

ISSN 0016-6758

Том 59, Номер 6

Июнь 2023



ГЕНЕТИКА



www.sciencejournals.ru



СОДЕРЖАНИЕ

Том 59, номер 6, 2023

Обзорные и теоретические статьи

Патогенетика кардиомиопатий

А. Н. Кучер, А. А. Слепцов, М. С. Назаренко 615

Сравнительный анализ линий Y-хромосомы древних и современных саха (якутов)

С. А. Федорова, В. Звенигорски, А. Н. Алексеев 633

Генетика микроорганизмов

Изучение генетического разнообразия возбудителя пирикулярриоза риса
Pyricularia oryzae Cav. по комплексу молекулярных
и морфологических признаков

*Д. В. Нартымов, Е. В. Дубина, М. Г. Рубан, Ю. В. Анискина,
С. В. Гаркуша, И. А. Шилов, Н. С. Велишаева,
О. С. Колобова, Н. К. Истомин* 640

Генетика растений

Динамика распространения ржаных транслокаций в генотипах российских
сортов мягкой пшеницы *Triticum aestivum* L.

А. В. Фисенко, О. А. Ляпунова, Е. В. Зуев, А. Ю. Новосельская-Драгович 648

Генетика человека

Анализ генетических факторов спорадических случаев шизофрении
в семейных трио с использованием метода полногеномного секвенирования

*Т. В. Андреева, Ф. А. Афанасьев, Ф. Е. Гусев,
А. Д. Патрикеев, С. С. Куннжева, Е. И. Рогаев* 659

Повышенный уровень лизосфинголипидов крови у пациентов с шизофренией

*А. Э. Копытова, Т. С. Усенко, А. И. Безрукова, К. С. Башарова, Т. В. Андреева,
Е. В. Волкова, А. Д. Манахов, Г. В. Байдакова, Е. И. Пальчикова,
Е. Ю. Захарова, Н. М. Залуцкая, Н. Г. Незнанов, С. Н. Пчелина* 670

Математические модели и методы

Влияние 3D-организации хроматина на суперэнхансер–промоторные
взаимодействия в эмбриональных стволовых и злокачественных клетках

Ю. А. Эйдельман, С. Г. Андреев 676

Сравнительный анализ инструментов для поиска структурных вариантов
на данных короткоридового полногеномного секвенирования

*А. А. Мкртчян, К. С. Грамматикати, П. Г. Казакова, С. И. Митрофанов,
П. Ю. Земский, А. А. Ивашечкин, М. Н. Пилипенко, Д. В. Светличный,
А. П. Сергеев, Е. А. Снигирь, Л. В. Фролова, Т. А. Шпакова,
В. С. Юдин, А. А. Кескинов, С. М. Юдин, В. И. Скворцова* 687

О генетической дивергенции миграционно-связанных популяций:
современное моделирование по результатам экспериментов
Ю.П. Алтухова с соавторами

О. Л. Жданова, Е. Я. Фрисман 708

Краткие сообщения

- Репарация ДНК не инициирует генетическую нестабильность дрожжевых клеток
Е. С. Евстратова, В. Г. Петин, С. А. Гераськин 718
- Идентификация гибридных особей скальных ящериц *Darevskia armeniaca* × *Darevskia valentini*
на основе микросателлитного генотипирования
А. Е. Гирнык, А. А. Вергун, А. П. Рысков 723
- Генетическая вариабельность локуса гена *MAOA* у агрессивных животных
неканонической поведенческой модели *Neogale vison*
А. Д. Манахов, Н. А. Дудко, Ф. Е. Гусев, Т. В. Андреева, О. В. Трапезов, Е. И. Розаев 728
-
-

Contents

Vol. 59, No. 6, 2023

Reviews and Theoretical Articles

Pathogenetics of Cardiomyopathy

A. N. Kucher, A. A. Sleptcov, and M. S. Nazarenko

615

The Y-Chromosome Lineage Variation in Ancient and Modern Populations of the Sakha (Yakuts)

S. A. Fedorova, V. Zvenigorosky, and A. N. Alekseev

633

Genetics of Microorganisms

Studying the Genetic Diversity of the Rice Break Cause *Pyricularia oryzae* Cav. on the Complex of Molecular and Morphological Features

D. V. Nartymov, E. V. Dubina, M. G. Ruban, Yu. V. Aniskina, S. V. Garkusha, I. A. Shilov, N. S. Velishaeva, O. S. Kolobova, and N. K. Istomin

640

Plant Genetics

Dynamics of Rye Translocations Frequency in Russian Common Wheat *Triticum aestivum* L. Cultivars

A. V. Fisenko, O. A. Lyapunova, E. V. Zuev, and A. Yu. Novoselskaya-Dravovich

648

Human Genetics

Analysis of Genetic Factors of Sporadic Schizophrenia in Family Trios Using Whole Genome Sequencing

T. V. Andreeva, Ph. A. Afanasiev, F. E. Gusev, A. D. Patrikeev, S. S. Kunizheva, and E. I. Rogaev

659

Elevated Level of Blood Lysosphingolipids in Patients with Schizophrenia

A. E. Kopytova, T. S. Usenko, A. I. Bezrukova, K. S. Basharova, T. V. Andreeva, E. V. Volkova, A. D. Manakhov, G. V. Baydakova, E. I. Palchikova, E. Yu. Zakharova, N. M. Zalutskaya, N. G. Neznanov, and S. N. Pchelina

670

Mathematical Models and Methods

Impact of Chromatin 3D-Organization on Promoter–Superenhancer Interactions in Embryonic Stem vs Cancer Cells

Yu. A. Eidelman and S. G. Andreev

676

Comparative Analysis of Structural Variant Callers on the Short-Read Whole-Genome Sequencing Data

A. A. Mkrchian, K. S. Grammatikati, P. G. Kazakova, S. I. Mitrofanov, P. U. Zemsky, A. A. Ivashechkin, M. N. Pilipenko, D. V. Svetlichny, A. P. Sergeev, E. A. Snigir, L. V. Frolova, T. A. Shpakova, V. S. Yudin, A. A. Keskinov, S. M. Yudin, and V. I. Skvortsova

687

About the Genetic Divergence of Populations Coupled by Migrations: Modern Modeling Based on Experimental Results of Yu.P. Altukhov and Coauthors

O. L. Zhdanova and E. Ya. Frisman

708

Short Communications

DNA Repair Does Not Trigger Genetic Instability of Yeast Cells

E. S. Evstratova, V. G. Petin, and S. A. Geraskin

718

Identification of Hybrid Specimens of Rock Lizards *Darevskia armeniaca* × *Darevskia valentini*
Based on Microsatellite Genotyping

A. E. Girnyk, A. A. Vergun, and A. P. Ryskov

723

Genetic Variability of *MAOA* Gene among Aggressive Animals
from the Non-Canonical Behavioral Model *Neogale vison*

*A. D. Manakhov, N. A. Dudko, F. E. Gusev,
T. V. Andreeva, O. V. Trapezov, and E. I. Rogaev*

728

ПАТОГЕНЕТИКА КАРДИОМИОПАТИЙ¹

© 2023 г. А. Н. Кучер¹, А. А. Слепцов¹, М. С. Назаренко¹, *

¹Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, 634050 Россия

*e-mail: maria.nazarenko@medgenetics.ru

Поступила в редакцию 22.08.2022 г.

После доработки 28.10.2022 г.

Принята к публикации 22.11.2022 г.

Анализируется значимость генетических факторов в развитии как первичных (или менделевских) кардиомиопатий (КМП), так и некоторых вторичных (приобретенных) форм КМП. Для первичных КМП описаны десятки генов с патогенными/вероятно патогенными вариантами. В большинстве случаев спектр причинных генетических вариантов для разных КМП специфичен, но регистрируются также общие гены и варианты. При этом, с одной стороны, не для всех случаев первичных КМП установлены генетические причины заболеваний, а с другой, патогенные варианты в генах менделевских КМП регистрируются и при вторичных КМП. Генетический компонент в развитии и первичных и вторичных КМП установлен также при проведении широкогеномных ассоциативных исследований (GWAS). Однонуклеотидные варианты (SNPs), ассоциированные с первичными, и со вторичными КМП, в большинстве случаев специфичны для разных КМП и вносят небольшой вклад в риск развития патологий. Для некоторых SNPs установлены ассоциации с ЭКГ- и Эхо-кардиографическими параметрами морфологически неизмененного сердца у человека. Большинство из SNPs, ассоциированных с КМП по данным GWAS, локализованы в некодирующих участках генома, но обладают регуляторным потенциалом, выступая в ткани сердца в качестве локусов, влияющих на уровень экспрессии (eQTL), сплайсинг (sQTL) или эпигенетические модификации. Примечательно, что эффекты генотипов eQTL и sQTL в ряде случаев неравнозначны для различных анатомических отделов сердца. В целом фенотип и клиническая картина КМП могут определяться широким спектром редких патогенных/вероятно патогенных вариантов с сильным эффектом и частых высокополиморфных вариантов с небольшим эффектом и модифицироваться эпигенетическими факторами.

Ключевые слова: кардиомиопатия, патогенетика, гены моногенных кардиомиопатий, широкогеномные ассоциативные исследования.

DOI: 10.31857/S0016675823050107, **EDN:** DRVHTN

Кардиомиопатии (КМП) — клинически гетерогенная группа заболеваний сердечной мышцы, характеризующаяся сложной этиологией. Используют различные подходы к классификации КМП, которые периодически пересматриваются. В связи с этим зачастую в публикациях приводятся обозначения типов КМП, не принятые в настоящее время. Однако в дальнейшем при изложении материала мы использовали те обозначения КМП, которые были указаны в первоисточниках.

Наиболее часто выделяют семейные и спорадические формы; первичные (генетически обусловленные), приобретенные, или вторичные (возникшие при воздействии широкого спектра эндогенных и экзогенных факторов) и смешанные (при комплексном воздействии генетических и средо-

вых факторов). Генетически детерминированные кардиомиопатии могут быть унаследованными или развиваться вследствие возникших *de novo* мутаций.

С учетом морфофункциональных нарушений сердца КМП классифицируют на гипертрофическую (ГКМП), дилатационную (ДКМП), аритмогенную (АРКМП), рестриктивную (РКМП) кардиомиопатии, к данному патологическому фенотипу относят также некомпактный миокард левого желудочка (НКМЛЖ), но иногда такое патологическое состояние рассматривают в качестве неклассифицированных КМП или варианта фенотипа [1, 2].

Первичные КМП различаются по частоте регистрации и для некоторых ее форм характерно широкое распространение: ДКМП встречается с частотой 1 : 250; ГКМП — с частотой 1 : 200–1 : 500; правожелудочковая АРКМП — с частотой 1 : 2000–1 : 5000; РКМП — наиболее редко встречающаяся

¹ Дополнительная информация для этой статьи доступна по doi 10.31857/S0016675823050107 для авторизованных пользователей.

КМП, на ее долю приходится 2–5% всех случаев КМП [3–6]. К категории вторичных (приобретенных) относят КМП, как правило, возникающие вследствие осложнения какого-либо заболевания (например, ишемической болезни сердца, сахарного диабета, инфекционных заболеваний и др.).

Классификации КМП в некотором смысле условны. С одной стороны, манифестация генетически обусловленных КМП может произойти при воздействии внешних средовых факторов или при физиологических изменениях в организме [7–10], а с другой, патогенные варианты в генах кардиомиопатий выявляются и при немэнделевских КМП и других патологиях сердца (например, при фибрилляции предсердий, миокардите и др.) [11–14].

Кроме того, наряду с первоначально выделенной правожелудочковой АРКМП, при которой происходит гибель кардиомиоцитов и их замещение фиброзно-жировыми рубцами в правом желудочке (что предрасполагает к аритмиям и внезапной смерти), в настоящее время выделяют формы с преобладающим поражением левого, а также обоих желудочков, причем спектр генов, мутации в которых приводят к развитию рубцовых изменений различной локализации, не идентичен [15, 16].

Иногда одна форма заболевания может “переходить” в другую (например, ГКМП – в ДКМП [17]) или различные клинические формы регистрируются у одного и того же пациента одновременно (например, сочетание ГКМП и ДКМП, аритмогенной дисплазии правого желудочка и НКМЛЖ) [18, 19]. Среди детей с фенотипом НКМ на долю изолированного НКМ приходится лишь 23%, а в большинстве случаев регистрируют смешанный фенотип: сочетание с ДКМП выявлено у 59% пациентов, с ГКМП – у 11% пациентов [20]. У детей ГКМП может сочетаться с РКМП или ДКМП (до 20% случаев) [21, 22]. Фенотип аритмогенной кардиомиопатии также может встречаться при других генетически обусловленных кардиомиопатиях [15].

Для КМП, в том числе и первичных и даже при наличии у индивидов одного и того же патологического варианта, характерны широкий возрастной диапазон манифестации патологии (от внутриутробного периода до пожилого возраста), разнообразные клинические проявления (от бессимптомного носительства патогенных вариантов до внезапной сердечной смерти в качестве первого признака патологии), тяжесть течения и клинические исходы [23–28]. С учетом серьезных жизнеугрожающих осложнений при различных КМП (существенное снижение качества жизни вследствие развития сердечной недостаточности, часто требующей трансплантации сердца; высокий риск внезапной сердечной смерти) важным представля-

ется выявление причин развития различных КМП и факторов, определяющих конкретный патологический фенотип и клиническую картину болезни. Генетические маркеры могут выступать в качестве и причины развития КМП, и модифицирующих факторов, а их выявление на ранних этапах позволит определить риски развития КМП и своевременно разработать индивидуализированные профилактические мероприятия.

Цель настоящего исследования заключалась в сравнительном анализе генетических факторов, лежащих в основе развития различных КМП на основе данных о генах, для которых известны патогенные варианты, и результатов широкогеномных ассоциативных исследований (GWAS).

МЕНДЕЛЕВСКИЕ ФОРМЫ КАРДИОМИОПАТИЙ

Наиболее хорошо с генетической точки зрения изучены и охарактеризованы моногенные/олигогенные формы КМП (табл. 1, составлено по [29]). В базе Simple ClinVar (дата входа 20.07.2022 г.) указан 41 ген, в которых локализованы патогенные и вероятно патогенные варианты, приводящие к развитию ГКМП, для ДКМП таких генов указано 50, для правожелудочковой АРКМП – 11, для рестриктивной – семь; в общей сложности с данными КМП ассоциированы патогенные/вероятно патогенные варианты 76 генов. При этом есть как специфичные для каждой из указанных КМП, так и общие гены для разных типов КМП. В частности, варианты в генах *ACTN2* и *TTN* могут приводить к четырем формам КМП – ДКМП, ГКМП, правожелудочковой АРКМП и РКМП; общими для ГКМП, ДКМП и правожелудочковой АРКМП являются гены *DES*, *LMNA* и *TMEM43*, для ГКМП, ДКМП и РКМП – гены *MYH7*, *TNNI3* и *TNNT2*, для ГКМП и ДКМП – гены *ACTC1*, *CSRP3*, *LAMP2*, *MT-TI*, *MYBPC3*, *MYH6*, *NEXN*, *PLN*, *RAF1*, *TCAP*, *TNNC1*, *TRPM1* и *VCL*, для ДКМП и правожелудочковой АРКМП – гены *DSG2* и *DSP*, для ДКМП и РКМП – гены *CRYAB* и *FLNC* [29, 30].

Большинство описанных на настоящий момент патогенных/вероятно патогенных вариантов являются миссенс-мутациями [29] и лишь начинаются исследования по оценке значимости в патогенезе КМП вариантов, локализованных в некодирующих участках генов и межгенных регионах [31]. Помимо патогенных/вероятно патогенных известны варианты с неопределенной значимостью, постоянно открываются новые гены КМП [32–35]. Так, в базе Simple ClinVar [29] варианты с неопределенной значимостью для ДКМП зарегистрированы в 110 генах, для ГКМП – в 110 генах, для правожелудочковой АРКМП – в 35 генах, для РКМП – в 22 генах (табл. 1); среди них есть как гены, несущие патогенные/вероятно

Таблица 1. Число генов с патогенными/вероятно патогенными вариантами и вариантами с неопределенной значимостью для различных первичных кардиомиопатий (КМП, по SimpleClinVar [29])

КМП	Число генов с вариантами		Общее число генов
	патогенными/вероятно патогенными	с неопределенной значимостью	
ГКМП	41	110	119
ДКМП	50	110	116
Правожелудочковая АРКМП	11	35	37
РКМП	7	22	24

патогенные варианты, описанные ранее для тех же или других КМП, так и новые. Если учесть гены, в которых локализованы патогенные и вероятно патогенные варианты, а также варианты с неопределенной значимостью в отношении КМП, то общее число генов данного заболевания возрастает более чем в 2 раза – до 159. При этом патогенетическая значимость редких генетических вариантов, выявленных у пациентов с КМП, периодически пересматривается с учетом новых данных о частоте их выявления среди пациентов и в популяциях, сегрегации в семьях и уточнения функциональных изменений при их наличии [36].

Для разных типов генетически обусловленных КМП установлены общие не только гены, но и патогенные варианты: в Simple ClinVar таких однонуклеотидных вариантов (SNVs) и вариаций числа копий участков ДНК (CNVs) в различных участках генов (экзоны, интроны) зарегистрировано более 200 [19, 29, 30]. Так, миссенс-вариант с.2513C>T (p.Pro838Leu) в гене *MYH7* может обуславливать формирование ГКМП, ДКМП и РКМП, а миссенс-замена с.451C>T (p.Arg151Trp) в гене *TNNT2* может приводить к развитию первичной ДКМП, семейной РКМП, семейной ГКМП и т.д.

Иногда разные формы КМП выявляются у членов одной семьи, несущих один и тот же генетический вариант в генах кардиомиопатий. S.C. Menon с соавт. [25] описана семья с аутосомно-доминантным заболеванием сердца, вызванным миссенс-вариантом, приводящим к замене p.Ile79Asn в сердечном тропонине T (кодирует ген *TNNT2*): все пораженные члены являлись гетерозиготными носителями данного варианта, но у двух из них диагностирована РКМП, у трех – необструктивная ГКМП, у двух – ДКМП, у одного – смешанная КМП (ГКМП, сочетающаяся с фенотипическими признаками ДКМП и рестриктивной физиологией), а еще у одного – легкая концентрическая гипертрофия левого желудочка (ЛЖ).

Общность генетической компоненты между разными формами КМП может определять сходство их фенотипических проявлений. При ГКМП и РКМП, которые обусловлены патогенными ва-

риантами генов саркомерных белков, в качестве преобладающего патофизиологического механизма выступает диастолическая дисфункция, а иногда выделяют гипертрофическую КМП с рестриктивной патофизиологией [37–39].

Несмотря на некоторые перекрытия по генам, разные клинические формы КМП различаются по спектру генов, патогенные варианты в которых объясняют большую долю генетических причинных факторов [29]. Для ДКМП 15–20% патогенетически значимых вариантов описаны в гене *TTN*, с частотой 3–6% регистрируются такие варианты в генах *LMNA*, *MYH7*, *FLNC*, *MYH6*, *BAG3*, *TNNT2* и *MYPN3* [40]. Наибольший вклад в структуру генетической компоненты ГКМП вносят патогенные варианты в генах саркомерных белков, причем на варианты в генах *MYH7* и *MYBPC3* приходится около 70%, а варианты в генах *TNNI3*, *TNNT2*, *TPM1*, *MYL2*, *MYL3* и *ACTC1* регистрируются с частотой 1–5%, без выраженных межэтнических различий [4]. К правожелудочковой АРКМП чаще приводят патогенные варианты в генах десмосом, интеркалированных дисков и ионных каналов: *PKP2* (до 50% выявляемых патогенных мутаций), *DSP*, *DSC2*, *DSG2*, *JUP*, *PLN* и *SCN5A* [41, 42].

Однако те генетические варианты, которые рассматриваются в качестве причинных для КМП, регистрируют в популяциях редко и в целом генетические причины удается установить не у всех пациентов с данной патологией. Частота выявления патогенных вариантов в генах при ДКМП составляет 15–40% [43–45], при ГКМП – 20–43% [46, 47], при правожелудочковой АРКМП – 50–63% [5, 41, 42]. Иными словами, до 50% и более пациентов с КМП остаются без молекулярно-генетического подтверждения диагноза. Следует, однако, отметить, что у пациентов с КМП выявляются еще и варианты с конфликтной и неопределенной значимостью, для которых нет однозначного мнения об их вовлеченности в развитие КМП.

Приведенные выше данные указывают на важность поиска как новых генов и новых вариантов, приводящих к развитию первичных КМП, так и генетических факторов, определяющих клиниче-

ский полиморфизм различных форм данной патологии. Одним из подходов к выявлению таких генетических факторов являются широкогеномные ассоциативные исследования (GWAS).

ХАРАКТЕРИСТИКА ГЕНЕТИЧЕСКИХ ВАРИАНТОВ, АССОЦИИРОВАННЫХ С КАРДИОМИОПАТИЯМИ ПО ДАННЫМ ШИРОКОГЕНОМНЫХ АССОЦИАТИВНЫХ ИССЛЕДОВАНИЙ

К настоящему времени GWAS проведены для ГКМП (в том числе для саркомер-позитивных и саркомер-негативных форм) [48] и ДКМП (включая идиопатическую) [49–53]; КМП при болезни Шагаса [54, 55]; КМП Такоцубо (стресс-индуцированная КМП) [56]; ишемической КМП [57]; КМП, связанной с приемом антрацилина у детей с онкозаболеваниями [58]; перипартальной КМП [59] и некоторых других [60]. Таким образом, GWAS выполнены не только для первичных КМП, но и КМП, причиной которых явились экзогенные и эндогенные факторы.

Исследования проводились для различных этно-территориальных групп населения: европеоидов (ГКМП [61], ДКМП [49, 53, 61], перипартальная КМП [59], КМП Такоцубо [56]), латиноамериканцев (КМП Шагаса [54, 55]), афроамериканцев (ДКМП [51]), испанцев (ишемическая КМП [57]), смешанных групп (ГКМП – включали африканцев, латиноамериканцев, европеоидов, азиатов и др. [48]; антрацилин-индуцированная КМП – неиспаноязычные европеоиды, смешанная этническая группа [58]). В то же время статистически значимые ассоциации (по критериям GWAS) были получены не во всех исследованиях.

В GWAS Catalog [60] приведены данные об ассоциациях 49 SNPs с ГКМП, 20 – с саркомер-позитивной ГКМП, 25 – с саркомер-негативной ГКМП, 22 – с ДКМП, 44 – с идиопатической ДКМП, 21 – с КМП при болезни Шагаса, пять – с КМП Такоцубо и три – с ишемической КМП у пациентов с ревматоидным артритом (Приложение, табл. 1, составлено по [60, 62, 63]).

В общей сложности с КМП установлены ассоциации 156 SNPs (табл. 2, составлена по [60, 62, 63]; Приложение, табл. 1), из которых 51 (33%) вариант локализован в межгенных регионах, 105 (67%) – в 102 генах (в том числе 10 SNPs, каждый из которых локализован в двух перекрывающихся генах), включая 76 белок-кодирующих генов, 26 генов некодирующих РНК (пять из которых – гены антисмысловых транскриптов: *CYP51A1-AS1*, *LY86-AS1*, *TRDN-AS1*, *TTN-AS1* и *SVIL-AS1*). Для ряда генов ассоциированными с КМП оказались от двух до пяти локализованных в них SNPs: по два варианта расположены в генах *FNDC3B*,

MITF, *PKD1*, *PRKCA*, *SMARCB1*, *SSPN*, *VTG1A*, три SNPs – в гене *BAG3*, пять SNPs – в гене *FHOD3*.

Как и в случае генов КМП, в результате GWAS выявлены специфичные (большинство) и общие гены и ассоциированные варианты (их число невелико) для разных типов патологии (табл. 2, 3; Приложение, табл. 1). Только уникальные ассоциированные гены были зарегистрированы для ишемической КМП и КМП Такоцубо, но в целом GWAS для данных КМП были немногочисленными и малоинформативными – выявлено небольшое число ассоциированных вариантов и генов (табл. 2).

