S., Yeong J., Alves A.C., Pekkanen J., Sun L.D., Irwanto A., Fairfax B.P., Naranbhai V., Common J.E., Tang M., Chuang C.K., Jarvelin M.R., Knight J.C., Zhang X., Chew F.T., Prabhakar S., Jianjun L., Wang Y., Zolezzi F., Poidinger M., Lane E.B., Meyaard L., Röttschke O. A functional SNP associated with atopic dermatitis controls cell type-specific methylation of the *VSTM1* gene locus. *Genome Med.* 2017;9(1):18. DOI 10.1186/s13073-017-0404-6.
- Kumar S., Ambrosini G., Bucher P. SNP2TFBS – a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* 2017;45(D1):D139-D144. DOI 10.1093/nar/gkw1064.
- Lee C.M., Barber G.P., Casper J., Clawson H., Diekhans M., Gonzalez J.N., Hinrichs A.S., Lee B.T., Nassar L.R., Powell C.C., Ranev B.J., Rosenbloom K.R., Schmelter D., Speir M.L., Zweig A.S., Haussler D., Haeussler M., Kuhn R.M., Kent W.J. UCSC Genome Browser enters 20th year. *Nucleic Acids Res.* 2020;48(D1):D756-D761. DOI 10.1093/nar/gkz1012.
- Levitsky V.G., Kulakovskiy I.V., Ershov N.I., Oshchepkov D.Y., Makeev V.J., Hodgman T.C., Merkulova T.I. Application of experimentally verified transcription factor binding sites models for computational analysis of ChIP-Seq data. *BMC Genom.* 2014;15(1):80. DOI 10.1186/1471-2164-15-80.
- Lewinsky R.H., Jensen T.G.K., Møller J., Stensballe A., Olsen J., Troelsen J.T. T₋₁₃₉₁₀ DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*. *Hum. Mol. Genet.* 2005;14(24):3945-3953. DOI 10.1093/hmg/ddi418.
- Li S., Li Y., Li X., Liu J., Huo Y., Wang J., Liu Z., Li M., Luo X.-J. Regulatory mechanisms of major depressive disorder risk variants. *Mol. Psychiatry*. 2020;25(9):1926-1945. DOI 10.1038/s41380-020-0715-7.
- Lupiáñez D.G., Kraft K., Heinrich V., Krawitz P., Brancati F., Kloppock E., Horn D., Kayserli H., Opitz J.M., Laxova R., Santos-Simarro F., Gilbert-Dussardier B., Wittler L., Borschiwer M., Haas S.A., Osterwalder M., Franke M., Timmermann B., Hecht J., Spielmann M., Visel A., Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015;161(5):1012-1025. DOI 10.1016/j.cell.2015.04.004.