Для КМП Шагаса лишь один ген из 19 ассоциированных был неуникальным: в гене *MITF* разные интронные SNPs были связаны с КМП Шагаса и саркомер-негативной ГКМП.

Общими для ГКМП и ДКМП были гены: *BAG3* (интронные варианты rs17099139 и rs72840788 ассоциированы с ГКМП, экзонный вариант rs2234962 – с ДКМП), *FHOD3* (четыре интронных SNPs – rs503274, rs118060942, rs4799426, rs617207 – ассоциированы с ГКМП, экзонный вариант rs2303510 – с ДКМП), *PKD1* (разные интронные SNPs ассоциированы с ДКМП и ГКМП – rs2519236 и rs9928278 соответственно), *SMARCB1* (разные аллели интронного варианта rs2186370 ассоциированы с ДКМП и ГКМП, интронный вариант rs7284877 – с ДКМП), *PRKCA* (один общий интронный SNP – rs9892651, но разные аллели ассоциированы с ДКМП и ГКМП; rs7210446 ассоциирован только с ГКМП) (табл. 2 и 3). Таким образом, несмотря на общность генов, лишь два SNPs были ассоциированы по данным GWAS и с ГКМП, и с ДКМП. Кроме того, с ГКМП ассоциированы по два разных интронных SNPs, локализованных в генах *FNDC3B* (rs9647379 и rs4894803), *SSPN* (rs17380837 и rs2003585), *VTG1A* (rs11196078 и rs11196085). Интересно, что спектр генов, варианты в которых приводили к развитию саркомер-позитивной и саркомер-негативной КМП, не идентичен (табл. 2).

Наблюдается также перекрывание между разными типами КМП, для которых установлены как редкие патогенные, так и частые (ассоциированные по данным GWAS) варианты в одних и тех же генах (табл. 3). Так, для генов *BAG3*, *FLNC*, *FHOD3*, *LMNA*, *TTN* и *TTN-AS1* описаны как патогенные/вероятно патогенные варианты для КМП, так и SNPs, ассоциированные с КМП по данным GWAS. Для ишемической КМП также зарегистрирована ассоциация с rs112941217-С, локализованным в гене *LMNA*, для которого установлены патогенные/вероятно патогенные варианты, приводящие к развитию ГКМП, ДКМП и правожелудочковой АРКМП.

Среди ассоциированных с КМП вариантов лишь 12 (8%) локализованы в экзонах белок-ко-

Таблица 2. Характеристика SNPs и их генов, ассоциированных с кардиомиопатиями (КМП) по данным GWAS [60]

КМП	Всего SNP	Межгенные SNP	Число генов ^{&}	Гены [§] , в которых локализованы ассоциированные SNPs (по [63])
Кардиомиопатия Шагаса при серопозитивности <i>Trypanosoma cruzi</i>	21	5	17 [#]	<i>ENSG00000226070</i> , <i>ENSG00000287303/PDGFRL</i> , <i>ILRUN</i> , <i>KAZN</i> , <i>LINC02895</i> , <i>LINC02100</i> , <i>MITF</i> , <i>NAALADL1</i> , <i>NRG1</i> , <i>PLA2G4F</i> , <i>POLRIC</i> , <i>SLCO1B1</i> , <i>SPTB</i> , <i>SYT9</i> , <i>TIAM1</i> , <i>ZNF208</i>
ДКМП	9	1	8	<i>BAG3</i> , <i>FHOD3</i> , <i>FLNC</i> , <i>NMB</i> , <i>PKD1</i> , <i>SMARCB1</i> , <i>SVIL-AS1</i> , <i>ZBTB17</i>
ДКМП (MTAG)	13	3	11 [#]	<i>BAG3</i> , <i>ENSG00000255491 (LOC105375743)</i> , <i>FHOD3</i> , <i>FLNC</i> , <i>NMB</i> , <i>PRKCA</i> , <i>SMARCB1</i> , <i>SVIL-AS1</i> , <i>TTN-AS1/TTN</i> , <i>ZBTB17</i>
Идиопатическая ДКМП	44	18	27 [#]	<i>ARHGEF3</i> , <i>BAG3</i> , <i>BOLL</i> , <i>CACNB4</i> , <i>CCDC149</i> , <i>DNAH17</i> , <i>ENSG00000241956 (LOC102546299)</i> , <i>ENSG00000249001</i> , <i>ENSG00000253535</i> , <i>ENSG00000287199</i> , <i>FAM120B</i> , <i>FGF12</i> , <i>IL19</i> , <i>JPH3</i> , <i>LINC00343</i> , <i>LINC02509</i> , <i>LINC02882</i> , <i>LRRDI/CYP51A1-AS1</i> , <i>MGAM</i> , <i>PHACTR1</i> , <i>RALY</i> , <i>SASH1</i> , <i>VMPI</i> , <i>WASHC2C</i> , <i>WWOX</i> , <i>ZBTB17</i>
ГКМП	33	12	14	<i>AHNAK</i> , <i>ARHGAP1</i> , <i>BAG3*</i> , <i>CDKN1A</i> , <i>ENSG00000287709 (LOC101928505)</i> , <i>FHOD3*</i> , <i>FNDC3B*</i> , <i>HSPB7</i> , <i>NSF</i> , <i>PRG3</i> , <i>PRKCA*</i> , <i>SMARCB1</i> , <i>SYNPO2L</i> , <i>VTT1A</i>
ГКМП (MTAG)	16	6	10	<i>BAG3</i> , <i>ENSG00000287709 (LOC101928505)</i> , <i>FHOD3</i> , <i>FNDC3B</i> , <i>NSF</i> , <i>PKD1</i> , <i>PRKCA</i> , <i>SMARCB1</i> , <i>SYNPO2L</i> , <i>VTT1A</i>
ГКМП (саркомер-позитивная)	20	5	18 [#]	<i>ENSG00000254746/ENSG00000255041</i> , <i>ENSG00000285939</i> , <i>ABCA5</i> , <i>AMBR1</i> , <i>CLCNKA</i> , <i>GTF2A1L/STON1-GTF2A1L</i> , <i>KCP</i> , <i>MTUS2</i> , <i>MUC6</i> , <i>NCKAP5</i> , <i>OR9Q1</i> , <i>P2RX3</i> , <i>PHF21A</i> , <i>TNKS1BP1</i> , <i>TXNRD1/ENSG00000257732</i>
ГКМП (саркомер-негативная)	25	6	20 [#]	<i>ADPRHL1</i> , <i>BAG3</i> , <i>CCDC112</i> , <i>CCDC141/ENSG00000287149</i> , <i>CDKN1A</i> , <i>CHRN4</i> , <i>ENSG00000259768</i> , <i>FHOD3*</i> , <i>FNDC3B</i> , <i>HSPB7</i> , <i>MITF</i> , <i>PRKCA</i> , <i>RNF207</i> , <i>SSPN*</i> , <i>SYMPK</i> , <i>TRDN/TRDN-AS1</i> , <i>ZSCAN2/ENSG00000275120</i>
Ишемическая КМП	3	0	3	<i>KIF26B</i> , <i>LMNA</i> , <i>WNT16</i>
КМП Такоцубо	5	2	3	<i>LINC02625</i> , <i>LY86-AS1</i> , <i>PIWIL2</i>
Всего:	156	51	101	

Примечание. [§] В круглых скобках приведены обозначения для новых генов по NCBI [64]. [&] Учтены гены, в которых локализованы SNPs, ассоциированные с КМП, по данным UCSC [62]. * В генах локализованы несколько ассоциированных с КМП SNPs. [#] С учетом перекрывающихся генов. Жирным шрифтом выделены гены, в которых локализованы варианты, ассоциированные с разными формами КМП. MTAG (multi-trait analysis of GWAS) – ассоциации установлены при проведении мультипризнакового анализа сводной статистики GWAS.

Таблица 3. Гены, для которых установлены ассоциации SNPs с разными типами кардиомиопатий (КМП) по данным GWAS, и патогенные/вероятно патогенные варианты для менделевских КМП

Ген	Данные GWAS [60]				Менделевские КМП (по [29])	
	SNP	частота аллеля*	локализация в гене	КМП {аллель}	КМП	число вариантов P LP
<i>PKD1</i>	rs2519236	0.82	Интрон	ДКМП {T}	–	–
	rs9928278	0.183	Интрон	ГКМП (MTAG) {C}	–	–
<i>SMARCB1</i>	rs2186370	0.809/0.219	Интрон	ДКМП (MTAG) {G}/ГКМП {A}, ГКМП (MTAG) {A}	–	–
	rs7284877	0.81	Интрон	ДКМП {C}	–	–
<i>PRKCA</i>	rs9892651	0.417/0.586	Интрон	ДКМП (MTAG) {C}/ГКМП {T}, ГКМП (MTAG) {T}	–	–
	rs7210446	0.419	Интрон	ГКМП {G}, ГКМП (саркомер-негативная) {G}	–	–
<i>MTF</i>	rs78356356	–	Интрон	КМП Шагаса {?}	–	–
	rs62253180	0.817	Интрон	ГКМП (саркомер-негативная) {A}	–	–
<i>BAG3</i>	rs2234962	0.779	Экзон / p.Cys151Arg	ДКМП {T}, ДКМП (MTAG) {T}, идиопатическая ДКМП {?}	ДКМП	34 22
	rs17099139	0.291	Интрон	ГКМП {G}, ГКМП (MTAG) {G}	–	–
<i>FHOD3</i>	rs72840788	0.792	Интрон	ГКМП {G}, ГКМП (саркомер-негативная) {G}	–	–
	rs2303510	0.679	Экзон / p.Val1134Ile	ДКМП {G}, ДКМП (MTAG) {G}	ГКМП	2 0
<i>FLNC</i>	rs503274	0.311	Интрон	ГКМП {C}, ГКМП (MTAG) {C}	–	–
	rs118060942	0.988	Интрон	ГКМП {A}, ГКМП (саркомер-негативная) {A}	–	–
<i>LMNA</i>	rs4799426	0.655	Интрон	ГКМП {G}	–	–
	rs617207	0.291	Интрон	ГКМП (саркомер-негативная) {A}	–	–
<i>FLNC</i>	rs2291569-G	0.92	Экзон / p.Arg1567Gln, p.Arg1567Leu, p.Arg1567Pro	ДКМП {G}, ДКМП (MTAG) {G}	ДКМП	112 35
	rs112941217-C	0.03	Интрон	Ишемическая КМП	ГКМП	116 37
<i>TTN</i>	rs2042995-T	0.78	Экзон/p.Ile10205Val	ДКМП (MTAG) {T}	ДКМП	32 45
	rs2042995-T	0.78	Интрон	ДКМП (MTAG) {T}	ГКМП	1 1
<i>TTN-AS1</i>	rs2042995-T	0.78	Интрон	ДКМП (MTAG) {T}	АРКМП	0 3
	rs2042995-T	0.78	Интрон	ДКМП (MTAG) {T}	ДКМП	63 992
<i>TTN-AS1</i>	rs2042995-T	0.78	Интрон	ДКМП (MTAG) {T}	ГКМП	2 5
	rs2042995-T	0.78	Интрон	ДКМП (MTAG) {T}	АРКМП	0 1
<i>TTN-AS1</i>	rs2042995-T	0.78	Интрон	ДКМП (MTAG) {T}	ДКМП	2 2

Примечание. В фигурных скобках – ассоциированный аллель. * Через / приведены частоты ассоциированных аллелей в случае разных оценок для разных КМП. АРКМП – правожелудочковая аритмогенная кардиомиопатия. P – варианты с патогенной значимостью; LP – варианты с вероятной патогенной значимостью.

дирующих генов (в том числе 11 SNPs приводят к несинонимичной, rs8051448 в гене *JPH3* – к синонимичной замене), три локализованы в 3'UTR, три – в экзонах некодирующих транскриптов, и 98 – в интронах белок-кодирующих и белок-некодирующих генов (Приложение, табл. 1). Таким образом, варианты, ассоциированные с различными КМП по данным GWAS, локализованы преимущественно либо в межгенных регионах, либо в некодирующих участках белок-кодирующих и чаще белок-некодирующих генов. В связи с этим представляет интерес оценка функциональной значимости данных вариантов.

РЕГУЛЯТОРНЫЙ ПОТЕНЦИАЛ ГЕНЕТИЧЕСКИХ ВАРИАНТОВ, АССОЦИИРОВАННЫХ С КАРДИОМИОПАТИЯМИ

Из числа ассоциированных по данным GWAS с КМП вариантов 37 SNPs обладают регуляторным потенциалом в различных отделах сердца, являясь eQTL (Expression Quantitative Trait Loci – локусы количественных признаков, определяющих уровень экспрессии) и/или sQTL (Splicing Quantitative Trait Loci – локусы количественных признаков альтернативного сплайсинга) для широкого спектра генов (34 SNPs являются eQTL и 21 – sQTL) [65, 66] или выступая в качестве меток для различных эпигенетических маркеров, значимых для регуляции экспрессии генов (68 SNPs) [63] (Приложение, табл. 1, 2).

В общей сложности 79 генов в сердце регулируются на уровне экспрессии или сплайсинга с участием SNPs, ассоциированных с КМП (для 72 генов ассоциированные варианты являются eQTL, для 16 – sQTL), причем один SNP может быть значим для регуляции нескольких генов, а один и тот же ген может регулироваться несколькими ассоциированными с КМП SNPs (рис. 1). К числу регулируемых относятся 13 генов, в которых локализованы SNPs, ассоциированные с КМП по данным GWAS (*CDKN1A*, *CLCNKA*, *FHOD3*, *HSPB7*, *NMB*, *NSF*, *PIWIL2* и др.), а также два гена, в которых известны патогенные/вероятно патогенные варианты для КМП (в гене *ALPK3* для ГКМП и в гене *FLNC* – для ДКМП) [29].

Экспрессия *CLCNKA* в сердце зависит от генотипов четырех SNPs: интронных rs2015509 и rs10927875, локализованных в генах *CLCNKA* и *ZBTB17* соответственно, экзонного rs1048302 гена *HSPB7* (эти три SNPs являются также sQTL для *HSPB7*), межгенного rs10927886 (является также eQTL и sQTL для *HSPB7*; см. рис. 1). При этом эффекты генотипов по rs1048302 (ассоциирован с саркомер-негативной ГКМП) и rs2015509 (саркомер-позитивная ГКМП) как sQTL для *HSPB7* различны в ЛЖ и ушке правого предсердия (Приложение, табл. 2). В этих отделах сердца для других

SNPs также зарегистрированы различные эффекты генотипов как sQTL (rs28768976 для гена *MAPT*, rs3740293 для гена *SYNPO2L*) и eQTL (rs13273616 для гена *PIWIL2*, rs6003909 и rs2186370 для гена *MMP11*, rs3779381 для гена *FAM3C*, rs748455 для гена *ALPK3*). Возможно, что именно камерная специфичность в эффектах sQTL и eQTL может объяснять “предпочтительное” поражение определенных отделов при некоторых КМП: при ГКМП – левый желудочек, а при аритмогенной КМП – правый желудочек.

Кроме того, различия регистрируются по спектру генов, редкие варианты в которых приводят к патологии сердца, в том числе и КМП различной локализации (т.е. гены, которые значимы в развитии КМП, зависят от поражаемого отдела сердца). Так, для аритмогенной КМП при поражении правого желудочка выявляются патогенные варианты в генах *PKP2*, *JUP*, *DSC2*, *DSG2*, *DSP* и *SCN5A*, белковые продукты которых контролируют функцию десмосом, интеркалированных дисков и ионных каналов, при поражении ЛЖ – в генах *LMNA*, *DSP*, *FLNC*, *TMEM43*, *BAG3*, *RBM20* и др. (относящихся к цитоскелету, саркоплазматическому ретикулуму, саркомеру, ионному каналу, митохондриям), при поражении обоих желудочков – в гене *PLN*, продукт которого важен для гомеостаза кальция в миокарде [41].

Помимо регуляции генов на уровне экспрессии и сплайсинга, ассоциированные по GWAS с КМП варианты выступают в качестве эпигенетических меток – показано для 68 SNPs (Приложение, табл. 1, 2). В частности, данные варианты локализованы в гиперчувствительных к ДНКазе сайтах (DNase-seq, 53 SNPs), в регионах доступа к транспозазе хроматина (ATAC-seq, 21 SNPs), в области модифицированных гистонов (H3K27me3, H3K27ac, H2AFZ – всего 49 SNPs) и различных эпигенетически значимых регионах для поддержания структуры и функции хроматина. Так, 12 SNPs расположены в области POLR2A – ДНК-зависимой субъединицы РНК-полимеразы II RPB1, девять SNPs – в области транскрипционного репрессора CTCFL, 12 SNPs – EP300-взаимодействующего ингибитора дифференцировки 1, шесть SNPs – гомолога белка репарации двухцепочечных разрывов rad21, два SNPs – центрального компонента когезина (SMC3), значимого для прохождения клеточного цикла. Например, генотипы rs3740293 (ассоциирован с ГКМП), локализованного в 3'UTR гена *SYNPO2L*, влияют на уровень экспрессии ряда генов, в том числе – гистонового шаперона *DNAJC9* (рис. 1; Приложение, табл. 1, 2). *DNAJC9* действует параллельно с классическими шаперонами гистонов, привлекая факторы теплового шока для защиты структурной целостности димеров гистонов H3–H4 на наиболее сложных стадиях их доставки и встраивания в хроматин, в том числе в участках, связанных с

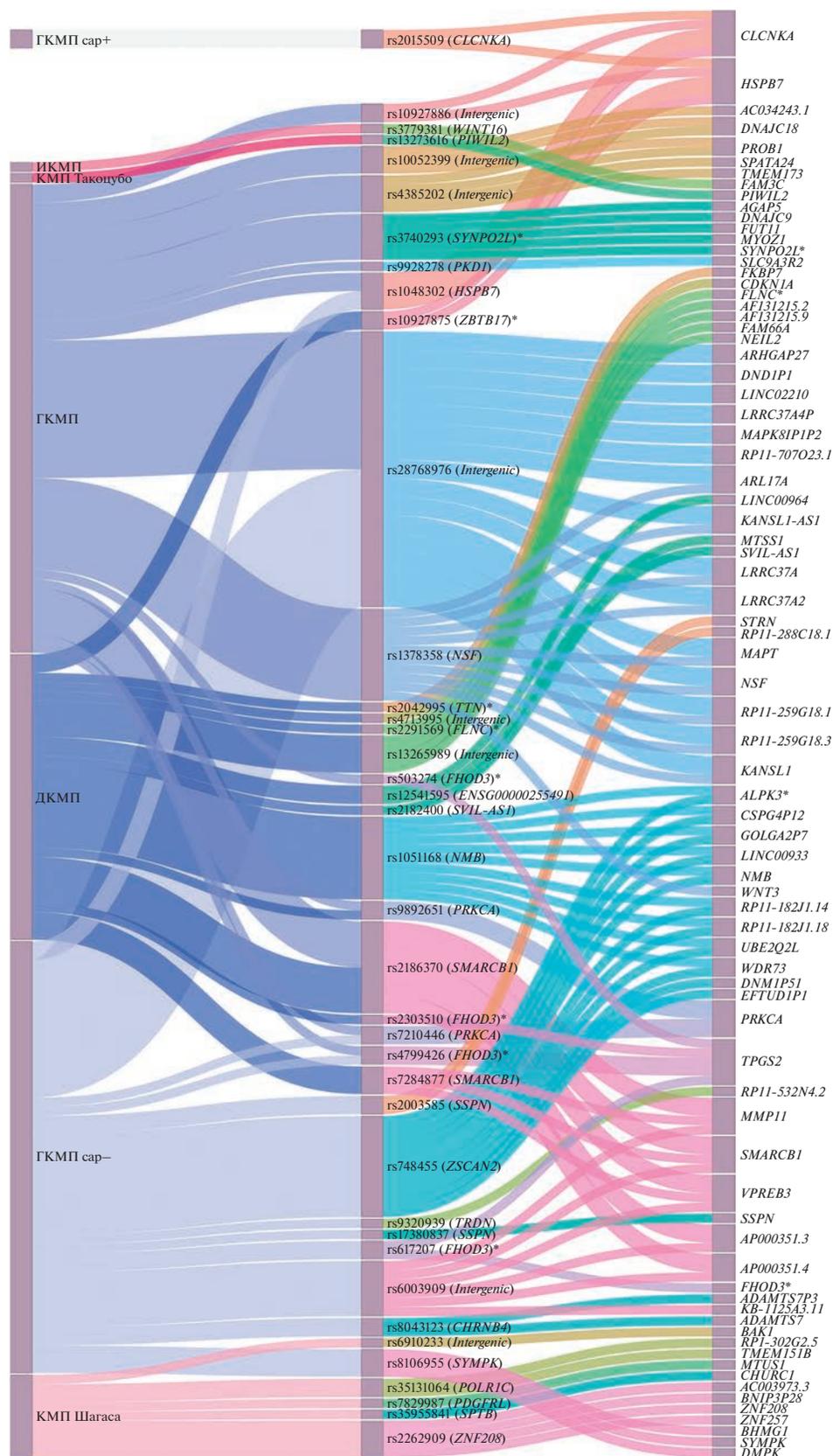


Рис. 1. Регуляторный потенциал SNPs, ассоциированных с кардиомиопатиями (КМП) по данным GWAS. Слева направо указаны фенотип КМП; SNP, ассоциированный с КМП (локализация генетического варианта); гены, на экспрессию которых в ушке предсердия или в левом желудочке сердца влияют ассоциированные SNP в качестве eQTL и/или sQTL. * Обозначены гены менделевских КМП.

репликацией и транскрипцией [67]. Примечательно, что уровень экспрессии и наличие сплайсинговых вариантов гена *HSPB7*, кодирующего один из белков теплового шока, зависят от четырех SNPs (eQTL и sQTL: межгенный rs10927886; sQTL: rs2015509, локализованный в интроне гена *CLCNKA*, rs1048302 – в экзоне/3'UTR гена *HSPB7*, rs10927875 – в интроне гена *ZBTB17*), ассоциированных с КМП по GWAS (ГКМП и ДКМП), и один из этих вариантов локализован в этом же гене (рис. 1; Приложение, табл. 1, 2). При этом связывание *HSPB7* с гистонами H3.1 и H4 зависит от DNAJC9 [67]. Значимость эпигенетических процессов показана для различных КМП, в том числе на уровне модификации гистонов, метилирования ДНК, а также изменения экспрессии широкого спектра регуляторных некодирующих РНК [68–73].

Вовлеченность в патогенез КМП регуляторных некодирующих РНК подтверждается, во-первых, локализацией ассоциированных с данной патологией SNPs в генах различных некодирующих РНК, во-вторых, регуляцией ассоциированных SNPs уровня экспрессии данного класса молекул в миокарде (рис. 1; Приложение, табл. 1). Как изменение структуры, так и изменение уровня регуляторных РНК могут повлиять на эффективность тех метаболических путей, которые обеспечивают работу сердца в норме и при развитии патологических процессов, в том числе и в воздействии неблагоприятных факторов (см., например, [68]). В частности, SNPs, ассоциированные с КМП, локализованы в экзоне *TTN* и в гене его антисмыслового транскрипта *TTN-AS1*, структурные и функциональные особенности которых могут приводить к развитию данной патологии [68]. Один из ассоциированных с ГКМП SNP локализован в *ENSG00000287149*, который кодирует антисмысловой транскрипт к *TTN* (в гене *TTN* известны патогенные/вероятно патогенные варианты для различных типов КМП – ДКМП, ГКМП, правожелудочковой АРКМП, РКМП) и *CCDC141* [29, 74].

Известно, что длинные некодирующие РНК (lncРНК) участвуют в регуляции активности генов, структурные варианты которых предрасполагают к развитию КМП. lncРНК *MIRF* в качестве конкурентной эндогенной регуляторной РНК (ceRNA) путем связывания с miR-26a влияет на уровень регулируемого данной микроРНК *Bak1*, тем самым способствуя апоптозу кардиомиоцитов [75]. В свою очередь, уровень экспрессии гена *BAK1* регулирует rs6910233, ассоциированный с КМП Шагаса по данным GWAS (рис. 1; Примечание, табл. 1, 2). lncРНК *LINC00964* отнесена к категории ключевых эндогенных РНК, вовлеченных в регуляцию генов, ассоциированных с фибрилляцией предсердия [76]. Уровень экспрессии данной lncРНК зависит от генотипов rs12541595 (интронный вариант *ENSG00000255491* (*LOC105375743*)),

который по данным GWAS ассоциирован с ДКМП (рис. 1; Приложение, табл. 1, 2). Ассоциированные с КМП SNPs локализованы и в генах других некодирующих РНК, которые открыты недавно, их роль и влияние структурной вариативности на функцию, в том числе в клетках миокарда, предстоит еще выяснить: в *ENSG00000257732* (кодирует антисмысловой транскрипт к *TXNRD1*), *ENSG00000285939* (кодирует антисмысловой транскрипт к *KIF2B*) и других [74].

В целом менее половины ассоциированных по данным GWAS с КМП SNPs (70 из 156 SNPs) не являются ни eQTL/sQTL, ни эпигенетическими метками, значимыми для поддержания структуры и функции хроматина. Патогенетическая значимость вариантов в некодирующих участках генома для развития КМП все больше привлекает исследователей. Ряд таких вариантов, локализованных в интронах, промоторах/энхансерах, 5'UTR и 3'UTR, в генах микроРНК, описаны при ГКМП, ДКМП и правожелудочковой АРКМП [31]. В данном случае важным является выявление тех генов и метаболических путей, которые оказываются под влиянием соответствующих регуляторных SNPs.

СВЯЗЬ ГЕНОВ, РЕГУЛИРУЕМЫХ АССОЦИИРОВАННЫМИ С КМП ВАРИАНТАМИ, С ГЕНАМИ МЕНДЕЛЕВСКИХ КАРДИОМИОПАТИЙ И С ПАТОЛОГИЕЙ СЕРДЦА

ALPK3, *FHOD3*, *FLNC*, *HSPB7* и некоторые другие гены, которые регулируются на уровне экспрессии и сплайсинга SNPs, ассоциированными с КМП по данным GWAS, относятся к числу генов, где часто выявляются новые варианты, приводящие к КМП, в том числе к семейным формам [77–83], а кодируемые ими белки важны для нормального функционирования сердца.

В ушке правого предсердия экспрессия гена *ALPK3* различается в зависимости от генотипов миссенс-варианта rs1051168 в гене *NMB* (ассоциирован с ДКМП), а в ушке правого предсердия и в ЛЖ – от генотипов rs748455, локализованного в *ZSCAN2/ENSG00000275120* (ассоциирован с саркомер-негативной ГКМП), причем характер влияния генотипов rs748455 на уровень экспрессии разнонаправлен для правого желудочка и ушка предсердия (рис. 1; Приложение, табл. 1, 2).

Ген *ALPK3* кодирует ядерную киназу, которая играет важную роль в ранней дифференцировке кардиомиоцитов [84]. J.C. Herkert с соавт. [85] показали, что редкие биаллельные варианты в гене *ALPK3* (сочетание вариантов с потерей функции и патогенных миссенс-вариантов) вызывают кардиомиопатию у детей, проявляющуюся переходом ДКМП в гипертрофию, часто с нарушением

сократительной функции. Варианты с потерей функции и патогенные миссенс-варианты в гетерозиготном состоянии данного гена регистрировались с повышенной частотой среди взрослых пациентов с КМП, у которых не были выявлены патогенные варианты в 83 протестированных генах КМП [85]. Кроме того, у пациентов с ГКМП с укороченным вариантом ALPK3 (ALPK3tv) наблюдали специфичность клинической картины: чаще регистрировались апикально-концентрическая гипертрофия и короткий интервал PR, вне зависимости от наличия у них патогенных вариантов в генах саркомерных белков [86]. При этом возраст на момент постановки диагноза и максимальная толщина стенки ЛЖ у таких пациентов были аналогичны саркомер-негативным пациентам, тогда как систолические нарушения ЛЖ и неустойчивая желудочковая тахикардия по частоте регистрации были сходны с таковыми для саркомер-позитивных пациентов.

Согласно данным GWAS четыре SNPs гена *FHOD3* ассоциированы с ГКМП (rs503274, rs118060942, rs4799426, rs617207) и один — с ДКМП (rs2303510) (рис. 1; Приложение, табл. 1). Интронный rs617207 гена *FHOD3*, ассоциированный с саркомер-негативной ГКМП, регулирует экспрессию гена своей локализации в ушке правого предсердия (рис. 1; Приложение, табл. 1, 2). Некоторые исследователи отмечают увеличение экспрессии *FHOD3* при ГКМП в ткани сердца [87]. Для этого гена не известны патогенные/вероятно патогенные варианты, но описаны варианты с неопределенной значимостью для ГКМП [29]. В то же время известно, что *FHOD3* играет решающую роль в развитии сердца на ранних этапах и в поддержании его функции во взрослом состоянии, участвуя в организации саркомеров [88–91], и нарушение структуры и функции данного белка может приводить к патологии сердца.

Ген *FHOD3* относят к новым генам ГКМП, редкие варианты в нем объясняют 1–2% случаев данной патологии [78]. По данным GWAS зарегистрирована ассоциация rs2303510 (p.Val1134Ile) гена *FHOD3* с ДКМП (рис. 1; Приложение, табл. 1), а в исследовании E.C. Wooten с соавт. [87] для этого же несинонимичного генетического варианта установлена связь с ГКМП. Ген *FHOD3* не только рассматривается в качестве причинного для развития ГКМП, но и определяет клинические особенности болезни, и наличие в нем патогенных вариантов представляет собой независимый риск смерти от сердечно-сосудистых заболеваний и смерти от всех причин при ГКМП [79].

Известны и другие варианты гена *FHOD3*, которые связывают с развитием КМП. Так, миссенс-вариант p.Tyr1249Asn, потенциально модифицирующий трехмерную структуру и нарушающий функцию *FHOD3*, препятствует сборке актиновых

филаментов в миокарде, тем самым способствуя развитию ДКМП [92]. Со сплайсинговым вариантом (rs2036163874; NM_001281740.3: c.1646+1G>A) в гене *FHOD3* косегрегирует некомпактность ЛЖ с аритмическими и тромботическими осложнениями, фиброзом миокарда и сердечной недостаточностью [93]. Кроме того, фосфорилирование белков *FHOD* (*FHOD1* и *FHOD3*) рассматривается в качестве одного из этапов патогенеза кардиомиопатий, вызванных мутациями в гене *LMNA* [94].

Генетические варианты филамина C (*FLNC*) связаны с различными сердечными и мышечными фенотипами, включая ГКМП и ДКМП, клиническими особенностями течения КМП (при ДКМП характерно злокачественное течение и высокий риск внезапной смерти) и др. [81, 83, 95–98]. Среди детей (средний возраст — 3 мес.) у обладателей редких вариантов в гене *FLNC* диагностированы РКМП (c.3557C>T, p.Ala1186Val; c.4916G>A, p.Cys1639Tyr) и ДКМП (сплайсинговый вариант c.2265+4del; c.1322G>T, p.Arg441Ile), а также аритмия (c.7942G>T, p.Ala2648Ser) [83]. Варианты сплайсинга в гене *FLNC* у пациентов с ДКМП описаны и в других исследованиях [97]. Укороченные варианты *FLNC* регистрировались у женщин с перипаретальной кардиомиопатией [99]. У пациентов с правожелудочковой АРКМП, несущих патогенные варианты в гене *FLNC*, экспрессия данного гена была снижена в ткани ЛЖ [98]. Более того, у данных пациентов на основании результатов транскриптомного анализа образцов тканей ЛЖ выявлено повышение уровня транскриптов 623 генов и снижение — 486 генов по сравнению с контрольными образцами. Помимо гена *FLNC* снижение экспрессии зарегистрировано и для других генов, ассоциированных с актином — *VCL*, *PARVB* и *MYL7*. На основании полученных результатов авторами процитированного исследования было высказано предположение о новом механизме правожелудочковой АРКМП, который отличается от классической формы, вызванной мутациями десмосомных генов [98].

Для некоторых генов, регулируемых ассоциированными с КМП SNPs, установлен протективный эффект в отношении данной патологии. У мышей в гипертрофированном сердце при стабильной общей экспрессии гена *Mtus1* регистрировали увеличение одного из сплайсинговых вариантов данного гена — *Mtus1A* (причем данный вариант оказывает ингибирующее влияние на гипертрофию сердца), что свидетельствует о специфичности сплайсинга при ГКМП [100]. У человека уровень экспрессии данного гена в ЛЖ зависит от rs7829987, ассоциированного с идиопатической ДКМП (рис. 1; Приложение, табл. 1, 2).

Сверхэкспрессия саркоспана (*SSPN*) облегчает клиническое течение кардиомиопатии, связанной с мышечной дистрофией Дюшенна [101].

Инtronный вариант rs17380837 гена *SSPN* выступает в качестве eQTL, и sQTL для своего гена, а уровень экспрессии данного гена возрастает с увеличением числа ассоциированных с ГКМП аллелей (рис. 1; Приложение, табл. 1, 2).

Связь с КМП отмечалась и для других генов, регулируемых SNPs, ассоциированными с КМП по данным GWAS. Локализованный в гене *HSPB7* rs1739843 ассоциирован с идиопатической ДКМП (минорный аллель *T*, регистрируемый с частотой 39%, обладал протективным эффектом) в европейских когортах [102]. А.А. Стрельцов с соавт. [103] отмечали, что генотип *TT* и аллель *T* по данному rs1739843 чаще, чем у здоровых индивидов, регистрировали у пациентов с ГКМП (в возрасте старше 45 лет) и хронической сердечной недостаточностью с сохраненной фракцией выброса ЛЖ ($\geq 50\%$), и у лиц с малосимптомным течением ГКМП (с более благоприятным прогнозом).

С предрасположенностью к фибрилляции предсердий, которая может рассматриваться как предсердная кардиомиопатия, связывают редкий вариант сайта сплайсинга (rs766868752), локализованный в интроне гена *SYNPO2L* [104], и изменение экспрессии гена *FKBP7* в ушке левого предсердия [105].

CDKN1A относится к числу генов, значимых для прогнозирования ДКМП в клинической практике [106]. С частым вариантом (rs4135240) этого гена также ассоциирована сердечная недостаточность у пациентов с неишемической КМП, данная ассоциация воспроизводилась в независимой выборке пациентов, сохранялась после коррекции факторов риска сердечной недостаточности и была значима при анализе показателей дисфункции ЛЖ у лиц без клинических проявлений сердечной недостаточности [107].

Также варианты с неопределенной значимостью для ДКМП выявлены для гена *ZBTB17*, в котором локализован ассоциированный с этим же типом КМП (ДКМП) по данным GWAS rs10927875, причем этот SNP является eQTL для гена *CLCNKA* в ЛЖ и sQTL — для гена *HSPB7* в ЛЖ и в ушке предсердия сердца (рис. 1; Приложение, табл. 2). Интересно, что 31 SNP, ассоциированный с КМП, изменяет экспрессию одного и более генов (всего 29 SNP) и/или является sQTL (всего 9 SNP) для ряда генов в ЛЖ (Приложение, табл. 2), 18 из них ассоциированы с ГКМП по данным GWAS.

ВОВЛЕЧЕННОСТЬ ГЕНЕТИЧЕСКИХ ВАРИАНТОВ, АССОЦИИРОВАННЫХ С КАРДИОМИОПАТИЯМИ, В РЕГУЛЯЦИЮ ВАРИАбельНОСТИ ЭКГ- И ЭхоКГ-ПАРАМЕТРОВ СЕРДЦА

Известно, что у пациентов с различными КМП регистрируются изменения на ЭКГ/ЭхоКГ и, более

того, они могут выступать в качестве диагностических признаков некоторых КМП [86, 108–110]. Это позволяет предположить, что варианты и гены, которые значимы для развития КМП, вносят вклад в вариабельность патогенетически значимых/диагностических ЭКГ- и ЭхоКГ-параметров сердца. Действительно, по данным GWAS из числа ассоциированных с КМП вариантов для 19 SNPs выявлена также связь с изменчивостью параметров, отражающих функциональное состояние сердца и сердечно-сосудистой системы (ЭКГ- и ЭхоКГ-параметры, биохимические показатели), или с другими патологическими состояниями сердца и сосудов (фибрилляция предсердий, пролапс митрального клапана и др.; Приложение, табл. 1). В частности, с продолжительностью комплекса QRS ассоциации зарегистрированы для шести SNPs. Из этих полиморфных вариантов два (rs17362588 и rs3176326) были ассоциированы с ГКМП, а еще четыре (rs12319113, rs1356410, rs182503338, rs10769783) показали не только связь с КМП Шага, но и были вовлечены в вариабельность указанного ЭКГ-параметра у пациентов, серопозитивных по *Tripinosoma cruzi* (провоцирует развитие данной КМП). Четыре SNPs участвовали в детерминации изменчивости интервала PR, включая rs41306688, rs3176326 и rs4894803 (ассоциированы также с саркомер-негативной ГКМП) и rs2042995 (ассоциирован с ДКМП). Два ассоциированных с ГКМП SNPs влияли на изменчивость интервала QT: rs846111 и rs28436726 (для второго из указанных SNPs ассоциация зарегистрирована с динамикой QT во время тренировки). Еще четыре SNPs оказались информативными при изучении в рамках GWAS структуры электрокардиограммы (амплитуды в разных временных точках) — это rs2234962 (ассоциирован с ДКМП), rs4894803 и rs846111 (ассоциированы с гипертрофической КМП) и rs2186370 (ассоциирован с ДКМП и с ГКМП).

Кроме того, для rs3176326 и rs17380837 (оба ассоциированы с ГКМП) зарегистрирована связь с фибрилляцией предсердий, а для rs17362588 (ассоциирован с ГКМП) — с частотой сердечных сокращений и реакцией сердечного ритма на восстановление после тренировки (Приложение, табл. 1). Следует отметить, что на начальной стадии проявления КМП может наблюдаться фибрилляция предсердий (ФП), а у 10% индивидов с ранней ФП выявлялись патогенный/вероятно патогенные варианты в генах КМП, и частота таких вариантов была выше у лиц с ранним началом ФП (до 30 лет), по сравнению с теми, у кого ФП регистрировалась в позднем возрасте (старше 60 лет) [111].

С рядом ЭхоКГ-параметров, такими как диастолический внутренний размер ЛЖ, фракция выброса ЛЖ, конечно-диастолический и конечно-систолический объем ЛЖ, общая окружная деформация ЛЖ и глобальная радиальная деформация

ЛЖ, зарегистрированы ассоциации rs72840788, rs12541595 и rs2042995 (первый из указанных SNPs ассоциирован с ГКМП, а два последних — с ДКМП; Приложение, табл. 1).

Среди этих SNPs есть и те, которые локализованы в общих генах, ассоциированных с ДКМП и ГКМП (*BAG3* — rs2234962, rs17099139, rs72840788; *SMARCB1* — rs2186370; *PRKCA* — rs9892651), и в генах менделевских КМП (*BAG3*, *TTN-AS1/TTN* — rs2042995; Приложение, табл. 1). К числу вовлеченных в вариабельность ЭКГ-/ЭхоКГ-параметров исследователи GWAS относили и другие SNPs и гены менделевских КМП (*MYH6* — rs452036), а также SNPs, которые локализованы в (или вблизи) генах, варианты которых были ассоциированы с различными КМП (*HSPB7/CLCNKA* — rs1763604 (локализован вблизи генов), *PRKCA* — rs9909004, *SYNPO2L* — rs3812629) [112, 113]. При этом ассоциации для rs2234962 (*BAG3*) и rs1763604 (вблизи *HSPB7/CLCNKA*) с ЭКГ-параметрами, зарегистрированные у пациентов с ДКМП, сохранялись у людей, у которых не диагностированы ДКМП или другие заболевания сердца, что указывает на то, что развитие патологии не определяет связи между этими вариантами и ЭКГ-параметрами [112].

На модельных объектах (мышь) показано, что нокаут гена *HSPB7* приводит к быстрому развитию сердечной недостаточности и внезапной смерти, причем перед гибелью у мышей регистрировали сердечную аритмию с аномальной проводимостью [114]. В то же время сверхэкспрессия данного белка защищает от ремоделирования сердца и тахикардии за счет ослабления пути RhoA GTPase на разных уровнях [115]. Согласно данным GWAS ряд ассоциированных с КМП SNPs выступают в тканях сердца для гена *HSPB7* в качестве sQTL (rs2015509 и rs1048302 — локализованы в генах *CLCNKA* и *HSPB7* соответственно, ассоциированы с ГКМП; rs10927875 — локализован в гене *ZBTB17*, ассоциирован с ДКМП), а межгенный rs10927886 является и sQTL, и eQTL для данного гена (рис. 1; Приложение, табл. 1, 2). Кроме того, rs1048302, rs10927886 и rs10927875 являются эпигенетическими метками, что также может влиять на функциональное состояние как гена *HSPB7*, так и других функционально значимых для сердца генов.

С различными ЭКГ-параметрами (интервал PR, продолжительность комплекса QRS и др.) и фибрилляцией предсердий зарегистрирована ассоциация интронного варианта rs3176326 гена *CDKN1A*, который ассоциирован с ГКМП (Приложение, табл. 1). При этом установлена корреляция между уровнем экспрессии *CDKN1A* и фракцией выброса левого желудочка [116] — важного показателя, отражающего функциональное состояние сердца, в том числе и при развитии КМП.

Межгенный rs28768976, ассоциированный по данным GWAS с ГКМП, определяет уровень экс-

прессии гена *ARHGAP27* в сердце (рис. 1; Приложение, табл. 1, 2), при этом на модельных объектах было показано, что транскрипция *Arhgap27* в сердце коррелировала с интервалом PR [116].

Таким образом, среди регулируемых генов, SNPs которых ассоциированы с КМП по данным GWAS, есть гены структурно и функционально значимые для нормальной работы сердца. Эти данные свидетельствуют о том, что локализованные в генах КМП как редкие патогенные варианты, так и частые, регулирующие их экспрессию, могут выступать в качестве причинных для развития данной патологии, а также в качестве генетических факторов, модифицирующих клиническую картину КМП.

ЗАКЛЮЧЕНИЕ

Приведенные в обзоре результаты работ свидетельствуют о значимости в развитии различных КМП не только редких патогенных/вероятно патогенных вариантов, но и частых вариантов генов, в том числе выявляемых при проведении GWAS.

Патогенные/вероятно патогенные варианты оказывают значительное влияние на формирование патологического фенотипа, но они редко встречаются в популяциях и не позволяют объяснить все выявляемые случаи КМП. Вклад частых полиморфных SNPs в риск развития КМП как правило невелик, но эти варианты могут формировать полигенную основу развития данной патологии. Действительно, накапливаются данные о том, что на риск развития КМП и патологический фенотип при данном заболевании может влиять совокупность частых полиморфных вариантов [48, 61, 117]. Особое место в выявлении вклада полиморфных вариантов в формирование риска развития различных кардиомиопатий принадлежит GWAS.

При проведении GWAS ГКМП установлено сильное полигенное влияние на риск развития данной патологии у лиц без патогенных мутаций в генах саркомерных белков (риск в 2 раза ниже в самом низком квинтиле и в 2 раза выше в самом высоком квинтиле), а также на тяжесть течения патологии у носителей патогенных вариантов в генах саркомера [48]. Значительная доля риска ГКМП была связана с аддитивными эффектами распространенных SNPs (частота минорного аллеля >0.01): оценки наследуемости (h_g^2) SNPs, рассчитанные с использованием GREML-LDMS, для выборки регистра КМП составили 0.35, для выборки BioResource for Rare Diseases — 0.68, при этом оценка наследуемости GREML для саркомер-негативной ГКМП превышала таковую для саркомер-позитивной ГКМП (0.34 и 0.16 соответственно) [48].

Полигенный фон может определять и то, какая форма КМП разовьется, а также особенности клинической картины. При проведении GWAS установлено полигенное влияние на тяжесть течения ГКМП у носителей патогенных вариантов в генах саркомера [48]. В другом GWAS для ГКМП, ДКМП и девяти параметров ЛЖ, полученных на основании магнитно-резонансной томографии сердца, выявлены соответственно 16, 13 и 23 ассоциированных локусов [61]. При этом, несмотря на то, что восемь из 16 локусов, ассоциированных с ГКМП, были связаны также с ДКМП, а 12 локусов ДКМП – с ГКМП, эти ассоциации имели разнонаправленное влияние при ГКМП и ДКМП (за исключением только локуса *DCM4*, локализованного рядом с геном ДКМП *TTN*). Важным представляется также то, что эти маркеры были ассоциированы с морфологическими параметрами ЛЖ в популяционной выборке [61].

При изучении генетической компоненты количественных показателей ЛЖ, полученных на основании магнитно-резонансной томографии сердца, с использованием подхода GWAS также показано, что полигенные факторы, участвующие в контроле систолического объема ЛЖ, не только связаны с возникновением ДКМП в общей популяции, но и влияют на размер и функцию сердца у обладателей редких патогенных вариантов, ведущих к возникновению преждевременного стоп-кодона в гене *TTN* (одна из самых частых из известных причин ДКМП) [117].

В целом как редкие, так и распространенные генетические варианты вносят существенный вклад в предрасположенность ГКМП в общей популяции и улучшают прогнозирование риска данной патологии по сравнению с тем, что достигается с помощью клинических факторов (таких как ожирение, артериальная гипертензия, фибрилляция предсердий и ишемическая болезнь сердца) [118]. Олигогенная модель (сочетание нескольких неблагоприятных вариантов) все чаще привлекается для объяснения как причин развития КМП, так и их клинической картины [119–121].

Локализация ассоциированных с КМП по GWAS полиморфных вариантов, в том числе и в генах менделевских КМП и в некодирующих участках генома, их функциональная значимость (вовлеченность в регулируя экспрессирующихся в сердце генов) и связь не только с КМП, но и со структурно-функциональными параметрами сердца (ЭхоКГ- и ЭКГ-параметры), свидетельствуют о неслучайности выявляемых ассоциаций.

Таким образом, к числу причинных/повышающих риск развития факторов КМП относятся как редкие, так и частые генетические варианты, а генетическая компонента КМП может реализовываться через моногенную, олигогенную и по-

лигенные модели. Следует отметить, что среди патогенетически значимых для развития КМП как редких, так и частых вариантов преимущественно анализируются SNPs, но практически нет сведений о таком функционально значимом классе маркеров как CNV (Copy number variation – вариации числа копий участков ДНК) [122]. Кроме того, все еще недостаточно исследована роль генетических вариантов, локализованных в некодирующих участках генома, вариантов, обладающих регуляторным потенциалом, в том числе и для метаболических путей, критически значимых для нормального функционирования сердца. Это указывает на актуальность проведения функциональных исследований, в том числе и на образцах тканей сердца. Лишь комплексный подход к оценке значимости структурно-функциональной вариабельности генома в норме и при развитии патологии позволит не только детализировать патогенез КМП, но и выявить те критические этапы/периоды, воздействием на которые можно предупредить или приостановить развитие патологии.