- Mathelier A., Shi W., Wasserman W.W. Identification of altered cis-regulatory elements in human disease. *Trends Genet.* 2015;31(2): 67-76. DOI 10.1016/j.tig.2014.12.003.
- Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E., Wang H., Reynolds A.P., Sandstrom R., Qu H., Brody J., Shafer A., Neri F., Lee K., Kutayin T., Stehling-Sun S., Johnson A.K., Canfield T.K., Giste E., Diegel M., Bates D., Hansen R.S., Neph S., Sabo P.J., Heimfeld S., Raubitschek A., Ziegler S., Cotsapas C., Sotoodehnia N., Glass I., Sunyaev S.R., Kaul R., Stamatoyannopoulos J.A. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190-1195. DOI 10.1126/science.1222794.
- McVicker G., van de Geijn B., Degner J.F., Cain C.E., Banovich N.E., Raj A., Lewellen N., Myrthil M., Gilad Y., Pritchard J.K. Identification of genetic variants that affect histone modifications in human cells. *Science.* 2013;342:747-749. DOI 10.1126/science.1242429.
- Meddens C., van der List A.C.J., Nieuwenhuis E.E.S., Mokry M. Non-coding DNA in IBD: from sequence variation in DNA regulatory elements to novel therapeutic potential. *Gut.* 2019;68(5):928-941. DOI 10.1136/gutjnl-2018-317516.
- Mei S., Ke J., Tian J., Ying P., Yang N., Wang X., Zou D., Peng X., Yang Y., Zhu Y., Gong Y., Zhong R., Chang J., Miao X. A functional variant in the boundary of a topological association domain is associated with pancreatic cancer risk. *Mol. Carcinog.* 2019;58(10): 1855-1862. DOI 10.1002/mc.23077.
- Merkulov V.M., Leberfarb E.Y., Merkulova T.I. Regulatory SNPs and their widespread effects on the transcriptome. *J. Biosci.* 2018;43(5): 1069-1075. DOI 10.1007/s12038-018-9817-7.
- Nan X., Ng H.H., Johnson C.A., Laherty C.D., Turner B.M., Eisenman R.N., Bird A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature.* 1998;393:386-389. DOI 10.1038/30764.
- Park C.-Y., Halevy T., Lee D.R., Sung J.J., Lee J.S., Yanuka O., Benvenisty N., Kim D.-W. Reversion of *FMR1* methylation and silencing by editing the triplet repeats in fragile X iPSC-derived neurons. *Cell. Rep.* 2015;13(2):234-241. DOI 10.1016/j.celrep.2015.08.084.
- Qunneville S., Verde G., Corsinotti A., Kapopoulou A., Jakobsson J., Offner S., Baglivo I., Pedone P.V., Grimaldi G., Riccio A., Trono D. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell.* 2011;44(3):361-372. DOI 10.1016/j.molcel.2011.08.032.
- Rahbar E., Waits C.M.K., Kirby E.H., Jr., Miller L.R., Ainsworth H.C., Cui T., Sergeant S., Howard T.D., Langefeld C.D., Chilton F.H. Allele-specific methylation in the *FADS* genomic region in DNA from human saliva, CD4+ cells, and total leukocytes. *Clin. Epigenetics.* 2018;10:46. DOI 10.1186/s13148-018-0480-5.
- Reddy T.E., Gertz J., Pauli F., Kucera K.S., Varley K.E., Newberry K.M., Marinov G.K., Mortazavi A., Williams B.A., Song L., Crawford G.E., Wold B., Willard H.F., Myers R.M. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res.* 2012;22(5):860-869. DOI 10.1101/gr.131201.111.
- Roadmap Epigenomics Consortium, Kundaje A., Meuleman W., Ernst J., Bilenky M., Yen A., Heravi-Moussavi A., Kheradpour P., Zhang Z., Wang J., Ziller M.J., ... Hirst M., Meissner A., Milosavljevic A., Ren B., Stamatoyannopoulos J.A., Wang T., Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015; 518(7539):317-330. DOI 10.1038/nature14248.
- Rozowsky J., Abyzov A., Wang J., Alves P., Raha D., Harmanci A., Leng J., Bjornson R., Kong Y., Kitabayashi N., Bhardwaj N., Rubin M., Snyder M., Gerstein M. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* 2011;7:522. DOI 10.1038/msb.2011.54.
- Schmitz R.J., Lewis Z.A., Goll M.G. DNA methylation: shared and divergent features across eukaryotes. *Trends Genet.* 2019;35(11): 818-827. DOI 10.1016/j.tig.2019.07.007.
- Shi W., Fornes O., Mathelier A., Wasserman W.W. Evaluating the impact of single nucleotide variants on transcription factor binding. *Nucleic Acids Res.* 2016;44(21):10106-10116. DOI 10.1093/nar/gkw691.
- Smith A.J.P., Deloukas P., Munroe P.B. Emerging applications of genome-editing technology to examine functionality of GWAS-associated variants for complex traits. *Physiol. Genomics.* 2018;50(7): 510-522. DOI 10.1152/physiolgenomics.00028.2018.
- Sun J.H., Zhou L., Emerson D.J., Phyto S.A., Titus K.R., Gong W., Gilgenast T.G., Beagan J.A., Davidson B.L., Tassone F., Phillips-Cremins J.E. Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell.* 2018;175(1):224-238. DOI 10.1016/j.cell.2018.08.005.
- Visser M., Palstra R.J., Kayser M. Allele-specific transcriptional regulation of *IRF4* in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the *IRF4* promoter. *Hum. Mol. Genet.* 2015;24(9):2649-2661. DOI 10.1093/hmg/ddv029.
- Vohra M., Sharma A.R., Prabhu B.N., Rai P.S. SNPs in sites for DNA methylation, transcription factor binding, and miRNA targets leading to allele-specific gene expression and contributing to complex disease risk: a systematic review. *Public Health Genomics.* 2020;23: 1-16. DOI 10.1159/000510253.
- Wang H., Lou D., Wang Z. Crosstalk of genetic variants, allele-specific DNA methylation, and environmental factors for complex disease risk. *Front. Genet.* 2019;9:695. DOI 10.3389/fgene.2018.00695.
- Ward L.D., Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(Database issue):D930-D934. DOI 10.1093/nar/gkr917.
- Waszak S.M., Kilpinen H., Gschwind A.R., Orioli A., Raghav S.K., Witwicki R.M., Migliavacca E., Yurovsky A., Lappalainen T., Hernandez N., Reymond A., Dermizakis E.T., Deplancke B. Identification and removal of low-complexity sites in allele-specific analysis of ChIP-seq data. *Bioinformatics.* 2014;30(2):165-171. DOI 10.1093/bioinformatics/btt667.
- Wingender E., Schoeps T., Dönitz J. TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Res.* 2013;41(D1):D165-D170. DOI 10.1093/nar/gks1123.
- Yates A.D., Achuthan P., Akanni W., Allen J., Allen J., Alvarez-Jarreta J., Amode M.R., Armean I.M., Azov A.G., Bennett R., Bhai J., ... Perry E., Ruffier M., Trevanion S.J., Cunningham F., Howe K.L., Zerbino D.R., Flicek P. Ensembl 2020. *Nucleic Acids Res.* 2020; 48(D1):D682-D688. DOI 10.1093/nar/gkz966.
- Younesy H., Möller T., Heravi-Moussavi A., Cheng J.B., Costello J.F., Lorincz M.C., Karimi M.M., Jones S.J.M. ALEA: a toolbox for allele-specific epigenomics analysis. *Bioinformatics.* 2014;30(8): 1172-1174. DOI 10.1093/bioinformatics/btt744.
- Zhang Y., Manjunath M., Zhang S., Chasman D., Roy S., Song J.S. Integrative genomic analysis predicts causative cis-regulatory mechanisms of the breast cancer-associated genetic variant rs4415084. *Cancer Res.* 2018;78(7):1579-1591. DOI 10.1158/0008-5472.CAN-17-3486.
- Zhao T., Hu Y., Zang T., Wang Y. Integrate GWAS, eQTL, and mQTL data to identify Alzheimer's disease-related genes. *Front. Genet.* 2019;10:1021. DOI 10.3389/fgene.2019.01021.

ORCID ID

E.V. Ignatieva orcid.org/0000-0002-8588-6511

Благодарности. Исследование поддержано из средств бюджетного проекта 0259-2021-0009.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 28.12.2020. После доработки 18.01.2021. Принята к публикации 18.01.2021.