Работа выполнена в рамках Государственного задания Министерства науки и высшего образования № 122020300041-7.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *McKenna W.J., Maron B.J., Thiene G.* Classification, epidemiology, and global burden of cardiomyopathies // *Circ. Res.* 2017. V. 121. № 7. P. 722–730. <https://doi.org/10.1161/CIRCRESAHA.117.309711>
2. *Salemi V.M.C., Mohty D., Altavilla S.L.L. et al.* Insights into the classification of cardiomyopathies: past, present, and future directions // *Clinics (Sao Paulo)*. 2021. V. 76. P. e2808. <https://doi.org/10.6061/clinics/2021/e2808>
3. *McKenna W.J., Judge D.P.* Epidemiology of the inherited cardiomyopathies // *Nat. Rev. Cardiol.* 2021. V. 18. № 1. P. 22–36. <https://doi.org/10.1038/s41569-020-0428-2>
4. *Ommen S.R., Mital S., Burke M.A. et al.* AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines // *Circulation*. 2020. V. 142. Iss. 25. P. e558–e631. <https://doi.org/10.1161/CIR.0000000000000937>
5. *Corrado D., Link M.S., Calkins H.* Arrhythmogenic right ventricular cardiomyopathy // *N. Engl. J. Med.* 2017. V. 376. № 1. P. 61–72. <https://doi.org/10.1056/NEJMra1509267>

6. Ciarambino T., Menna G., Sansone G., Giordano M. Cardiomyopathies: an overview // *Int. J. Mol. Sci.* 2021. V. 22. Iss. 14. <https://doi.org/10.3390/ijms22147722>
7. Hey T.M., Rasmussen T.B., Madsen T. et al. Clinical and genetic investigations of 109 index patients with dilated cardiomyopathy and 445 of their relatives // *Circ. Heart Fail.* 2020. V. 13. № 10. P. e006701. <https://doi.org/10.1161/circheartfailure.119.006701>
8. Robertson J., Lindgren M., Schaufelberger M. et al. Body mass index in young women and risk of cardiomyopathy: A long-term follow-up study in Sweden // *Circulation.* 2020. V. 141. Iss. 7. P. 520–529. <https://doi.org/10.1161/circulationaha.119.044056>
9. Peters S., Johnson R., Birch S. et al. Familial dilated cardiomyopathy // *Heart Lung Circ.* 2020. V. 29. Iss. 4. P. 566–574. <https://doi.org/10.1016/j.hlc.2019.11.018>
10. Замараева Д.В., Трунина И.И., Котлукова Н.П. и др. Дебют генетически обусловленной дилатационной кардиомиопатии в исходе перенесенного миокардита (клинический случай) // *Клин. и эксперим. хирургия. Журн. им. акад. Б.В. Петровского.* 2020. Т. 8. № 3(29). С. 110–118. <https://doi.org/10.33029/2308-1198-2020-8-3-110-118>
11. Povysil G., Chazara O., Carss K.J. et al. Assessing the role of rare genetic variation in patients with heart failure // *JAMA Cardiol.* 2021. V. 6. № 4. e206500. <https://doi.org/10.1001/jamacardio.2020.6500>
12. Tiron C., Campuzano O., Fernández-Falgueras A. et al. Prevalence of pathogenic variants in cardiomyopathy-associated genes in myocarditis // *Circ. Genom. Precis. Med.* 2022. V. 15. № 3. <https://doi.org/10.1161/CIRCGEN.121.003408>
13. Patel A.P., Dron J.S., Wang M. et al. Association of pathogenic DNA variants predisposing to cardiomyopathy with cardiovascular disease outcomes and all-cause mortality // *JAMA Cardiol.* 2022. V. 7. № 7. P. 723–732. <https://doi.org/10.1001/jamacardio.2022.0901>
14. Lazarte J., Jurgens S.J., Choi S.H. et al. LMNA variants and risk of adult-onset cardiac disease // *J. Am. Coll. Cardiol.* 2022. V. 80. № 1. P. 50–59. <https://doi.org/10.1016/j.jacc.2022.04.035>
15. Cipriani A., Perazzolo Marra M., Bariani R. et al. Differential diagnosis of arrhythmogenic cardiomyopathy: phenocopies versus disease variants // *Minerva Med.* 2021. V. 112. № 2. P. 269–280. <https://doi.org/10.23736/S0026-4806.20.06782-8>
16. Mattesi G., Cipriani A., Bauce B. et al. Arrhythmogenic left ventricular cardiomyopathy: Genotype-phenotype correlations and new diagnostic criteria // *J. Clin. Med.* 2021. V. 10. Iss. 10. <https://doi.org/10.3390/jcm10102212>
17. Lee T.M., Hsu D.T., Kantor P. et al. Pediatric cardiomyopathies // *Circ. Res.* 2017. V. 21. № 7. P. 855–873. <https://doi.org/10.1161/CIRCRESAHA.116.309386>
18. Лутохина Ю.А., Благова О.В., Шестак А.Г. и др. Сочетание аритмогенной дисплазии правого желудочка и некомпактного миокарда левого желудочка как особая форма кардиомиопатии: клиника, диагностика, генетическая природа, течение // *Вестн. РАМН.* 2020. Т. 75. № 6. С. 594–604. <https://doi.org/10.15690/vramn1245>
19. Blagova O., Alieva I., Kogan E. et al. Mixed hypertrophic and dilated phenotype of cardiomyopathy in a patient with homozygous in-frame deletion in the MYBPC3 gene treated as myocarditis for a long time // *Front. Pharmacol.* 2020. V. 11. <https://doi.org/10.3389/fphar.2020.579450>
20. Jefferies J.L., Wilkinson J.D., Sleeper L.A. et al. Cardiomyopathy phenotypes and outcomes for children with left ventricular myocardial noncompaction: results from the Pediatric Cardiomyopathy Registry // *J. Card. Fail.* 2015. V. 21. № 1. P. 877–884. <https://doi.org/10.1016/j.cardfail.2015.06.381>
21. Webber S.A., Lipshultz S.E., Sleeper L.A. et al. Outcomes of restrictive cardiomyopathy in childhood and the influence of phenotype: A report from the Pediatric Cardiomyopathy Registry // *Circulation.* 2012. V. 126. Iss. 10. P. 1237–1244. <https://doi.org/10.1161/CIRCULATIONAHA.112.104638>
22. Lipshultz S.E., Orav E.J., Wilkinson J.D. et al. Risk stratification at diagnosis for children with hypertrophic cardiomyopathy: An analysis of data from the Pediatric Cardiomyopathy Registry // *Lancet.* 2013. V. 382. № 9908. P. 1889–1897. [https://doi.org/10.1016/S0140-6736\(13\)61685-2](https://doi.org/10.1016/S0140-6736(13)61685-2)
23. Кучер А.Н., Валуахметов Н.Р., Салахов Р.Р. и др. Фенотипическая вариабельность гипертрофической кардиомиопатии у носителей патогенного варианта р.Arg870His гена MYH7 // *Бюл. сиб. медицины.* 2022. Т. 21. № 3. С. 205–216. <https://doi.org/10.20528/1682-0363-2022-3-205-216>
24. Кучер А.Н., Слепцов А.А., Назаренко М.С. Генетический ландшафт дилатационной кардиомиопатии // *Генетика.* 2022. Т. 58. № 4. С. 371–387. <https://doi.org/10.31857/S0016675822030080>
25. Menon S.C., Michels V.V., Pellikka P.A. et al. Cardiac troponin T mutation in familial cardiomyopathy with variable remodeling and restrictive physiology // *Clin. Genet.* 2008. V. 74. Iss. 5. P. 445–454. <https://doi.org/10.1111/j.1399-0004.2008.01062.x>
26. Norrish G., Cleary A., Field E. et al. Clinical features and natural history of preadolescent nonsyndromic hypertrophic cardiomyopathy // *J. Am. Coll. Cardiol.* 2022. V. 79. № 20. P. 1986–1997. <https://doi.org/10.1016/j.jacc.2022.03.347>
27. Wu W., Lu C.X., Wang Y.N. et al. Novel phenotype–genotype correlations of restrictive cardiomyopathy with myosin-binding protein c (MYBPC3) gene mutations tested by next-generation sequencing // *J. Am. Heart Assoc.* 2015. V. 4. № 7. <https://doi.org/10.1161/JAHA.115.001879>
28. Курушко Т.В., Вайханская Т.Г., Булгак А.Г. и др. Ламин А/С ассоциированная дилатационная кардиомиопатия: вариабельность клинических проявлений // *Кардиология в Беларуси.* 2018. Т. 10. № 6. С. 892–903.
29. Simple ClinVar [Electronic resource]. URL: <https://simple-clinvar.broadinstitute.org/>. Accessed 03.2022.
30. Pérez-Palma E., Gramm M., Nürnberg P. et al. Simple ClinVar: An interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database // *Nucl. Acids Res.* 2019. V. 47. № W1. P. W99–W105. <https://doi.org/10.1093/nar/gkz411>
31. Salman O.F., El-Rayess H.M., Abi Khalil C. et al. Inherited cardiomyopathies and the role of mutations in non-coding regions of the genome // *Front. Cardiovasc. Med.* 2018. V. 5. № 77. <https://doi.org/10.3389/fcvm.2018.00077>

32. Комиссарова С.М., Чакова Н.Н., Ниязова С.С. Гипертрофическая кардиомиопатия: анализ связи генотипа и фенотипа у пациентов с высоким риском внезапной смерти // *Мед. генетика*. 2018. Т. 17. № 11. С. 34–42.
33. Walsh R., Offerhaus J.A., Tadros R., Bezzina C.R. Minor hypertrophic cardiomyopathy genes, major insights into the genetics of cardiomyopathies // *Nat. Rev. Cardiol.* 2022. V. 9. № 3. P. 151–167. <https://doi.org/10.1038/s41569-021-00608-2>
34. Ingles J., Goldstein J., Thaxton C. et al. Evaluating the clinical validity of hypertrophic cardiomyopathy genes // *Circ. Genom. Precis. Med.* 2019. V. 12. № 2. <https://doi.org/10.1161/CIRCGEN.119.002460>
35. Zigova M., Bernasovska J., Boronova I. et al. Finding the candidate sequence variants for diagnosis of hypertrophic cardiomyopathy in East Slovak patients // *J. Clin. Lab. Anal.* 2018. V. 32. № 3. <https://doi.org/10.1002/jcla.22303>
36. Richmond C.M., James P.A., Pantaleo S. et al. Clinical and laboratory reporting impact of ACMG-AMP and modified ClinGen variant classification frameworks in *MYH7*-related cardiomyopathy // *Genet. Med.* 2021. V. 23. Iss. 6. P. 1108–1115. <https://doi.org/10.1038/s41436-021-01107-y>
37. Kubo T., Gimeno J.R., Bahl A. et al. Prevalence, clinical significance, and genetic basis of hypertrophic cardiomyopathy with restrictive phenotype // *J. Am. Coll. Cardiol.* 2007. V. 49. № 25. P. 2419–2426. <https://doi.org/10.1016/j.jacc.2007.02.061>
38. Vio R., Angelini A., Basso C. et al. Hypertrophic cardiomyopathy and primary restrictive cardiomyopathy: similarities, differences and phenocopies // *J. Clin. Med.* 2021. V. 10. № 9. Article 1954. <https://doi.org/10.3390/jcm10091954>
39. Li S., Wu B., Yin G. et al. MRI characteristics, prevalence, and outcomes of hypertrophic cardiomyopathy with restrictive phenotype // *Radiol. Cardiothorac. Imaging.* 2020. V. 2. № 4. P. e190158. <https://doi.org/10.1148/ryct.2020190158>
40. Hershberger R.E., Jordan E. Dilated Cardiomyopathy Overview. 2007 Jul 27 [updated 2022 Apr 7] // *GeneReviews®* [Internet] / Eds Adam M.P., Ardinger H.H., Pagon R.A. et al. Seattle (WA): Univ. Washington, Seattle, 1993–2022.
41. Towbin J.A., McKenna W.J., Abrams D.J. et al. HRS expert consensus statement on evaluation, risk stratification, and management of arrhythmogenic cardiomyopathy // *Heart Rhythm.* 2019. V. 16. № 11. P. e301–e372. <https://doi.org/10.1016/j.hrthm.2019.05.007>
42. Groeneweg J.A., Bhonsale A., James C.A. et al. Clinical presentation, long-term follow-up, and outcomes of 1001 arrhythmogenic right ventricular dysplasia/cardiomyopathy patients and family members // *Circ. Cardiovasc. Genet.* 2015. V. 8. № 3. P. 437–446. <https://doi.org/10.1161/CIRCGENETICS.114.001003>
43. Pugh T.J., Kelly M.A., Gowrisankar S. et al. The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing // *Genet. Med.* 2014. V. 16. Iss. 8. P. 601–608. <https://doi.org/10.1038/gim.2013.204>
44. Jordan E., Hershberger R.E. Considering complexity in the genetic evaluation of dilated cardiomyopathy // *Heart.* 2021. V. 107. № 2. P. 106–112. <https://doi.org/10.1136/heartjnl-2020-316658>
45. Morales A., Kinnamon D.D., Jordan E. et al. Variant Interpretation for Dilated Cardiomyopathy: Refinement of the American College of Medical Genetics and Genomics/ClinGen Guidelines for the DCM Precision Medicine Study // *Circ. Genom. Precis. Med.* 2020. V. 13. № 2. <https://doi.org/10.1161/CIRCGEN.119.002480>
46. Lopes L.R., Zekavati A., Syrris P. et al. Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing // *J. Med. Genet.* 2013. V. 50. № 4. P. 228–239. <https://doi.org/10.1136/jmedgenet-2012-101270>
47. Tran Vu M.T., Nguyen T.V., Huynh N.V. et al. Presence of hypertrophic cardiomyopathy related gene mutations and clinical manifestations in Vietnamese patients with hypertrophic cardiomyopathy // *Circ. J.* 2019. V. 83. № 9. P. 1908–1916. <https://doi.org/10.1253/circj.CJ-19-0190>
48. Harper A.R., Goel A., Grace C. et al. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity // *Nat. Genet.* 2021. V. 53. № 2. P. 135–142. <https://doi.org/10.1038/s41588-020-00764-0>
49. Villard E., Perret C., Gary F. et al. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy // *Eur. Heart J.* 2011. V. 32. Iss. 9. P. 1065–1076. <https://doi.org/10.1093/eurheartj/ehr105>
50. Meder B., Rühle F., Weis T. et al. A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy // *Eur. Heart J.* 2014. V. 35. Iss. 16. P. 1069–1077. <https://doi.org/10.1093/eurheartj/eh251>
51. Xu H., Dorn G.W. 2nd, Shetty A. et al. A Genome-wide association study of idiopathic dilated cardiomyopathy in African Americans // *J. Pers. Med.* 2018. V. 8. № 1. Article 11. <https://doi.org/10.3390/jpm8010011>
52. de Denus S., Mottet F., Korol S. et al. A genetic association study of heart failure: More evidence for the role of *BAG3* in idiopathic dilated cardiomyopathy // *ESC Heart Fail.* 2020. V. 7. № 6. P. 4384–4389. <https://doi.org/10.1002/ehf2.12934>
53. Garnier S., Harakalova M., Weiss S. et al. Genome-wide association analysis in dilated cardiomyopathy reveals two new players in systolic heart failure on chromosomes 3p25.1 and 22q11.23 // *Eur. Heart J.* 2021. V. 42. Iss. 20. P. 2000–2011. <https://doi.org/10.1093/eurheartj/ehab030>
54. Deng X., Sabino E.C., Cunha-Neto E. et al. Genome wide association study (GWAS) of Chagas cardiomyopathy in *Trypanosoma cruzi* seropositive subjects // *PLoS One.* 2013. V. 8. № 11. P. e79629. <https://doi.org/10.1371/journal.pone.0079629>
55. Casares-Marfil D., Strauss M., Bosch-Nicolau P. et al. A genome-wide association study identifies novel susceptibility loci in chronic chagas cardiomyopathy // *Clin. Infect. Dis.* 2021. V. 73. № 4. P. 672–679. <https://doi.org/10.1093/cid/ciab090>
56. Eitel I., Moeller C., Munz M. et al. Genome-wide association study in takotsubo syndrome – Preliminary results and future directions // *Int. J. Cardiol.* 2017. V. 236. P. 335–339. <https://doi.org/10.1016/j.ijcard.2017.01.093>
57. López-Mejías R., Carmona F.D., Genre F. et al. Identification of a 3'-untranslated genetic variant of *RARB* associated with carotid intima-media thickness in

- rheumatoid arthritis: A Genome-Wide Association Study // *Arthritis Rheumatol.* 2019. V. 71. № 3. P. 351–360.
<https://doi.org/10.1002/art.40734>
58. Wang X., Sun C.L., Quiñones-Lombrana A. et al. *CELF4* variant and anthracycline-related cardiomyopathy: A Children's Oncology Group Genome-Wide Association Study // *J. Clin. Oncol.* 2016. V. 34. Iss. 8. P. 863–870.
<https://doi.org/10.1200/JCO.2015.63.4550>
 59. Horne B.D., Rasmuson K.D., Alharethi R. et al. Genome-wide significance and replication of the chromosome 12p11.22 locus near the *PTHLH* gene for peripartum cardiomyopathy // *Circ. Cardiovasc. Genet.* 2011. V. 4. № 4. P. 359–366.
<https://doi.org/10.1161/CIRCGENETICS.110.959205>
 60. GWAS Catalog [Electronic resource]. URL: <https://www.ebi.ac.uk/gwas/>. Accessed 05.2022.
 61. Tadros R., Francis C., Xu X. et al. Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect // *Nat. Genet.* 2021. V. 53. № 2. P. 128–134.
<https://doi.org/10.1038/s41588-020-00762-2>
 62. UCSC Genome Browser on Human (GRCh38/hg38) <https://genome.ucsc.edu/>.
 63. VannoPortal Index [Electronic resource]. URL: <http://www.mulinlab.org/vportal/index.html/>. Accessed 05.2022.
 64. National Center for Biotechnology Information Electronic resource]. URL: <https://www.ncbi.nlm.nih.gov/>. Accessed 06.2022.
 65. Genotype-Tissue Expression (GTEx) Portal [Electronic resource]. URL: <https://gtexportal.org/home/>. Accessed 06.2022.
 66. Ahn J., Wu H., Lee K. Integrative analysis revealing human heart-specific genes and consolidating heart-related phenotypes // *Front Genet.* 2020. V. 11.
<https://doi.org/10.3389/fgene.2020.00777>
 67. Hammond C.M., Bao H., Hendriks I.A. et al. DNAJC9 integrates heat shock molecular chaperones into the histone chaperone network // *Mol. Cell.* 2021. V. 81. № 12. P. 2533–2548.e9.
<https://doi.org/10.1016/j.molcel.2021.03.041>
 68. Кучер А.Н., Назаренко М.С. Регуляторный потенциал ко-локализованных с генами кардиомиопатий некодирующих РНК // *Генетика.* 2023. Т. 59. № 4. С. 381–402.
 69. Кучер А.Н., Назаренко М.С. Эпигенетика кардиомиопатий: модификации гистонов и метилирование ДНК // *Генетика.* 2023. Т. 59. № 3. С. 266–282.
 70. Scolari F.L., Faganello L.S., Garbin H.I. et al. A systematic review of microRNAs in patients with hypertrophic cardiomyopathy // *Int. J. Cardiol.* 2021. V. 327. P. 146–154.
<https://doi.org/10.1016/j.ijcard.2020.11.004>
 71. Gao J., Collyer J., Wang M. et al. Genetic dissection of hypertrophic cardiomyopathy with myocardial RNA-Seq // *Int. J. Mol. Sci.* 2020. V. 21. № 9.
<https://doi.org/10.3390/ijms21093040>
 72. Huang G., Liu J., Yang C. et al. RNA sequencing discloses the genome wide profile of long noncoding RNAs in dilated cardiomyopathy // *Mol. Med. Rep.* 2019. V. 19. № 4. P. 2569–2580.
<https://doi.org/10.3892/mmr.2019.9937>
 73. Hailu F.T., Karimpour-Fard A., Toni L.S. et al. Integrated analysis of miRNA-mRNA interaction in pediatric dilated cardiomyopathy // *Pediatr. Res.* 2022. V. 92. № 1. P. 98–108.
<https://doi.org/10.1038/s41390-021-01548-w>
 74. Ensembl Genome Browser [Electronic resource]. URL: <https://www.ensembl.org/index.html/>. Accessed 06.2022.
 75. Su X., Lv L., Li Y. et al. lncRNA MIRF promotes cardiac apoptosis through the miR-26a-Bak1 axis // *Mol. Ther. Nucl. Acids.* 2020. V. 20. P. 841–850.
<https://doi.org/10.1016/j.omtn.2020.05.002>
 76. Liu Y., Liu N., Bai F., Liu Q. Identifying ceRNA networks associated with the susceptibility and persistence of atrial fibrillation through weighted gene co-expression network analysis // *Front. Genet.* 2021. V. 12.
<https://doi.org/10.3389/fgene.2021.653474>
 77. Esslinger U., Garnier S., Korniat A. et al. Exome-wide association study reveals novel susceptibility genes to sporadic dilated cardiomyopathy // *PLoS One.* 2017. V. 12. № 3.
<https://doi.org/10.1371/journal.pone.0172995>
 78. Ochoa J.P., Sabater-Molina M., García-Pinilla J.M. et al. Formin homology 2 domain containing 3 (*FHOD3*) is a genetic basis for hypertrophic cardiomyopathy // *J. Am. Coll. Cardiol.* 2018. V. 72. № 20. P. 2457–2467.
<https://doi.org/10.1016/j.jacc.2018.10.001>
 79. Wu G., Ruan J., Liu J. et al. Variant spectrum of formin homology 2 domain-containing 3 gene in Chinese patients with hypertrophic cardiomyopathy // *J. Am. Heart Assoc.* 2021. V. 10. № 5.
<https://doi.org/10.1161/JAHA.120.018236>
 80. Fernlund E., Kissopoulou A., Green H. et al. Hereditary hypertrophic cardiomyopathy in children and young adults – the value of reevaluating and expanding gene panel analyses // *Genes (Basel).* 2020. V. 11. № 12.
<https://doi.org/10.3390/genes11121472>
 81. Qin X., Li P., Qu H.Q. et al. *FLNC* and *MYLK2* gene mutations in a Chinese family with different phenotypes of cardiomyopathy // *Int. Heart J.* 2021. V. 62. № 1. P. 127–134.
<https://doi.org/10.1536/ihj.20-351>
 82. Brun F., Gigli M., Graw S.L. et al. *FLNC* truncations cause arrhythmogenic right ventricular cardiomyopathy // *J. Med. Genet.* 2020. V. 57. № 4. P. 254–257.
<https://doi.org/10.1136/jmedgenet-2019-106394>
 83. Xiao F., Wei Q., Wu B. et al. Clinical exome sequencing revealed that *FLNC* variants contribute to the early diagnosis of cardiomyopathies in infant patients // *Transl. Pediatr.* 2020. V. 9. № 1. P. 21–33.
<https://doi.org/10.21037/tp.2019.12.02>
 84. Phelan D.G., Anderson D.J., Howden S.E. et al. *ALPK3*-deficient cardiomyocytes generated from patient-derived induced pluripotent stem cells and mutant human embryonic stem cells display abnormal calcium handling and establish that *ALPK3* deficiency underlies familial cardiomyopathy // *Eur. Heart J.* 2016. V. 37. Iss. 33. P. 2586–2590.
<https://doi.org/10.1093/eurheartj/ehw160>
 85. Herkert J.C., Verhagen J.M.A., Yotti R. et al. Expanding the clinical and genetic spectrum of *ALPK3* variants: Phenotypes identified in pediatric cardiomyopathy patients and adults with heterozygous variants // *Am. Heart J.* 2020. V. 225. P. 108–119.
<https://doi.org/10.1016/j.ahj.2020.03.023>
 86. Lopes L.R., Garcia-Hernández S., Lorenzini M. et al. Alpha-protein kinase 3 (*ALPK3*) truncating variants are a cause of autosomal dominant hypertrophic car-

- diomyopathy // *Eur. Heart J.* 2021. V. 42. Iss. 32. P. 3063–3073.
<https://doi.org/10.1093/eurheartj/ehab424>
87. *Wooten E.C., Hebl V.B., Wolf M.J. et al.* Formin homology 2 domain containing 3 variants associated with hypertrophic cardiomyopathy // *Circ. Cardiovasc. Genet.* 2013. V. 6. № 1. P. 10–18.
<https://doi.org/10.1161/CIRCGENETICS.112.965277>
 88. *Kan-O M., Takeya R., Abe T. et al.* Mammalian formin Fhod3 plays an essential role in cardiogenesis by organizing myofibrillogenesis // *Biol. Open.* 2012. V. 1. № 9. P. 889–896.
<https://doi.org/10.1242/bio.20121370>
 89. *Fujimoto N., Kan-O M., Ushijima T. et al.* Transgenic expression of the formin protein Fhod3 selectively in the embryonic heart: Role of actin-binding activity of Fhod3 and its sarcomeric localization during myofibrillogenesis // *PLoS One.* 2016. V. 11. № 2.
<https://doi.org/10.1371/journal.pone.0148472>
 90. *Ushijima T., Fujimoto N., Matsuyama S. et al.* The actin-organizing formin protein Fhod3 is required for postnatal development and functional maintenance of the adult heart in mice // *J. Biol. Chem.* 2018. V. 293. Iss. 1. P. 148–162.
<https://doi.org/10.1074/jbc.M117.813931>
 91. *Matsuyama S., Kage Y., Fujimoto N. et al.* Interaction between cardiac myosin-binding protein C and formin Fhod3 // *Proc. Natl Acad. Sci. USA.* 2018. V. 115. Iss. 19. P. E4386–E4395.
<https://doi.org/10.1073/pnas.1716498115>
 92. *Arimura T., Takeya R., Ishikawa T. et al.* Dilated cardiomyopathy-associated *FHOD3* variant impairs the ability to induce activation of transcription factor serum response factor // *Circ. J.* 2013. V. 77. № 12. P. 2990–2996.
<https://doi.org/10.1253/circj.cj-13-0255>
 93. *Myasnikov R., Bukaeva A., Kulikova O. et al.* A case of severe left-ventricular noncompaction associated with splicing altering variant in the *FHOD3* gene // *Genes (Basel).* 2022. V. 13. № 2.
<https://doi.org/10.3390/genes13020309>
 94. *Antoku S., Wu W., Joseph L.C. et al.* ERK1/2 phosphorylation of FHOD connects signaling and nuclear positioning alternations in cardiac laminopathy // *Dev. Cell.* 2019. V. 51. № 5. P. 602–616.e12.
<https://doi.org/10.1016/j.devcel.2019.10.023>
 95. *Verdonschot J.A.J., Vanhoutte E.K., Claes G.R.F. et al.* A mutation update for the *FLNC* gene in myopathies and cardiomyopathies // *Hum. Mutat.* 2020. V. 41. Iss. 6. P. 1091–1111.
<https://doi.org/10.1002/humu.24004>
 96. *Ortiz-Genga M.F., Cuenca S., Dal Ferro M. et al.* Truncating *FLNC* mutations are associated with high-risk dilated and arrhythmogenic cardiomyopathies // *J. Am. Coll. Cardiol.* 2016. V. 68. № 22. P. 2440–2451.
<https://doi.org/10.1016/j.jacc.2016.09.927>
 97. *Begay R.L., Tharp C.A., Martin A. et al.* *FLNC* gene splice mutations cause dilated cardiomyopathy // *JACC Basic Transl. Sci.* 2016. V. 1. № 5. P. 344–359.
<https://doi.org/10.1016/j.jacbts.2016.05.004>
 98. *Hall C.L., Gurha P., Sabater-Molina M. et al.* RNA sequencing-based transcriptome profiling of cardiac tissue implicates novel putative disease mechanisms in *FLNC*-associated arrhythmogenic cardiomyopathy // *Int. J. Cardiol.* 2020. V. 302. P. 124–130.
<https://doi.org/10.1016/j.ijcard.2019.12.002>
 99. *Goli R., Li J., Brandimarto J. et al.* Genetic and phenotypic landscape of peripartum cardiomyopathy // *Circulation.* 2021. V. 143. Iss. 19. P. 1852–1862.
<https://doi.org/10.1161/CIRCULATIONAHA.120.052395>
 100. *Ito S., Asakura M., Liao Y. et al.* Identification of the *Mtus1* splice variant as a novel inhibitory factor against cardiac hypertrophy // *J. Am. Heart Assoc.* 2016. V. 5. № 7.
<https://doi.org/10.1161/JAHA.116.003521>
 101. *Parvatiyar M.S., Brownstein A.J., Kanashiro-Takeuchi R.M. et al.* Stabilization of the cardiac sarcolemma by sarcospan rescues DMD-associated cardiomyopathy // *JCI Insight.* 2019. V. 5. № 11.
<https://doi.org/10.1172/jci.insight.123855>
 102. *Stark K., Esslinger U.B., Reinhard W. et al.* Genetic association study identifies *HSPB7* as a risk gene for idiopathic dilated cardiomyopathy // *PLoS Genet.* 2010. V. 6. № 10.
<https://doi.org/10.1371/journal.pgen.1001167>
 103. *Стрельцова А.А., Гудкова А.Я., Полякова А.А. и др.* Полиморфный вариант rs1739843 гена белка теплового шока 7 (*HSPB7*) и его связь с вариантами клинического течения и исходами у пациентов с гипертрофической кардиомиопатией (результаты 10-летнего наблюдения) // *Рос. кардиол. журн.* 2019. № 10. С. 7–15.
<https://doi.org/10.15829/1560-4071-2019-10-7-15>
 104. *Clausen A.G., Vad O.B., Andersen J.H., Olesen M.S.* Loss-of-function variants in the *SYNPO2L* gene are associated with atrial fibrillation // *Front. Cardiovasc. Med.* 2021. V. 8. Article 650667.
<https://doi.org/10.3389/fcvm.2021.650667>
 105. *Hsu J., Gore-Panter S., Tchou G. et al.* Genetic control of left atrial gene expression yields insights into the genetic susceptibility for atrial fibrillation // *Circ. Genom. Precis. Med.* 2018. V. 11. № 3.
<https://doi.org/10.1161/CIRCGEN.118.002107>
 106. *Ma X., Mo C., Huang L. et al.* An robust rank aggregation and least absolute shrinkage and selection operator analysis of novel gene signatures in dilated cardiomyopathy // *Front. Cardiovasc. Med.* 2021. V. 8.
<https://doi.org/10.3389/fcvm.2021.747803>
 107. *Aragam K.G., Chaffin M., Levinson R.T. et al.* Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery // *Circulation.* 2019. V. 139. P. 489–501.
<https://doi.org/10.1161/CIRCULATIONAHA.118.035774>
 108. *Celeghin R., Cipriani A., Bariani R. et al.* Filamin-C variant-associated cardiomyopathy: A pooled analysis of individual patient data to evaluate the clinical profile and risk of sudden cardiac death // *Heart Rhythm.* 2022. V. 19. № 2. P. 235–243.
<https://doi.org/10.1016/j.hrthm.2021.09.029>
 109. *Finocchiaro G., Merlo M., Sheikh N. et al.* The electrocardiogram in the diagnosis and management of patients with dilated cardiomyopathy // *Eur. J. Heart Fail.* 2020. V. 22. Iss. 7. P. 1097–1107.
<https://doi.org/10.1002/ejhf.1815>
 110. *Lee Y., Choi B., Lee M.S. et al.* An artificial intelligence electrocardiogram analysis for detecting cardiomyopathy in the peripartum period // *Int. J. Cardiol.* 2022. V. 352. P. 72–77.
<https://doi.org/10.1016/j.ijcard.2022.01.064>
 111. *Yoneda Z.T., Anderson K.C., Quintana J.A. et al.* Early-onset atrial fibrillation and the prevalence of rare vari-

- ants in cardiomyopathy and arrhythmia genes // *JAMA Cardiol.* 2021. V. 6. № 12. P. 1371–1379. <https://doi.org/10.1001/jamacardio.2021.3370>
112. *Verweij N., Benjamins J.W., Morley M.P. et al.* The genetic makeup of the electrocardiogram // *Cell Syst.* 2020. V. 11. Iss. 3. P. 229–238.e5. <https://doi.org/10.1016/j.cels.2020.08.005>
113. *Weng L.C., Hall A.W., Choi S.H. et al.* Genetic determinants of electrocardiographic P-wave duration and relation to atrial fibrillation // *Circ. Genom. Precis. Med.* 2020. V. 13. № 5. P. 387–395. <https://doi.org/10.1161/CIRCGEN.119.002874>
114. *Liao W.C., Juo L.Y., Shih Y.L. et al.* HSPB7 prevents cardiac conduction system defect through maintaining intercalated disc integrity // *PLoS Genet.* 2017. V. 13. № 8. <https://doi.org/10.1371/journal.pgen.1006984>
115. *Ke L., Meijering R.A., Hoogstra-Berends F. et al.* HSPB1, HSPB6, HSPB7 and HSPB8 protect against RhoA GTPase-induced remodeling in tachypaced atrial myocytes // *PLoS One.* 2011. V. 6. № 6. <https://doi.org/10.1371/journal.pone.0020395>
116. *Adriaens M.E., Lodder E.M., Moreno-Moral A. et al.* Systems genetics approaches in rat identify novel genes and gene networks associated with cardiac conduction // *J. Am. Heart Assoc.* 2018. V. 7. № 21. <https://doi.org/10.1161/JAHA.118.009243>
117. *Pirruccello J.P., Bick A., Wang M. et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy // *Nat. Commun.* 2020. V. 11. № 1. Article 2254. <https://doi.org/10.1038/s41467-020-15823-7>
118. *Biddinger K.J., Jurgens S.J., Maamari D. et al.* Rare and common genetic variation underlying the risk of hypertrophic cardiomyopathy in a National Biobank // *JAMA Cardiol.* 2022. V. 7. № 7. P. 715–722. <https://doi.org/10.1001/jamacardio.2022.1061>
119. *Marian A.J.* Oligogenic cardiomyopathy // *J. Cardiovasc. Aging.* 2022. V. 2. Article 3. <https://doi.org/10.20517/jca.2021.27>
120. *Pourebahim K., Marian J.G., Tan Y. et al.* A combinatorial oligogenic basis for the phenotypic plasticity between late-onset dilated and arrhythmogenic cardiomyopathy in a single family // *J. Cardiovasc. Aging.* 2021. V. 1. Article 12. <https://doi.org/10.20517/jca.2021.15>
121. *Баулина Н.М., Киселёв И.С., Чумакова О.С., Фаворова О.О.* Гипертрофическая кардиомиопатия как олигогенное заболевание: аргументы транскриптомики // *Мол. биология.* 2020. Т. 54. № 6. С. 955–967. <https://doi.org/10.31857/S0026898420060026>
122. *Alimohamed M.Z., Johansson L.F., Posafalvi A. et al.* Diagnostic yield of targeted next generation sequencing in 2002 Dutch cardiomyopathy patients // *Int. J. Cardiol.* 2021. V. 332. P. 99–104. <https://doi.org/10.1016/j.ijcard.2021.02.069>

Pathogenetics of Cardiomyopathy

A. N. Kucher^a, A. A. Sleptcov^a, and M. S. Nazarenko^{a, *}

^a*Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, 634050 Russia*

**e-mail: maria.nazarenko@medgenetics.ru*

This review summarizes the current state of knowledge on the genetic factors of both primary or Mendelian cardiomyopathies (CMPs) and some of its secondary forms. Dozens of genes with pathogenic/probably pathogenic variants have been described for primary CMPs. In most cases, the spectrum of causal genetic variants is specific for different CMPs, but shared genes and variants are also discovered. On the one hand genetic causes of diseases have not been established for all cases of primary CMPs, but on the other hand pathogenic variants in Mendelian disease genes are also found for its secondary forms. The genetic component in the development of both primary and secondary CMPs was also established during genome-wide association studies (GWAS). Single nucleotide polymorphisms (SNPs) associated with both primary and secondary CMPs are in most cases specific for different types of disease and make a small contribution to an individual's overall risk. The link between some SNPs and electro- or echocardiogram features of the normal heart has been reported in the population. Most of the CMPs-associated SNPs are localized in non-coding regions of the genome, but they have a regulatory potential, acting in the heart as loci that affect the level of expression (eQTL), splicing (sQTL) or epigenetic modifications. It is noteworthy that the effects of the eQTL and sQTL genotypes in some cases are not equivalent for different anatomical regions of the heart. The phenotype and clinical presentation of CMPs in general can be determined by a wide range of rare pathogenic/probably pathogenic variants with a strong effect and common polymorphisms with a small effect and modified by epigenetic factors.

Keywords: cardiomyopathy, pathogenetics, genes of monogenic cardiomyopathies, genome-wide association studies.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЛИНИЙ Y-ХРОМОСОМЫ ДРЕВНИХ И СОВРЕМЕННЫХ САХА (ЯКУТОВ)

© 2023 г. С. А. Федорова^{1, 2, *}, В. Звенигорски³, А. Н. Алексеев^{1, 4}

¹Северо-Восточный федеральный университет им. М.К. Аммосова, Якутск, 677013 Россия

²Якутский научный центр комплексных медицинских проблем, Якутск, 677018 Россия

³Университет Страсбурга, Страсбург, 67081 Франция

⁴Институт гуманитарных исследований и проблем малочисленных народов Севера
Сибирского отделения Российской академии наук, Якутск, 677027 Россия

*e-mail: sardaanafedorova@mail.ru

Поступила в редакцию 19.09.2022 г.

После доработки 12.10.2022 г.

Принята к публикации 25.10.2022 г.

В обзоре представлены результаты палеогенетических исследований древнего населения Якутии эпохи средневековья. Рассмотрен спектр гаплогрупп и STR-гаплотипов Y-хромосомы 74 мужчин из якутских погребений XV–XIX вв. в сравнении с данными по современным этногеографическим группам якутов. Проведено сопоставление генетических реконструкций с историческими преданиями народа саха.

Ключевые слова: древняя ДНК, саха (якуты), Y-хромосома.

DOI: 10.31857/S001667582306005X, **EDN:** SSBTGC

Сравнение геномных последовательностей современных и древних людей играет центральную роль в понимании происхождения отдельных популяций, демографических тенденций и влияния на них различных заболеваний (процессов адаптации и давления естественного отбора). К настоящему времени наиболее успешно исследованы древние популяции Америки и Европы, в то время как по анализу ДНК древнего населения северо-востока Евразии известно небольшое количество публикаций, которые в основном касаются якутских погребений XV–XIX вв. [1–9].

Саха (якуты) – тюркоязычный народ Восточной Сибири, сумевший адаптироваться к суровым климатическим условиям Арктики традиционные для тюрков виды хозяйства – коневодство и разведение крупного рогатого скота. Прародина якутов по мнению большинства исследователей находится на территориях, прилегающих к оз. Байкал с запада. Начало миграции южных предков народа саха на север до сих пор является темой дебатов среди археологов. Одни авторы – сторонники относительно позднего переселения тюрков на Среднюю Лену начиная с XIII в. [10], другие склонны сдвинуть временные рамки появления скотоводческих племен на территории современной Якутии к X в. и даже ранее, вплоть до III–IV вв. н.э. [11, 12]. У якутов выделяют несколько этногеографических групп, различающихся территорией

расселения, особенностями языка и культуры: центральную (амгино-ленскую), вилюйскую и северную. В хозяйстве центральных и вилюйских якутов ярко представлен скотоводческий комплекс, у северных якутов помимо разведения лошадей и коров развито оленеводство.

Интенсивные археологические исследования древних погребений якутов с участием палеогенетиков начались с 2002 г. членами Саха-Французской экспедиции Якутского государственного университета им. М.К. Аммосова (с 2011 г. Северо-Восточный федеральный университет им. М.К. Аммосова). Объектом изучения являлись, с одной стороны, сопроводительный материал (артефакты) и, с другой стороны, человеческие и животные костные останки, погребенные в слое вечной мерзлоты. В результате этих исследований вначале были опубликованы статьи по генетическому анализу 11 останков из якутских погребений XVII и XVIII вв. [1–3], позднее проведен анализ останков 58 человек из могильников Центральной Якутии [4] и 130 человек из погребений Амгино-Ленского междуречья, Вилюя и Верхоянья [6]. Эти опубликованные данные наряду с новыми, полученными в ходе раскопок в последние годы [8], позволили осуществить анализ ДНК 74 мужчин из древней популяции якутов и провести междисциплинарное исследование палеогенетических, исторических и археологических данных с использовани-

Таблица 1. Частоты гаплогрупп Y-хромосомы в древних и современных субпопуляциях якутов

Гаплогруппа	Древние			Современные					
	ЦЯ ¹ n = 46	ВЯ ¹ n = 9	СЯ ¹ n = 19	ЦЯ ¹ n = 89	ЦЯ ² n = 92	ВЯ ¹ n = 77	ВЯ ² n = 58	СЯ ¹ n = 85	СЯ ² n = 66
N3(M1982)	91.3	88.9	94.7	92.1	85.9	85.7	93.1	82.4	71.2
Ht1	43.5	77.8	47.4	32.6	—	13.0	—	40.0	—
Ht3	2.2	11.1	0	4.5		19.5		3.5	
Ht2	10.9	0	0	30.3		7.8		0	
N2(CTS6380)	2.2	11.1	0	1.1	0	1.3	0	5.9	0
N2(B478)	—	—	—	—	2.2	—	0	—	10.6
C3*(M217)	4.3	0	0	3.4	0	2.6	0	2.4	3.0
C3*(F3985)	0	0	5.3	0	0	0	0	0	—
C3c(M48)	—	—	—	—	0	—	3.4	—	6.1
C3d(M407)	—	—	—	—	1.1	—	0	—	4.5
O(P49)	0	0	0	0	0	2.6	0	0	0
I1(M253)	0	0	0	0	2.2	1.3	1.7	0	0
I2a(L621)	0	0	0	1.1	0	2.6	0	1.1	0
J2a1(Z7671)	0	0	0	0	0	0	0	1.1	0
R1a	0	0	0	1.1	0	3.9	0	5.9	0
R1a1*(M198)	0	0	0	—	4.3	—	1.7	—	4.5
R1a1a7(M458)	0	0	0	—	2.2	—	0	—	0
R1b1b2(M269)	0	0	0	0	2.2	0	0	1.1	0
н.и.	2.2	0	0	1.1	0	0	0	0	0

Примечание. ЦЯ – центральные якуты, ВЯ – вилюйские якуты, СЯ – северные якуты (¹ [8], ² [13, 14]). Образцы древней ДНК были генотипированы по 17 STR-локусам набора Y-Filer AmpFLSTR, принадлежность STR-линий к гаплогруппам определена с использованием предиктора Nevgen Y-ДНК (www.nevgen.org). Гаплогруппы азиатского происхождения выделены серым цветом. Обозначения ветвей гаплогруппы N соответствуют филогенетической номенклатуре, использованной в работе [15]. н.и. – гаплогруппы не идентифицированы; “—” – гаплогруппы не тестировали по указанным маркерам.

ем различных подходов с целью понять эволюцию культурных норм и общественной структуры народа саха на протяжении последних пяти веков.

В данном обзоре описаны результаты исследований разнообразия линий Y-хромосомы у древних якутов [1–8] в сравнении с данными по современным популяциям [8, 13, 14].

ОСОБЕННОСТИ СТРУКТУРЫ ГЕНОФОНДА СОВРЕМЕННЫХ САХА (ЯКУТОВ) ПО ЛИНИЯМ Y-ХРОМОСОМЫ

Структура генофонда современных якутов к настоящему времени хорошо изучена. Отличительной и наиболее яркой особенностью якутского этноса является доминирование мужчин с гаплогруппой N3(M1982) (от 71 до 93% в различных этногеографических группах), среди которых

в результате эффекта основателя выделяются носители одного STR-гаплотипа (Ht1) [8, 13–17] (табл. 1). Помимо доминирующей N3(M1982) в мужском генофонде якутов были обнаружены в небольшом количестве (от 1 до 6%) и другие азиатские гаплогруппы: N2(B478) (характерна для популяций Западной и Южной Сибири), C3*(M217) и C3c(M48) (в большей степени распространены среди тунгусоязычных этносов) и C3d(M407) (обнаружена у народов Южной Сибири – бурятов, телеутов, тувинцев, сойотов, монголов, хамниган). Доля линий западноевразийского происхождения (R1a1*, R1a1a7, R1b1b2, J2a1 и I1) составляет от 3 до 11% в различных популяционных выборках (табл. 1).

Популяции якутов отличаются низкой генетической гетерогенностью: степень генетических различий между тремя этногеографическими груп-

Таблица 2. Мажорные гаплотипы Y-хромосомы в древней популяции якутов [8]

Гаплотип	DYS456	DYS389 I	DYS390	DYS389 II	DYS458	DYS19	DYS385a	DYS385b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Ht1	14	14	23	32	16	14	11	13	14	11	10	22	16	12	14	11	19
Ht2	14	14	23	31	16	14	11	13	14	11	10	22	15	12	14	11	19
Ht3	14	14	23	31	16	14	11	13	14	11	10	22	16	12	14	11	19

пами саха по Y-хромосоме составляет всего 4.1% [16]. Оценка генетических взаимоотношений показала более тесное генетическое родство центральных и вилюйских якутов и отдаленность от них северных. У северных якутов снижена частота гаплогруппы N3 (71–82%) и обнаружены более высокие частоты гаплогрупп, характерных для коренных малочисленных этносов Якутии – юкагиров, эвенов и эвенков (C3*, C3c) (табл. 1).

Из народов, проживающих в сопредельных с Якутией регионах, по своим генетическим характеристикам к якутам более близки народы Южной Сибири – прежде всего буряты и тувинцы, а также алтайцы, хакасы и монголы [14].

СПЕКТР ГАПЛОГРУПП Y-ХРОМОСОМЫ ДРЕВНИХ САХА

Степень изученности отдельных этнотерриториальных групп древних саха существенно различается: наибольшее количество генетических данных получено для центральных якутов Лено-Амгинского междуречья ($n = 46$); вилюйские ($n = 9$) и северные якуты ($n = 19$) исследованы в меньшей степени [8]. Тем не менее установлено, что во всех трех группах древних якутов наблюдается крайне низкий уровень разнообразия линий Y-хромосомы и необычайно высокая частота гаплогруппы N3 (от 89 до 95%) (табл. 1), что полностью соответствует данным по современным популяциям и указывает на стабильность генетических характеристик мужского генофонда якутов в течение последних столетий. Кроме N3, в древних популяциях центральных и вилюйских якутов были идентифицированы линии гаплогруппы N2(CTS6380), а у центральных и северных якутов – единичные линии гаплогруппы C3*(M217), которые встречаются во всех трех этнографических группах современных саха.

В выборке древних якутов не обнаружено ни одной линии западноевразийского происхождения, которые в современных популяциях представлены хотя и в небольшом количестве, но достаточно разнообразным спектром линий (табл. 1). Ранее нами было высказано предположение, что западноевразийские линии R1a1a7, R1b1b2 и I1 были привнесены в генофонд якутов русскоязычными этносами, заселившими территорию Якутии начи-

ная с XVII в., а линии J2a1 и R1a1*, широко распространенные на прародине народа саха в Южной Сибири, имеют более древнее происхождение, связанное с доэтническим этапом формирования якутов как отдельного народа [13, 14]. Полное отсутствие западноевразийских линий Y-хромосомы в исследованной выборке древних якутов либо может указывать на тотальное ограничение браков женщин-саха с европейцами вплоть до XIX в., либо может объясняться эффектом выборки – малочисленностью образцов или избирательностью мужских грунтовых захоронений, из которых был собран материал исследования, что, по нашему мнению, более вероятно.

ФИЛОГЕНЕТИЧЕСКИЙ АНАЛИЗ STR-ГАПЛОТИПОВ ДРЕВНИХ ЯКУТОВ

При типировании Y-хромосом древних якутов по 17 STR-маркерам в общей выборке ($n = 74$) идентифицирован 21 гаплотип, из которых доминировали два N3-гаплотипа, распространенных также в современных популяциях: Ht1, носителями которого являются большинство мужчин, и эволюционно более древний гаплотип Ht2, более характерный для центральных якутов (табл. 1, 2). Кроме того, в древней популяции был обнаружен гаплотип Ht3, в большей степени распространенный у вилюйских якутов (как древних, так и современных) [8]. Сравнительный анализ древних STR-гаплотипов с линиями современных якутов показал, что они либо идентичны, либо находятся в тесном родстве и различаются не более чем на три мутации.

При поиске сходных гаплотипов в базе данных, содержащей более 200000 как современных, так и древних Y-STR-гаплотипов, эволюционно близкие к якутским N3-гаплотипы были найдены у древних тюрков и монголов и у современных бурят [8]. N2-гаплотип, общий для двух индивидов из Центральной Якутии и Вилюйского региона, захороненных до XVII в., обнаружен у современных хакасов, древних тюрков и монголов. Гаплотип C*(M217) двух недатированных индивидов, похороненных в Центральной Якутии, был найден у современных бурят и более чем у 100 современных монголов, и не обнаружен в базе данных

Таблица 3. Частоты гаплогрупп Y-хромосомы в древней популяции якутов по временным периодам [8]

Гаплогруппа	до 1700 г <i>n</i> = 21	XVIII в. <i>n</i> = 32	XIX в. <i>n</i> = 17	XXI в. <i>n</i> = 251
N3(TAT)	85.7	100	100	86.9
Ht1	33.3	65.6	47.1	29.1
Ht2	0	9.4	11.8	13.1
Ht3	9.5	0	0	8.8
другие	42.9	25.0	41.1	35.9
N2(CTS6380)	9.5	0	0	2.8
C3(M217)	0	0	0	2.8
C3*(F3985)	4.8	0	0	0
O(P49)	0	0	0	0.8
I1(M253)	0	0	0	0.4
I2a(L621)	0	0	0	1.6
J2a1(Z7671)	0	0	0	0.4
R1a	0	0	0	3.6
R1b1b2(M269)	0	0	0	0.4
н.и.	0	0	0	0.4

Примечание. н.и. – гаплогруппы не идентифицированы.

по древним гаплотипам. Гаплотип мужчины, похороненного в Верхоянске до XVII в., C3*(F3985) не был обнаружен ни в одной современной популяции, но филогенетически близкие линии найдены у шести древних индивидов из европейской степи, что, возможно, указывает на алтайское происхождение этой вымершей линии.

Таким образом, данные филогенетического анализа древних STR-гаплотипов полностью коррелируют с результатами исследований генома современных популяций, которые указывают на южносибирское происхождение якутов к западу от Байкала в Восточно-Саянском регионе и на генетическую близость саха к народам Южной Сибири [14, 17].

ИЗБИРАТЕЛЬНОСТЬ МУЖСКИХ ГРУНТОВЫХ ЗАХОРОНЕНИЙ ЯКУТОВ В XVIII–XIX вв.

В табл. 3 представлены частоты гаплогрупп Y-хромосомы в древней популяции якутов, разделенной по временным периодам: 21 образец датирован временем до 1700 г., 32 образца относятся к XVIII в., 17 образцов – к XIX в., четыре – не датированы [8].

Распределение гаплотипов по периодам (до 1700 г., 1700–1800 гг., 1800–1900 гг. и после 2000 г.) показало доминирование гаплотипа Ht1 в любом из указанных периодов времени. В образцах XVIII в.

Ht1 присутствует в абсолютном большинстве (>60%) археологических образцов. Линия Ht2, более характерная для центральных якутов, не обнаружена в погребениях до XVIII в., но впоследствии становится вторым по распространенности гаплотипом у мужчин-саха. Линия Ht3, распространенная в большей степени у вилюйских якутов, обнаружена в двух образцах, датированных временем до 1700 г., но отсутствует в погребениях XVIII–XIX вв., хотя в современной популяции она является третьей по распространенности.

Разнообразие STR-гаплотипов в XVIII–XIX вв. по сравнению с более древним (до 1700 г.) и современным периодами оказалось значительно ниже ожидаемого, в могилах XVIII–XIX вв. были обнаружены в основном носители Ht1 и Ht2 (табл. 3). Сниженный уровень разнообразия STR-гаплотипов указывает на то, что мужские захоронения были избирательными и это были главным образом представители доминирующей линии Ht1. Проведенные палеогенетиками тесты показали, что гипотетическое разнообразие гаплотипов в XVIII–XIX вв. должно было быть значительно выше наблюдаемых значений, следовательно доминирование линии Ht1 (и, возможно, Ht2) в археологическом материале является аномальным и не может быть результатом случайных захоронений в демографической истории этноса [8].

Следует отметить, что якутские погребения до конца XVII в. – это не традиционные кладбища, где обычно хоронят умерших, такая практика фактически не существовала в средневековой Якутии, где рытье могил для захоронения – сложная и трудоемкая задача из-за вечной мерзлоты. Погребальные памятники представлены одиночными грунтовыми могилами, которые располагаются на расстоянии десятков километров друг от друга. В силу объективных трудностей наиболее распространенным способом захоронения вплоть до массовой христианизации, т.е. до середины XVIII в., было так называемое воздушное захоронение – помещение тела в гроб, установленный на двух или четырех столбах (арангас). Тела покойников в таких арангасах обычно полностью разлагались всего за несколько лет. До конца XVII в. существовал также обычай кремации покойников [18]. В научной литературе высказывались противоречивые точки зрения о происхождении грунтового способа захоронения. Многие исследователи считали его навязанным русской администрацией в период христианизации якутов. Однако серия грунтовых погребений, датируемых ранее XVII в., дает основание отнести их к местным практикам захоронения [19].

Избирательность мужских грунтовых захоронений в XVIII–XIX вв. возможно свидетельствует о культурном влиянии русских на некоторые якутские роды после вхождения территории Яку-

тии в состав Российской империи в XVII в. [8]. В Центральной Якутии в этот период хоронили носителей линий Ht1 и Ht2. Линия Ht3 до прихода русских уже присутствовала, но ни один из ее носителей не был обнаружен в погребениях периода 1700–1900 гг., даже в Вилюйском регионе, где эта линия доминирует в настоящее время. Представителей линии Ht3, по-видимому, не хоронили в период господства линии Ht1 либо потому, что ее носители меньше сталкивались с русской культурой, либо потому, что им было отказано в привилегии быть захороненными, либо они предпочитали быть похороненными другим способом.

ГЕНЕТИЧЕСКИЕ ДАННЫЕ В СРАВНЕНИИ С ИСТОРИЧЕСКИМИ ПРЕДАНИЯМИ НАРОДА САХА

Согласно историческим преданиям народа саха преобладающая часть крупных якутских родов (хангаласцы, мегинцы, батурсцы, борогонцы и др.) произошли от прародителя Эллэя [20, 21]. Эллэй Боотур этнически позиционируется в преданиях как сын татарского царя Татаар Таймы, который под напором врагов бежал с юга на Среднюю Лену. В долине Туймаада он встретил Омогой Баая – владельца многочисленной челяди и огромного количества коров и лошадей. От Эллэя и старшей дочери Омогой Баая родились множество сыновей (от 4 до 11 согласно разным вариантам легенды), от которых произошли большинство якутских родов [21]. По мнению этнографа Г.В. Ксенофонтова описанная в преданиях история народа саха может быть сравнима с историей еврейского народа, который согласно Ветхому Завету произошел от одного прародителя (Иакова, внука Авраама) и подразделялся по мужской линии на 12 колен Израилевых [22]. От Омогой Баая согласно некоторым единичным легендам происходят батулинцы, намцы, баягантайцы [20, 21, 23, 24]. Кроме того, часть родов народа саха (хоринцы или хоролоры) связывают с именем третьего прародителя Улуу Хоро, явившегося в Якутию с низовьев Амура верхом на быке [23, 24].

В ранних исследованиях нами было отмечено соответствие генетических данных историческим преданиям народа саха: выраженный эффект N3-основателя у современных якутов указывает на то, что в предковой популяции преобладали мужчины одного рода, прародителем которого мог быть человек, относящийся к верхушке социальной иерархии, у которого было много сыновей [13]. С другой стороны, демографические модели показывают, что доминирование какого-либо гаплотипа в популяции может быть не связано с какой-либо исторической личностью, имевшей социальные и культурные привилегии, но скорее является результатом естественного отбора [25].

В случае якутской популяции было проведено генетическое исследование с целью определить действительно ли существование конкретных исторических и/или описанных в преданиях якутских правителей может быть связано с событиями последних столетий в генетической истории этноса. В работе В. Звенигорски и соавт. [8] описан редкий пример исследования древней ДНК для установления личности отдельных индивидов и их роли в эволюции культуры и системы власти в Якутии в недавнем прошлом. В генетических исследованиях останков знатных людей из богатых захоронений в Центральной Якутии было доказано, что род Мазары Бозекова, предводителя хангаласцев, имел гаплотип Ht1. Известно, что Мазары Бозеков был внуком Тыгын Дархана, правителя якутов в конце XVI–первой трети XVII в., который в свою очередь, согласно преданиям, был прямым потомком Эллей Боотура. Таким образом, результаты генетического анализа подтвердили существование связи между линией Тыгын Дархана/Эллей Боотура и гаплотипом Ht1 [8].

У центральных якутов Лено-Амгинского междуречья (как древних, так и современных) доминируют две мужские линии: в большей степени Ht1 (гаплотип Тыгын Дархана/Эллей Боотура) и в меньшей – Ht2 (табл. 1). У северных якутов Ht1 имеет наибольшие частоты (47% в древней популяции, 40% в современной), тогда как Ht2 распространен с низкой частотой. Третий гаплотип Ht3 преобладает у вилюйских якутов. Поскольку в якутской этнографии предполагается существование трех якутских предков-основателей (Эллей, Омогой и Улуу Хоро), было высказано предположение, что эволюционно более древний гаплотип Ht2, распространенный в Центральной Якутии, может быть связан с Омогоем, а “вилюйский” гаплотип Ht3 – с Улуу Хоро, этнонимы которого часто встречаются у западных якутов [8]. По преданиям Улуу Хоро прибыл вначале в Центральную Якутию, откуда его потомки ушли на запад под натиском хангаласского вождя Тыгын Дархана и основали свои наследи в Верхневилуйском и Сунтарском улусах [26]. Однако гипотеза о связи Ht3 с потомками Улуу Хоро является дискуссионной и подлежит проверке по мере накопления новых генетических сведений. Распространение и доминирование гаплотипа Ht1 в XVIII–XIX вв., по мнению В. Звенигорски и соавт., является непосредственным результатом консолидации власти в Якутии вокруг рода хангаласцев, проживавших на довольно ограниченной территории в Центральной Якутии и на Вилюе в XVII в., но уже спустя продвинувшихся до далеких северо-восточных районов.

Следует отметить, что в случае Омогоя и Улуу Хоро прямых доказательств их связей с линиями Ht2 и Ht3 не получено и гипотеза В. Звенигорски основана на географическом распределении

могил и современных гаплотипов. По его мнению, Омогой мог быть легендарным прародителем мужских кланов, имеющих гаплотип Ht2, побежденных кланом Тыгына еще до прихода русских. Улуу Хоро мог представлять собой главу западного клана или племени (с гаплотипом Ht3), также находившегося в конфликте с Тыгыном Дарханом, запретившим их захоронения во время господства его сыновей и внуков [8]. С точки зрения Э. Крюбези, эти генетические реконструкции совместимы с некоторыми легендами о том, что Тыгын Дархан был потомком самого Эллей Боотура и что он враждовал с потомками Омогоя (намцами, баягантайцами и хоринцами, включая западные вилюйские кланы) [8, 27].

ЗАКЛЮЧЕНИЕ

Сравнительный ДНК-анализ древней и современной якутской популяции показал преэмульсивность линий Y-хромосомы и относительную стабильность характеристик мужского генофонда якутской популяции в течение последних веков. STR-гаплотипы древних и современных мужчин саха оказались либо идентичны, либо находились в тесном родстве. Филогенетически близкие к якутским STR-гаплотипы найдены у древних тюрков и монголов, современных бурят, монголов и хакасов, что указывает на тесные генетические связи саха с народами Южной Сибири по отцовским линиям.

Палеогенетическое исследование древней популяции якутов позволило установить, что в XVIII–XIX вв. выбор мужчин для грунтовых захоронений в значительной степени был неслучаен и определялся общественными предпочтениями. При детальном анализе N3-линий Y-хромосомы в древней и современной популяциях якутов были выявлены три доминирующих гаплотипа: Ht1, носителями которого является большинство мужчин; Ht2, распространенный в центральной части Якутии; и Ht3, преобладающий на западе Якутии в Вилюйском регионе. Распределение гаплотипов по четырем периодам (до 1700 г., 1700–1800 гг., 1800–1900 гг. и после 2000 г.) показывает доминирование гаплотипа Ht1 во все изученные периоды времени. Результаты исследований подтверждают существование связи между линией могущественного вождя якутов Тыгын Дархана, потомка Эллей Боотура, и гаплотипом Ht1. По данным ДНК-анализа установлено, что члены других крупных древних кланов, предположительно связанные с именами двух других легендарных прародителей народа саха – Омогоя (Ht2) и Улуу Хоро (Ht3), реже подвергались грунтовым захоронениям, хотя проживали в тех же регионах.

Таким образом, совместные исследования археологов и генетиков позволили частично реконструировать генетическую историю народа саха в

Восточной Сибири. На ее обширном пространстве еще остается множество необнаруженных захоронений, а новые геномные подходы несомненно предоставят больше доказательств и помогут лучше понять генетический аспект заселения Якутии в древности.

Работа выполнена в рамках Государственного задания Министерства науки и высшего образования РФ (FSRG-2023-0003) “Генетические особенности населения Северо-Востока России: реконструкция генетической истории, механизмы адаптации и старения, возраст-зависимые и наследственные заболевания”, НИР ЯНЦ КМП “Изучение генетической структуры и груза наследственной патологии в популяциях Республики Саха (Якутия)”.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Ricaut F.-X., Kolodesnikov S., Keyser-Tracqui C. et al. Genetic analysis of human remains found in two eighteenth century Yakut graves at At-Dabaan // *Int. J. Legal Med.* 2004. V. 118. P. 24–31. <https://doi.org/10.1007/s00414-003-0411-6>
2. Ricaut F.-X., Kolodesnikov S., Keyser-Tracqui C. et al. Molecular genetic analysis of 400-year-old human remains found in two Yakut burial sites // *Am. J. Phys. Anthr.* 2006. V. 129. P. 55–63. <https://doi.org/10.1002/ajpa.20195>
3. Федорова С.А., Степанов А.Д., Адоаян М. и др. Анализ линий древней митохондриальной ДНК в Якутии // *Мол. биология.* 2008. Т. 42. № 3. С. 445–453.
4. Crubézy E., Amory S., Keyser C. et al. Human evolution in Siberia: From frozen bodies to ancient DNA // *BMC Evol. Biol.* 2010. V. 10: 25. <https://doi.org/10.1186/1471-2148-10-25>
5. Thèves C., Balaesque P., Evdokimova L.E. et al. Population genetics of 17 Y-chromosomal STR loci in Yakutia // *Forensic Sci. Int. Genet.* 2010. V. 4. e129–e130. <https://doi.org/10.1016/j.fsigen.2010.01.018>
6. Keyser C., Hollard C., Gonzalez A. et al. The ancient Yakuts: A population genetic enigma // *Phil. Trans. R. Soc. B.* 2015. V. 370: 20130385. <https://doi.org/10.1098/rstb.2013.0385>
7. Zvenigorosky V., Crubézy E., Gibert M. et al. The genetics of kinship in remote human groups // *Forensic Sci. Int. Genet.* 2016. V. 25. P. 52–62. <https://doi.org/10.1016/j.fsigen.2016.07.018>
8. Zvenigorosky V., Duchesne S., Romanova L. et al. The genetic legacy of legendary and historical Siberian chieftains // *Communication Biology.* 2020. V. 3: 581. <https://doi.org/10.1038/s42003-020-01307-3>
9. Zvenigorosky V., Sabbagh A., Gonzalez A. et al. The limitations of kinship determinations using STR data in ill-defined populations // *Intern. J. Legal Med.* 2020.

- V. 134. P. 1981–1990.
<https://doi.org/10.1007/s00414-020-02298-w>
10. *Гоголев А.И.* Якуты: проблемы этногенеза и формирования культуры. Якутск: Национальное изд-во РС(Я), 1993. 136 с.
 11. *Алексеев А.Н.* Древняя Якутия. Железный век и эпоха средневековья. Новосибирск: Изд-во ИАЭТ СО РАН, 1996. 95 с.
 12. *Алексеев А.Н., Бравина Р.И.* Формирование якутского народа // История Якутии. Т. 1. Новосибирск: Наука, 2020. С. 223–230.
 13. *Федорова С.А.* Генетические портреты народов Республики Саха (Якутия): анализ линий митохондриальной ДНК и Y-хромосомы. Якутск: Изд-во ЯНЦ СО РАН, 2008. 235 с.
 14. *Fedorova S.A., Reidla M., Metspalu E. et al.* Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): Implications for the peopling of Northeast Eurasia // BMC Evolutionary Biol. 2013. V. 13: 127.
<https://doi.org/10.1186/1471-2148-13-127>
 15. *Iltmäe A.M., Reidla M., Chukhryaeva M. et al.* Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families // Am. J. Hum. Genet. 2016. V. 99. P. 163–173.
<https://doi.org/10.1016/j.ajhg.2016.05.025>
 16. *Федорова С.А., Барашков Н.А., Ушницкий В.В. и др.* Этнотерриториальные группы якутов: особенности генетической структуры // Якутский мед. журн. 2014. № 2. С. 28–31.
 17. *Pugach I., Matveev R., Spitsyn V. et al.* The complex admixture history and recent southern origins of siberian populations // Mol. Biol. Evol. 2016. V. 33(7). P. 1777–1795.
<https://doi.org/10.1093/molbev/msw055>
 18. Якуты. Саха. Серия Народы и культуры. М.: Наука, 2013. 599 с.
 19. *Бравина Р.И., Дьяконов В.М.* Раннеякутские средневековые погребения XIV–XVII вв.: совокупность отличительных признаков // Северо-Вост. гуманитарный вестник. 2015. № 3. С. 27–31.
 20. *Эргис Г.У.* Исторические предания и рассказы якутов: в 2 частях. М.; Л.: Изд-во АН СССР, 1960. Ч. 1. 322 с.
 21. *Ксенофонтов Г.В.* Элэйада: материалы по мифологии и легендарной истории якутов. Новосибирск: Наука, 1977. 245 с.
 22. *Ксенофонтов Г.В.* Ураанхай-сахалар: Очерки по древней истории якутов. Иркутск: ОГИЗ Восточносиб. областное изд-во, 1937. Т. I. 576 с.
 23. *Алексеев Н.А., Емельянов Н.В., Петров В.Т.* Предания, легенды и мифы саха (якутов). Серия “Памятники фольклора народов Сибири и Дальнего Востока”. Новосибирск: Наука, 1995. 400 с.
 24. *Ушницкий В.В.* Легенды о прародителях якутского этноса как исторический источник // Монголоведение. 2014. № 7. С. 81–88.
 25. *Guillot E.G., Cox M.P.* High frequency haplotypes are expected events, not historical figures // F1000Res. 2015. V. 4: 666.
<https://doi.org/10.12688/f1000research.7023.2>
 26. *Зориктуев Б.Р.* Якутские хоролоры – кто они? (К проблеме этнической идентификации) // Археология, этнография и антропология Евразии. 2011. № 2(46). С. 119–127.
 27. *Боло С.И.* Прошлое якутов до прихода русских на Лену: по преданиям якутов бывшего Якутского округа. Якутск: Якутгосиздат, 1938. 227 с.

The Y-Chromosome Lineage Variation in Ancient and Modern Populations of the Sakha (Yakuts)

S. A. Fedorova^{a, b, *}, V. Zvenigorosky^c, and A. N. Alekseev^{a, d}

^a*Ammosov North East Federal University, Yakutsk, 677013 Russia*

^b*Yakut Science Center of Complex Medical Problems, Yakutsk, 677018 Russia*

^c*University of Strasbourg, Strasbourg, 67081 France*

^d*Institute for Humanities Research and Indigenous Studies of the North, Siberian Branch of the Russian Academy of Sciences, Yakutsk, 677027 Russia*

*e-mail: sardaanafedorova@mail.ru

This review presents the results of paleogenetic studies on the ancient population of Yakutia in the Middle Ages. The diversity of haplogroups and STR haplotypes of the Y chromosome of 74 men from Yakut burials from the 15th–19th centuries was considered and compared with the data on modern ethnogeographical groups of the Yakuts. A comparison of genetic reconstructions with historical data and the legends of the Sakha people was performed.

Keywords: ancient DNA, Sakha (Yakuts), Y chromosome.

УДК 004.655.3(075.8)

ИЗУЧЕНИЕ ГЕНЕТИЧЕСКОГО РАЗНООБРАЗИЯ ВОЗБУДИТЕЛЯ ПИРИКУЛЯРИОЗА РИСА *Pyricularia oryzae* Cav. ПО КОМПЛЕКСУ МОЛЕКУЛЯРНЫХ И МОРФОЛОГИЧЕСКИХ ПРИЗНАКОВ

© 2023 г. Д. В. Нартымов¹, Е. В. Дубина^{1, 2, *}, М. Г. Рубан¹, Ю. В. Анискина³, С. В. Гаркуша^{1, 2}, И. А. Шилов³, Н. С. Велишаева³, О. С. Колобова³, Н. К. Истомин¹

¹Федеральный научный центр риса, Краснодар, пос. Белозерный, 350921 Россия

²Кубанский государственный аграрный университет им. И.Т. Трубилина, Краснодар, 350044 Россия

³Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва, 127550 Россия

*e-mail: lenakrug1@rambler.ru

Поступила в редакцию 28.09.2022 г.

После доработки 29.11.2022 г.

Принята к публикации 16.12.2022 г.

На основе молекулярно-генетических подходов проведено изучение биоразнообразия фитопатогенного гриба *Pyricularia oryzae* Cav. на юге России и проведена классификация морфолого-культуральных и генетических признаков патогена. За 2020–2022 гг. проведен мониторинг и выделено 24 штамма патогена из пораженного гербарного материала, собранного с полей пяти экологических рисосеющих зон Краснодарского края (Красноармейского, Калининского, Абинского, Славянского районов, Краснодар). С использованием мультиплексной системы ПЦР на основе фрагментного анализа среди изученных штаммов *Pyricularia oryzae* Cav. за весь период выполнения проекта выявлено пять генотипов, каждый из которых характеризуется уникальным генетическим профилем. Составлены их “ДНК-паспорта”. Изученные изоляты возбудителя пирикуляриоза охарактеризованы по морфолого-культуральным признакам. Выделено шесть морфотипов патогена.

Ключевые слова: генетическое биоразнообразие, фитопатогенные грибы, пирикуляриоз (*Pyricularia oryzae* Cav.), патотип, штаммы, SSR-локусы, ПЦР, ДНК-паспорта.

DOI: 10.31857/S0016675823060127, **EDN:** SUNFYL

Пирикуляриоз риса остается одной из самых проблематичных особенностей в технологии выращивания риса на территории Краснодарского края. Широкое применение фунгицидов, направленная селекция и предварительная подготовка почвы во многом помогают сельхозтоваропроизводителям снизить вероятность возникновения патогенеза. Однако изменчивость патогена позволяет ему достаточно быстро приспособиться к таким условиям и развиваться на растениях риса.

Для проведения направленных работ по селекции риса необходимо учитывать тенденции развития признаков патогена *Pyricularia oryzae* Cav., его генотипа и морфологических признаков, позволяющих ему, в том числе, поражать различные части растения. С целью мониторинга динамики развития штаммов на территории юга России на базе лаборатории информационных цифровых и биотехнологий ФГБНУ ФНЦ риса была создана база данных “Биоразнообразие штаммов фитопатогенного гриба *Pyricularia oryzae* Cav. для рисосеющих хозяйств юга России” [1]. База данных позволяет систематизировать данные по генетиче-

ским и морфолого-культуральным признакам штаммов пирикуляриоза риса, полученные в результате проведения научно-исследовательских работ.

Систематизация данных по признакам штаммов пирикуляриоза риса позволит выявить закономерности распределения популяций патогена на территории рисосеяния и установить взаимосвязь между популяциями, выявленными в результате проведения исследований. Определение родственного сходства генотипов и морфолого-культуральных признаков позволит достоверно установить обусловленность локализации популяций на территории рисосеяния Краснодарского края. Выявление таких связей создаст условия для имплементации данных в модели селекционных работ в рисоводстве.

МАТЕРИАЛЫ И МЕТОДЫ

Штаммы патогена *Pyricularia oryzae* Cav. выделяли из пораженных тканей риса (узлов стебля, листьев, частей метелки), собранных с рисовых

полей, в соответствии с методическими указаниями, разработанными во ВНИИ фитопатологии [2]. Образцы пораженной ткани промывали стерильной водой и размещали на стерильную морковно-сахарную агаровую среду в чашки Петри, которые затем устанавливали во влажную (100%) камеру при температуре $27 \pm 1^\circ\text{C}$, оптимальной для образования конидий гриба *P. oryzae*. Нарастивание мицелия для выделения ДНК проводили в чашках Петри на жидкой морковной агаризированной среде. Полученную культуру культивировали до необходимой степени чистоты.

При регистрации отбираемых штаммов патогена фиксировали зону [3] и район отбора гербарного материала, а также сорт и часть растения, пораженного пирикулярриозом.

После получения культуры проводили описание фенотипических свойств в соответствии с классификацией морфолого-культуральных признаков штаммов *P. oryzae* [4].

Генотипирование штаммов возбудителя пирикулярриоза проводили в лабораториях информационных, цифровых и биотехнологий ФГБНУ ФНЦ риса и анализа генома ФГБНУ ВНИИСБ молекулярно-генетическими методами. Грибную ДНК выделяли из смеси мицелия и конидий моноконидиальной культуры гриба *P. oryzae* методом СТАВ (Murray and Thompson, 1980). Метод заключается в использовании гексадецилтриметиламмония бромида в качестве детергента в лизирующем буфере. Состав лизирующего СТАВ-буфера: 0.2 М Tris-HCl, pH 8; 2 М NaCl; 0.05 М EDTA; 2% СТАВ.

При достижении необходимой биомассы при культивировании мицелий снимали с чашек, растирали его в 700 мкл лизирующего СТАВ-буфера в стерильной керамической ступке. Затем полученную смесь переносили в пробирку Eppendorf и инкубировали при 65°C в течение часа. К полученной смеси добавляли 500 мкл хлороформа и центрифугировали 10 мин при 13000 об./мин. После с помощью пипетмана на 1000 мкл отбирали 500 мкл надосадочной жидкости, добавляли к ней еще 500 мкл хлороформа и повторяли центрифугирование. К полученному супернатанту добавляли 400 мкл изопропанола и ставили в морозильную камеру на 2 ч. После центрифугировали 10 мин при 13000 об./мин. Выпавший осадок ДНК промывали охлажденным 70%-ным этанолом и ресуспендировали в 50 мкл буфера TE для хранения при температуре -20°C .

Аmplификацию ДНК фитопатогенного гриба *P. oryzae* проводили в объеме 25 мкл, содержащем однократный ПЦР-буфер, 1.25 ед. *Taq* ДНК-полимеразы, 100 мкмоль dNTP (ООО НПО “ДНК-Технология”, Россия), по 10 пкмоль каждого праймера и по 5 мкл раствора ДНК в детектирующем термоциклере (CFX-BioRad, США) по

программе: $95^\circ\text{C} - 10$ мин; 30 циклов: $94^\circ\text{C} - 30$ с, $55^\circ\text{C} - 30$ с, $72^\circ\text{C} - 30$ с.

Визуализация продуктов ПЦР выполняется на автоматическом генетическом анализаторе НАНОФОР-05 (Институт аналитического приборостроения РАН, Россия) при поддержке центра коллективного пользования “Биотехнология” ФГБНУ ВНИИСБ.

Обработку данных выполняли с использованием программного обеспечения ДНК ФА (ИАП, Россия).

Набор данных, полученных при генотипировании и изучении морфолого-культуральных признаков, был перенесен в базу данных “Биоразнообразие штаммов фитопатогенного гриба *P. oryzae* для рисосеющих хозяйств юга России” [3], согласно установленной структуре. Таким образом, устанавливалась явная связь между особенностями образца и его изучаемыми свойствами, а систематизированные данные позволяли проводить глубокий статистический анализ.

Анализ данных представлял собой выявление совокупности однородных признаков штаммов по генотипу и фенотипу культуры путем сопоставления присутствия отдельного признака. Близость штаммов описывается иерархической кластеризацией по методу Ward [5], а полученная дендрограмма выявляет родственные связи. Иерархическую кластеризацию по методу Ward проводили при помощи алгоритмов модуля SciPy 1.8.0 Python 3.10.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

За период проведения исследовательской работы было получено 9 штаммов в 2020 г. и 15 штаммов *Pyricularia oryzae* Cav. в 2021 г. (табл. 1).

При наращивании культуры были получены 24 образца, которые проявляли повторяемые признаки развития колонии. Полученные колонии были отобраны для описания морфолого-культуральных признаков патогена и генотипирования его штаммов (рис. 1, 2).

Результаты генотипирования и изучения морфолого-культуральных признаков приведены в табл. 2.

Важно отметить наличие большого количества образцов с зонированным характером роста, а в культурах штаммов 2021 г. преобладание этого характера роста выше, чем в культурах штаммов 2020 г.

Также в каждой временной группе образцов присутствуют штаммы с уникальными морфолого-культуральными признаками. Так, в 2020 г. выявлен штамм 1.20 с колонией разнонаправленного характера роста клочковатой структуры и коническим рельефом низкого профиля, а в 2021 г. — штамм 11.21 с колонией равнонаправленного ха-

Таблица 1. Штаммы *Pyricularia oryzae* Cav.

Номер штамма	Зона	Район	Сорт	Часть
2020 г.				
1.20	В	Красноармейский район	Чайка	Лист
5.20	В	»	Фаворит	»
7.20	В	»	Полевик	»
10.20	В	»	Рапан	»
11.20	В	Калининский район	Исток	»
12.20	В	Красноармейский район	»	»
13.20	В	»	»	»
16.20	А	Славянский район	Рапан	Ножка метелки
20.20	В	г. Краснодар	КП-20	Узел
2021 г.				
1.21	С	Абинский район	Рапан	Лист
3.21	В	Красноармейский район	Фаворит	»
5.21	А	Славянский район	Регул	»
6.21	А	»	Рапан	»
7.21	А	»	Полевик	»
8.21	С	Абинский район	Рапан	»
9.21	В	Красноармейский район	Рапан 2	»
11.21	В	»	Исток	»
12.21	В	»	Полевик	»
13.21	В	»	Патриот	»
14.21	В	»	Рапан	»
15.21	В	»	Патриот	»
17.21	С	Северский район	Фаворит	»
18.21	С	Тахтамукайский район	КП 34-16	»
19.21	А	Славянский район	Сонет	Узел

рактера роста однородной структуры и равномерным рельефом низкого профиля.

Данные внесены в базу данных “Биоразнообразие штаммов фитопатогенного гриба *Pyricularia oryzae* Cav. для рисосеющих хозяйств юга России” и инициирована процедура обновления базы данных. Внесена информация о 24 штаммах пирикуляриоза риса, выявленных в 2020 и 2021 гг., и связанных с ними генотипах и фенотипах.

Иерархический кластерный анализ данных образцов изучаемых штаммов был проведен отдельно для штаммов, выделенных в разные годы, для выявления миграции и временной переходности генетического материала между зонами рисосеяния юга России.

Результаты кластеризации представлены на рис. 3 и 4.



Рис. 1. Культуры штаммов *Pyricularia oryzae* Cav., полученных в 2020 г.

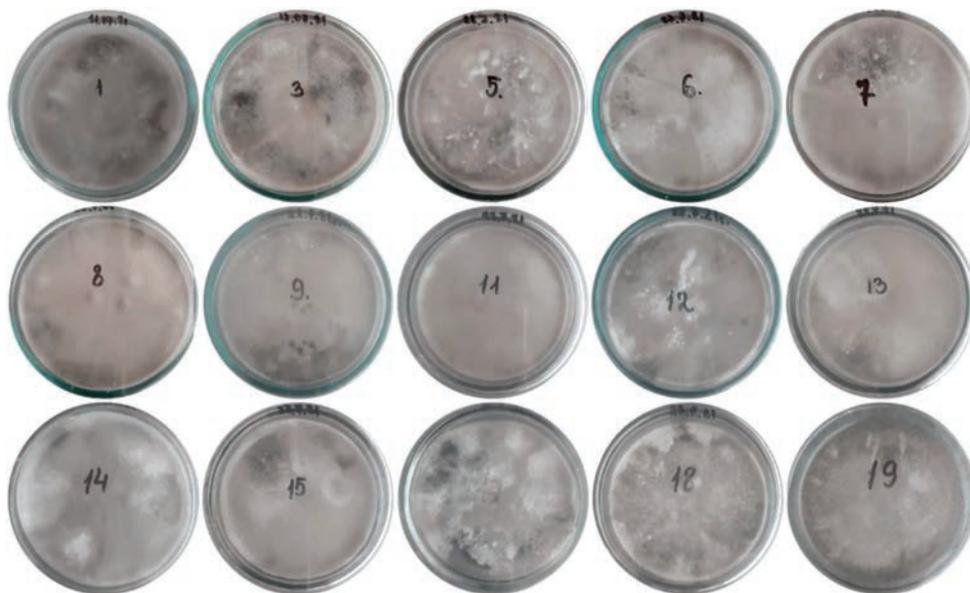


Рис. 2. Культуры штаммов *Pyricularia oryzae* Cav., полученных в 2021 г.

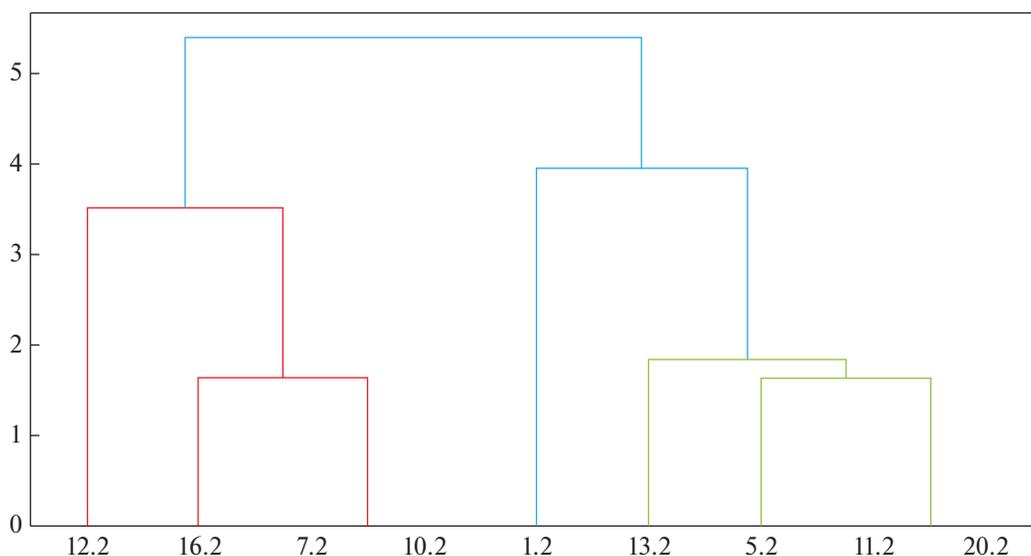


Рис. 3. Дендрограмма кластеризации штаммов, выделенных в 2020 г.

Из дендрограммы рис. 3 видно, что на дальнем расстоянии набор штаммов образует четыре кластера, которые объединены преимущественно по близости генетических признаков и высокой однородности морфолого-культуральных признаков. Так, штаммы второго кластера (16.20, 7.20, 10.20) имеют одинаковый генотип, а штаммы четвертого (13.20, 5.20, 11.20, 20.20) – сходный фенотип (табл. 3). Штамм 1.2 имеет уникальный набор признаков.

На дендрограмме рис. 4 выявлены пять кластеров. Первый (5.21, 19.21, 9.21, 18.21) и второй (12.21, 17.21, 1.21, 7.21) кластеры показывают близ-

кое сходство по всему ряду признаков. Признаки штамма 11.21 уникальны. Описание кластеров представлено в табл. 3.

Среди изученных образцов выделяются штаммы 5.20, 11.20, 13.20, 20.20 (2020 г.) и 3.21, 6.21, 14.21, 15.21 (2021 г.). Они имеют очень близкий генотип и сильно сходны по морфолого-культуральным признакам. Это обусловлено принадлежностью к одной популяции штаммов патогена, локализованной на условно выделенной территории. Эти штаммы были получены в зонах А и В [3].

Таблица 2. Описание выявленных кластеров

Номер штамма	Генотип A:B:C:D:E:F	Культуральные признаки			
		характер	структура	рельеф	профиль
2020 г.					
1.20	137:172:177:209:186:220	Разнонаправленный	Клочковатая	Конический	Низкий
5.20	137:172:180:205:211:220	Зонированный	»	Бугристый	»
7.20	139:172:180:203:211:220	Равнонаправленный	Войлочная	Конический	Средний
10.20	139:172:180:203:211:220	»	»	»	»
11.20	137:172:180:203:211:220	Зонированный	Клочковатая	Бугристый	Низкий
12.20	137:172:180:203:211:218	Равнонаправленный	Войлочная	Равномерный	»
13.20	139:172:180:203:211:220	Зонированный	Клочковатая	Бугристый	»
16.20	139:172:180:203:211:220	Концентрический	Войлочная	Конический	Средний
20.20	137:172:180:203:211:220	Зонированный	Клочковатая	Бугристый	Низкий
2021 г.					
1.21	135:172:180:207:211:218	Зонированный	Клочковатая	Бугристый	Низкий
3.21	137:172:180:203:211:220	»	Войлочная	»	»
5.21	137:174:180:203:211:218	»	Клочковатая	»	»
6.21	137:172:180:203:211:220	»	Войлочная	»	»
7.21	157:172:180:205:211:218	»	Клочковатая	»	»
8.21	135:172:180:203:211:220	Равнонаправленный	Войлочная	Равномерный	»
9.21	139:172:180:203:211:218	Зонированный	Клочковатая	Бугристый	Средний
11.21	129:174:177:217:219:222	Равнонаправленный	Однородная	Равномерный	Низкий
12.21	137:172:180:205:211:222	Зонированный	Клочковатая	Бугристый	»
13.21	141:172:180:203:211:220	Равнонаправленный	Пушистая	Равномерный	»
14.21	139:172:180:203:211:220	Зонированный	Войлочная	Бугристый	»
15.21	137:172:180:203:211:220	»	»	»	»
17.21	137:172:180:205:211:222	»	Клочковатая	»	»
18.21	139:174:180:203:211:220	»	»	»	»
19.21	137:172:180:203:211:218	»	»	»	»

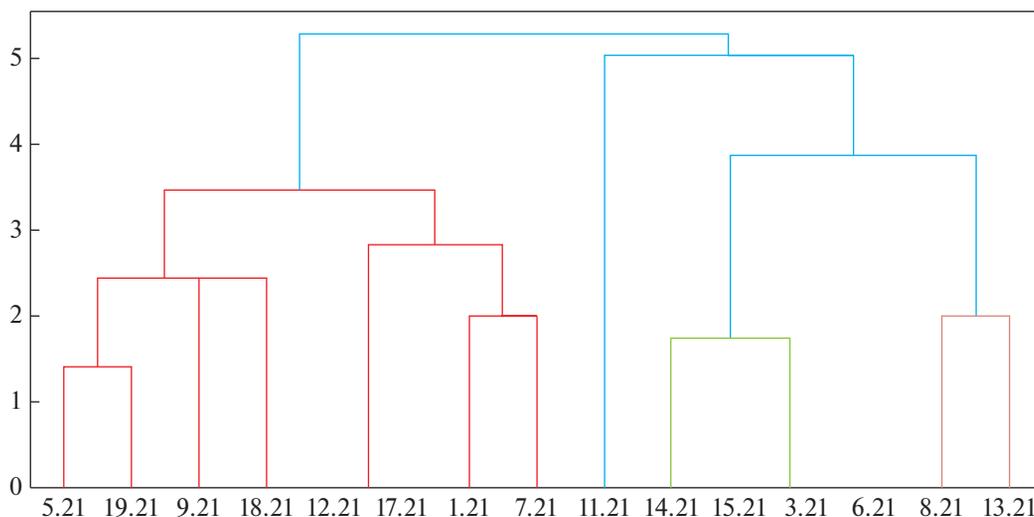


Рис. 4. Дендрограмма кластеризации штаммов, выделенных в 2021 г.

Таблица 3. Описание выявленных кластеров

2020 г.		
1	12.20	Генотип: (137):(172):(180):(203):(211):(218) Фенотип: колонии с равнонаправленным характером роста войлочной структуры и равномерным рельефом низкого профиля
2	16.20 7.20 10.20	Генотип: (139):(172):(180):(203):(211):(220) Фенотип: колонии с равнонаправленным или концентрическим характером роста войлочной структуры и коническим рельефом среднего профиля
3	1.20	Генотип: (137):(172):(177):(209):(186):(220) Фенотип: колонии с разнонаправленным характером роста клочковатой структуры и коническим рельефом низкого профиля
4	13.20 5.20 11.20 20.20	Генотип: (137,139):(172):(180):(203,205):(211):(220) Фенотип: колонии с зонированным характером роста клочковатой структуры и бугристым рельефом низкого профиля
2021 г.		
1	5.21 19.21 9.21 18.21	Генотип: (137,139):(172,174):(180):(203):(211):(218,220) Фенотип: колонии с зонированным характером роста клочковатой структуры и бугристым рельефом низкого или среднего профиля
2	12.21 17.21 1.21 7.21	Генотип: (135,137,157):(172):(180):(205,207):(211):(218,222) Фенотип: колонии с зонированным характером роста клочковатой структуры и бугристым рельефом низкого профиля
3	11.21	Генотип: (129):(174):(177):(217):(219):(222) Фенотип: колонии с равнонаправленным характером роста однородной структуры и равномерным рельефом низкого профиля
4	14.21 15.21 3.21 6.21	Генотип: (137,139):(172):(180):(203):(211):(220) Фенотип: колонии с зонированным характером роста войлочной структуры и бугристым рельефом низкого профиля
5	8.21 13.21	Генотип: (141,135):(172):(180):(203):(211):(220) Фенотип: колонии с равнонаправленным характером роста пушистой или войлочной структуры и равномерным рельефом низкого профиля

Таблица 4. Популяции патогена *Pyricularia oryzae* Cav., распространенные на территории рисосеяния Краснодарского края [3]

Номер популяции	Зона	Признаки
I	А, В, С	Генотип: (135,137,139,157):(172,174):(180):(203,205,207):(211):(218,220,222) Фенотип: колонии с зонированным характером роста клочковатой структуры и бугристым рельефом низкого профиля
II	А, В	Генотип: (137,139):(172):(180):(203,205):(211):(220) Фенотип: колонии с зонированным характером роста клочковатой структуры и бугристым рельефом низкого профиля
III	А, В	Генотип: (139):(172):(180):(203):(211):(220) Фенотип: колонии с равнонаправленным или концентрическим характером роста войлочной структуры и коническим рельефом среднего профиля
IV	В, С	Генотип: (141,135):(172):(180):(203):(211):(220) Фенотип: колонии с равнонаправленным характером роста пушистой или войлочной структуры и равномерным рельефом низкого профиля
V	В	Генотип: (137):(172):(180):(203):(211):(218) Фенотип: колонии с равнонаправленным характером роста войлочной структуры и равномерным рельефом низкого профиля
VI	В	Генотип: (137):(172):(177):(209):(186):(220) Фенотип: колонии с разнонаправленным характером роста клочковатой структуры и коническим рельефом низкого профиля
VII	В	Генотип: (129):(174):(177):(217):(219):(222) Фенотип: колонии с равнонаправленным характером роста однородной структуры и равномерным рельефом низкого профиля

На протяжении всего периода исследований в зоне В выявлялись уникальные штаммы с отличным от других набором генетических и морфологических признаков (штаммы 1.20 и 11.20). Это может быть обусловлено высокой степенью естественного полиморфизма патогена.

Также на протяжении двух лет выявлялись группы штаммов с близкими признаками и неселективные по зоне отбора. Достаточно близкое сходство штаммов по генотипу показывает низкий уровень изменчивости патогена на террито-

рии рисосеяния и, скорее всего, обусловлено локализацией популяций.

Таким образом можно выделить семь популяций патогена, распространенных на территории рисосеяния Краснодарского края (табл. 4).

Исследовательская работа выполнена при финансовой поддержке Кубанского научного фонда в рамках научного проекта № МФИ-20.1/3.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Дубина Е.В., Харченко Е.С., Серая Л.И. и др. База данных “Штаммы патогена *Pyricularia oryzae* Cav. юга России”. 2019. Свидетельство № 2019620149.
2. Фролова В.С., Коваленко Е.Д., Наскидашвили Ж.Г. и др. Методические указания по оценке устойчивости сортообразцов риса к пирикулярриозу в инфекционном питомнике. М.: ВАСХНИЛ, 1983. 14 с.
3. Nartymov D., Dubina E., Kharitonov E. et al. Zoning of rice growing territory of Krasnodar region by the set of features of *Pyricularia oryzae* Cav. isolation // BIO Web of Conferences. 2022. V. 51. 04006. 9 с. <https://doi.org/10.1051/bioconf/20225104006>
4. Nartymov D., Kharitonov E., Dubina E. et al. Studying of cultural properties of *Pyricularia oryzae* Cav. strains in the south of Russia // Microbiol. Res. 2021. V. 12. № 1. P. 21–28. <https://doi.org/10.3390/microbiolres1201000310.3390/microbiolres12010003>
5. Истомин Н.К., Нартымов Д.В., Дубина Е.В. и др. Изучение наиболее оптимальных методов кластеризации морфологических и культуральных признаков // Научно-практ. журн. “Рисоводство”. 2021. № 3(52). С. 22–26. <https://doi.org/10.33775/1684-2464-2021-52-3-22-26>

Studying the Genetic Diversity of the Rice Break Cause *Pyricularia oryzae* Cav. on the Complex of Molecular and Morphological Features

D. V. Nartymov^a, E. V. Dubina^{a, b, *}, M. G. Ruban^a, Yu. V. Aniskina^c, S. V. Garkusha^{a, b}, I. A. Shilov^c, N. S. Velishaeva^c, O. S. Kolobova^c, and N. K. Istomin^a

^aFederal Scientific Rice Centre, Krasnodar, 350921 Russia

^bTrubilin Kuban State Agrarian University, Krasnodar, 350004 Russia

^cAll-Russian Research Institute of Agricultural Biotechnology, Moscow, 127550 Russia

*e-mail: lenakrug1@rambler.ru

On the basis of molecular genetic approaches, the study of the biodiversity of the phytopathogenic fungus *Pyricularia oryzae* Cav. in the south of Russia and a classification of the morphological and cultural characteristics of the pathogen was carried out. For 2020–2022 monitoring was carried out and 24 strains of the pathogen were isolated from the affected herbarium material collected from the fields of five ecological rice growing zones of the Krasnodar Territory (Krasnoarmeisky, Kalininsky, Abinsky, Slavyansky districts, Krasnodar). Using the multiplex PCR system developed by us based on fragment analysis, five genotypes were identified among the studied *Pyricularia oryzae* Cav. strains over the entire period of the project, each of which is characterized by a unique genetic profile. Compiled their “DNA passport”. The studied isolates of the causative agent of blast were characterized by morphological and cultural characteristics. Six morphotypes of the pathogen have been identified.

Keywords: genetic biodiversity, phytopathogenic fungi, *Pyricularia oryzae* Cav. (blast), pathotype, strains, SSR loci, PCR, DNA passports.

ДИНАМИКА РАСПРОСТРАНЕНИЯ РЖАНЫХ ТРАНСЛОКАЦИЙ В ГЕНОТИПАХ РОССИЙСКИХ СОРТОВ МЯГКОЙ ПШЕНИЦЫ *Triticum aestivum* L.

© 2023 г. А. В. Фисенко¹ *, О. А. Ляпунова², Е. В. Зуев², А. Ю. Новосельская-Драгович¹ **

¹Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

²Федеральный исследовательский центр, Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, 190000 Россия

*e-mail: fisenko800@mail.ru

**e-mail: dragova@mail.ru

Поступила в редакцию 25.10.2022 г.

После доработки 16.11.2022 г.

Принята к публикации 18.11.2022 г.

Широкое использование в селекции пшеницы сортов с транслокациями 1RS/1BL и 1RS/1AL было обусловлено их высокой комплексной устойчивостью за счет генов резистентности на 1RS-хромосомах. Многочисленные сведения о появлении новых вирулентных рас патогенов и потере генами, локализованными на 1RS, своих защитных свойств делают актуальным решение вопроса о перспективности использования обеих ржаных транслокаций в современных селекционных программах. Кроме того, использование в интрогрессивной гибридизации различных генетических источников ржаной хромосомы 1RS ставит задачу изучения разнообразия групп сцепления 1RS. Нами исследована динамика распространения ржаных транслокаций в генотипах 240 российских сортов мягкой пшеницы из трех селекционных центров РФ. Для маркирования 1RS использовались множественные аллели генов запасных белков, специфичных для 1RS. Выявлено, что гены устойчивости на 1RS по-прежнему эффективны против ряда патогенов, в частности против возбудителей мучнистой росы и стеблевой ржавчины. Прослежена динамика изменения качества зерна сортов с TR:1RS/1BL. Подтверждено статистически достоверное его улучшение у озимых сортов начиная с 2000-х гг., а у яровых сортов присутствие в генотипах TR:1RS/1BL не препятствует получению зерна с высоким уровнем качества. Показано, что в результате этих процессов во всех трех селекцентрах идет накопление TR:1RS/1BL во времени. Показано высокое аллельное разнообразие секалин-кодирующего локуса (*Gli-B1*) на хромосоме 1RS. Однако все сорта с TR:1RS/1BL независимо от происхождения 1RS несли один и тот же аллель. Обсуждаются возможные причины такого единообразия. Транслокация 1RS/1AL в исследованных нами сортах не выявлена, обсуждаются возможные причины.

Ключевые слова: пшеница, ржаные транслокации, гены запасных белков.

DOI: 10.31857/S0016675823050053, **EDN:** FNOUIG

Мягкая пшеница является одной из важнейших сельскохозяйственных культур мира. Поэтому от ее продуктивности зависит благосостояние населения большинства стран мира. Критическим фактором в получении высокого и стабильного урожая зерна является устойчивость культивируемых сортов к биотическим и абиотическим стрессам. При поиске новых генов для улучшения пшеницы большое внимание уделяется ее диким и культурным родственникам. Одним из широко используемых доноров генов устойчивости к заболеваниям является рожь [1]. Несмотря на то, что рожь относится к видам третичного генофонда (виды, геномы которых не гомологичны геномам

пшеницы), анализ геномных последовательностей пшеницы и ржи выявил их высоко консервативную колинеарность. Однако только хромосома 1R является полностью синтетической своим гомеологам у пшеницы [2]. Вероятно поэтому на фоне большого количества линий пшеницы [3] с генетическим материалом ржи, только две из них нашли широкое применение в практике. Это транслокации, в которых участвует короткое плечо хромосомы 1R ржи — TR:1RS.1BL и TR:1RS.1AL.

Первые официальные российские сорта с TR:1RS.1BL — Кавказ и Аврора (районированные в 1970-е гг.) произошли от линии Neuzucht 14/14 (с 1RS от сорта ржи Petkus) и стали донорами этой

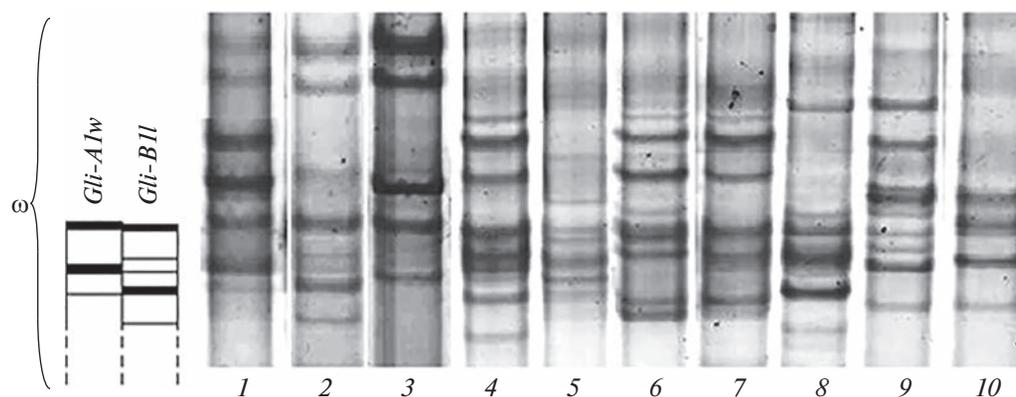


Рис. 1. ω – зона электрофоретических спектров различных сортов пшеницы (1–3) и тритикале (4–10): 1 – сорт Богданка (несет аллельный вариант *Gli-A1w* локуса, контролирующего секалины и расположенного на коротком плече хромосомы 1R в транслокации 1RS.1AL); 2 – сорт Прохоровка (несет аллель *Gli-B11* локуса на 1RS в транслокации 1RS.1BL); 3 – сорт Безостая 1 (несет аллель *Gli-B1b* локуса, расположенного на коротком плече пшеничной хромосомы 1B); 4 – АД Зеленый; 5 – Ставропольский 2; 6 – Ставропольский 1; 7 – Ставропольский 3; 8 – Донской 2; 9 – АД 196; 10 – АД 206. Схематически представлены блоки компонентов, контролируемые *Gli-A1w*, *Gli-B11*.

транслокации для многих отечественных и зарубежных сортов [3, 4]. Сегмент 1RS обладает большим числом генов устойчивости¹. В результате, первые сорта с 1RS характеризовались комплексным иммунитетом, обеспечивающим высокую урожайность, однако эти положительные свойства новых сортов сопровождались значительным снижением их качества [6–8].

Источником транслокации 1RS.1AL является сорт Amigo (1970-е гг., США), 1RS у которого получен от сорта ржи Insave F.A. [9]. С появлением TR:1RS.1AL генофонд пшеницы обогатился дополнительными генами устойчивости². Сейчас в мире создано около 100 сортов и линий с этой транслокацией [4].

Низкое хлебопекарное качество сортов с TR:1RS/1BL связывается с отсутствием в геноме короткого плеча 1B-хромосомы (1BS), на котором находятся локусы *Glu-3* и *Gli-1*. Они кодируют низкомолекулярные глютеины и глиадины соответственно [10, 11]. Считается, что оба локуса вносят значительный вклад в качество пшеницы, положительно влияя на время развития и вязкость теста, а также на объем хлеба [12, 13].

Локусы *Glu-3* и *Gli-1* представляют собой кластеры сцепленных генов [14], характеризующиеся множественным аллелизмом. Аллельные варианты одного локуса контролируют синтез полипептидов,

которые на электрофореграмме выглядят как блоки сцепленных компонентов. Блоки различаются числом, интенсивностью, а также подвижностью входящих в них компонентов, что определяется различием в числе экспрессирующихся генов, молекулярной массе и заряде полипептидов. Для локуса *Gli-B1* на 1BS-хромосоме известно минимум 16 аллельных вариантов [15]. Такое высокое разнообразие в сочетании с разработанной интерпретацией результатов [16] используется для маркирования участков хромосом. Присутствие ржаного плеча 1RS легко выявляется по аллельному варианту *Gli-B11*, в случае TR:1RS.1BL [17] (1RS передано сортом ржи Petkus) и по *Gli-A1w* – в случае TR:1RS.1AL [18] (1RS передано рожью Insave). Белковые продукты двух аллельных вариантов легко выявляются на электрофореграмме в виде характерных различающихся между собой блоков компонентов, которые не имеют ничего общего с любым другим пшеничным блоком, так как это продукты секалин-кодирующего локуса, расположенного на 1RS, характеризующегося, как и глиадинкодирующие локусы, множественным аллелизмом [15, 17, 19] (рис. 1).

К 2000-х гг. генетикам и селекционерам удалось “разорвать” негативную сцепку “устойчивость к болезням–качество” [20], но с 1990-х гг. уже появляются сведения о потере генами на 1RS в 1RS.1BL и 1RS/1AL транслокациях своей резистентности [21, 22]. Для усиления защитных свойств новых сортов пшеницы в селекционном процессе стали использовать альтернативные источники 1R-хромосомы. Поскольку рожь является перекрестно опыляющейся культурой и характеризуется чрезвычайно высоким генетическим разнообразием [23], то в ее геноме предполагалось найти 1RS с другими аллелями генов устойчивости,

¹ *Lr26*, *Sr31*, *Yr9*, *Pm8* и *Dn7*, гены, контролирующие иммунитет к бурой *Puccinia triticina* f. sp. *tritici*, стеблевой *P. graminis* f. sp. *tritici*, желтой *P. striiformis* f. sp. *tritici*, ржавчинам, мучнистой росе *Blumeria graminis* f. sp. *tritici* и русской пшеничной тле *Diuraphis noxia* соответственно [5].

² *Pm17* и *SrAmi* – гены устойчивости к стеблевой ржавчине и мучнистой росе, *Gb2*, *Gb6* и *Cmc3* – гены устойчивости к тле *Schizaphis graminum* и клещу *Aceria tosicheilla* соответственно [5].

которые бы эффективно работали с современными расами патогенов. Для этой цели использовались межвидовые скрещивания пшеницы с тритикале — вида, у которого один из геномов ржаной. Учитывая высокое генетическое разнообразие ржи, можно предположить, что новые группы сцепления 1RS будут маркироваться аллелями, отличающимся от двух уже известных — *Gli-B11* и *Gli-A1w*.

В отличие от обширных исследований динамики накопления чужеродных транслокации в генофонде пшеницы из различных стран мира [1, 24, 25] подобные исследования на отечественных сортах носят ограниченный и несистемный характер, несмотря на высокую актуальность таких работ [18, 20, 26]. Было показано небольшое увеличение частоты встречаемости TR:1RS.1BL с 40 до 44% в селекционном материале Краснодарского НИИСХ за период 2003–2014 гг. А их сорта, внесенные в Госреестр в 2014 г., на 36.6% состояли из носителей этой транслокации. В селекционном материале также отмечено повышение частоты встречаемости транслокации 1RS.1AL с 2.7 до 8.5% с 2011 по 2014 гг. [20]. Исследование 20 сортов селекции БелНИИСХ позволило выявить три сорта с TR:1RS.1BL и один сорт с TR:1RS.1AL [18]. Среди яровых сортов СибНИИСХа, в родословных которых присутствовал сорт Кавказ, обнаружено только три сорта с TR:1RS.1BL [26], однако эта выборка не отражала разнообразия сибирских сортов. Значительная часть отечественных сортов остается не изученной.

Задачами нашей работы были: 1) исследование динамики чистоты встречаемости транслокаций 1RS.1AL и 1RS.1BL в современных российских сортах мягкой пшеницы в трех крупных селекционных центрах РФ, 2) выявление взаимосвязей между присутствием транслокаций и актуальной устойчивостью к болезням, а также показателями качества зерна, 3) поиск неизвестных ранее аллелей локуса *Gli-1*, маркирующих группу сцепления короткого плеча 1R-хромосомы ржи, полученной из альтернативных источников.

МАТЕРИАЛЫ И МЕТОДЫ

Проанализировано 240 сортов, созданных в регионах со значительной долей пшеницы в общей площади посевов. Это: Северокавказский регион (г. Краснодар), Нижняя Волга (г. Саратов) и Западная Сибирь (г. Омск). Большая часть изученных в работе образцов была получена от оригинаторов сортов из соответствующих научных учреждений. Образцы сорта Богданка предоставлены д. б. н. В.П. Нецветаевым. Часть современных сортов, районированных в Российской Федерации в последние годы, получена из ФИЦ Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова (ВИР) Министерство науки и высшего образования.

Для идентификации ржаных транслокаций 1RS.1AL и 1RS.1BL, а также поиска неизвестных ранее аллелей глиадинкодирующих локусов использован метод электрофореза спирторастворимых запасных белков — глиадинов. Приготовление образцов и электрофорез в 8.3%-ном полиакриламидном геле (ПААГ) в кислом лактат-алюминиевом буфере проводили по стандартному методу [15]. В каждом сорте методом электрофореза исследовано по 100 отдельных зерновок.

Для изучения взаимосвязи между локусами *Gli-1* и *Glu-3* у сортов с транслокацией 1RS.1BL была использована база данных GRIS [27].

Для расчета показателей устойчивости к болезням (бурая, желтая и стеблевая ржавчины, септориоз, мучнистая роса, фузариоз) были использованы база данных испытаний сортов на искусственных инфекционных фонах [28, 29]. Устойчивость сортов оценивалась в баллах от 1 до 5: 1 — восприимчивый сорт, 5 — устойчивый. Для двух сравниваемых групп сортов (со ржаной транслокацией и без нее) рассчитывались усредненные баллы устойчивости к шести основным заболеваниям. Разделение сортов пшеницы на группы качества — сильные, ценные, филлеры — проведено в соответствии с описаниями сортов в Госреестре [30].

Достоверность различий групп сортов по качеству оценивалась по критерию Стьюдента, а по устойчивости — по χ -критерию Ван-дер-Вардена.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Транслокация 1RS.1BL (TR:1RS.1BL)

Транслокация 1RS.1BL в озимых сортах мягкой пшеницы, созданные в Национальном центре зерна им. П.П. Лукьяненко (г. Краснодар)

Изучено 120 озимых сортов мягкой пшеницы, созданных с 1958 по 2021 гг. Сорта разделены на четыре группы: 1) сорта советской селекции (1958–1989 гг.), 2) сорта, созданные в 1990-е годы, 3) сорта, созданные в 2000-е гг. и 4) сорта селекции 2010-х гг. (табл. 1). Выявлено, что частота встречаемости 1RS.1BL-транслокации, маркируемой аллелем *Gli-B11* неуклонно растет: от 20% в сортах 1-й группы до 59.5% в сортах 4-й группы (табл. 1). Как отмечалось выше, TR:1RS.1BL попала в краснодарские сорта из четырех разных источников: это сорта Кавказ и Аврора (1RS от ржи “Petkus”) и три разных сорта тритикале (см. примеч. табл. 1). Исследования электрофоретических спектров запасных белков у 15 сортов озимых тритикале, позволило показать их широкое разнообразие, которое контролируется разными аллельными вариантами локуса на 1RS (рис. 1). Учитывая эти данные, а также высокое разнообразие сортов ржи [23, 31], ожидалось, что 1R-хромосомы, полученные из разных источников, будут отличаться аллельными вариантами от *Gli-B11*. Однако все

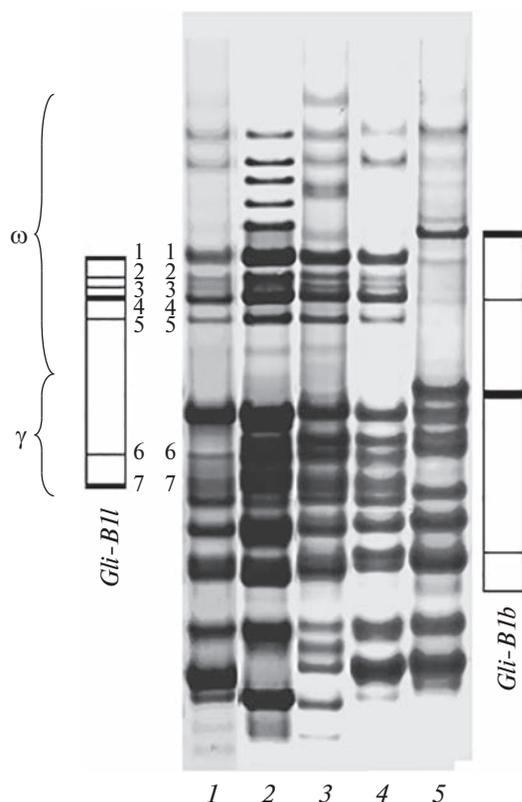


Рис. 2. Электрофоретические спектры сортов Яшуклянка (1), Княжна (2), Тая (3), Кавказ (4). Сорстандарт Безостая 1 (5). Схематически указаны компоненты спектра, контролируемые аллельным вариантом *Gli-B1I* на коротком плече хромосомы 1R и *Gli-B1b* (Безостая 1), контролируемые пшеничным коротким плечом 1RS. γ и ω – общепринятые обозначения зон спектра.

проанализированные сорта независимо от источника получения 1R-хромосомы, характеризовались блоком электрофоретических компонентов, контролируемым аллельным вариантом локуса *Gli-B1I*, который маркирует короткое плечо хромосомы 1R, полученной от ржи “Petkus” (рис. 2)³. Было показано, что из 23 наиболее известных в мире в 1980-е годы сортов с 1RS, предположительно полученной из разных источников, только у некоторых зерновок сорта Weique наблюдался аллель, отличный от аллеля *Gli-B1I* [32].

Для уточнения группы сцепления, маркируемой аллелем *Gli-B1I*, нами проведено сопоставление аллелей локусов *Gli-B1* и *Glu-B3* у 19 транслокационных сортов 1RS.1BL с использованием базы данных GRIS [27]. Показано, что аллелю *Gli-B1I* во всех случаях соответствует аллель *Glu-B3j*, образуя группу

³ Генетический контроль компонентов, контролируемых локусом *Gli-B1I*, был определен с помощью гибридологического исследования [17].

сцепления *Gli-B1I/Glu-B3j*, служащую надежным маркером транслокации 1RS.1BL.

Один и тот же аллель локуса *Gli-B1*, наблюдаемый у всех исследованных сортов с TR:1RS.1BL, несмотря на разное происхождение 1RS в них позволяет сделать ряд предположений о причинах возникновения такого единообразия. И наиболее очевидное из них – это межгеномные взаимодействия, ограничивающие существующее аллельное разнообразие *Gli-B1* ржи и тритикале единственным конкурентоспособным аллелем *Gli-B1I* для полноценного функционирования генетического материала ржи в геноме пшеницы. Косвенным подтверждением этого предположения является тот факт, что несмотря на активное привлечение рода *Secale* в селекционные программы и получение большого числа линий с генетическим материалом ржи [4, 31], только две из них (кроме 1RS.1BL и 1RS.1AL) получили некоторое распространение – это 4BS.4BL-2RL с генами *Lr25* и *Pm7*, отмеченная у ряда североамериканских сортов [33], и 6D.6RL, обнаруженная у китайского сорта Xiao Yan 6 [34, 35].

Еще одна теоретическая возможность появления одного и того же 1RS плеча у 1RS.1BL сортов – это случайное участие в сложных межвидовых гибридах линий пшеницы с *Gli-B1I* аллелем.

Преодоление патогенами защитных свойств растений с 1RS.1BL и 1RS.1AL [21, 22] должно было повлиять на постепенное их вымывание из популяции современных сортов. Однако мы наблюдаем существенное увеличение частоты их встречаемости во времени (табл. 1), что ставит вопрос об их реальной устойчивости к патогенам. Нами проанализирована устойчивость сортов, районированных начиная с 1990-х гг. с TR:1RS.1BL, и ее отсутствие к шести основным заболеваниям по данным [28, 29] (табл. 2). Для каждой группы сортов рассчитаны усредненные баллы устойчивости к каждому заболеванию, при этом расы патогенов, меняющиеся во времени, не учитывались. Согласно данным табл. 2 устойчивость к двум заболеваниям – стеблевой ржавчине и мучнистой росе, у группы сортов с TR:1RS/1BL статистически достоверно превосходит таковую у сортов без TR:1RS.1BL. То есть с высокой долей вероятности можно предположить, что в Северо-Кавказском регионе гены *Sr31* и *Pm8* остаются эффективными против большинства рас обоих типов патогенов.

Положительное влияние 1RS.1BL на устойчивость к мучнистой росе отмечено также у современных сортов Беларуси с этой транслокацией [36]. Что касается остальных четырех заболеваний, то наши данные подтверждают, что устойчивость к ним, определяемая генами на 1RS.1BL, уже преодолена новыми вирулентными биотипами соответствующих возбудителей [1].

Таблица 1. Наличие ржаной транслокации 1RS.1BL, маркируемой аллелем *Gli-B11*, в краснодарских, саратовских и омских сортах разных периодов селекции

Годы районирования	Названия сортов		Сортов всего/с <i>Gli-B11</i>	Частота сортов с 1RS.1BL (%)
	без транслокации	с транслокацией		
КРАСНОДАР				
1958–1989	Безостая 1, Краснодарские 6, 39, 46, 57, Исток, Замена, Дружина, Криница, Олимпия, Павловка, Ранняя 12, Полукарликовая 49, Северокубанка, Северокубанская 43, Спартанка	Аврора, Безостая 2, Кавказ, Массив	20/4	20.0
1990–1999	Дельта, Деметра, Зимородок, Краснодарские 70, 90, Даха, Крошка, Леда, Ника Кубани, Олимпия 2, Офелия, Победа 50, Руфа, Скифянка, Соратница, Уманка, Эхо, Югина, Юна	Горлица, Княжна, Купава, Половчанка, Хазарка	24/5	20.8
2000–2009	Батько, Виза, Горянка, Грация, Есаул, Зимница, Зимтра, Краснодарская 99, Кума, Ласточка, Лебедь***, Лига 1, Лира, Москвич, Нота, Палпич, Память, Русса, Старшина, Фишт, Юбилейная 100, Юнона	Айвина, Веда*, Дея, Афина, Вита, Дока, Восторг, Иришка, Красота, Мафэ, Тая, Миллениум, Первица, Селянка, Фортуна, Файл, Яшкулянка	39/17	43.6
2010–2021	Бригада, Вершина*, Гром, Доля, Морозко, Протон, Сила, Табор, Трио, Этнос, Юмпа, Дуплет, Сварог, Собербаш, Еланчик	Адель, Алексеич, Баграт, Безостая 100, Васса, Гурт**, Курень*, Курс, Юка*, Стан, Лауреат, Ольхон**, Уруп, Вид, Антонина, Ваня, Веха, Жива**, Степь, Велена, Каравай, Тимирязевка 150	37/22	59.5
САРАТОВ				
1920–1940	Альбидум 604, 721, 24, Лютесценс 62, 55/11, 53/12, Эритроспермум 341, Сарроза, Саррубра, Полтавка, Русская, Селивановский Русак, Междуреченская, Милоградовка	–	14/0	0
1940–1976	Альбидум 21, 43, 1616, 1697, Лютесценс 53/12, 758, Стекловидная 1, Саратовские 210, 29, 33, 36, 38, 39, 42, 44	–	15/0	0
1979–1989	Саратовские 45, 46, 48, 49, 50, 52, 54, 55, Альбидум 28	–	9/0	0
1991–1999	Саратовские 58, 60, 62, Л-503, Л-505, Самсар, Белянка, Альбидум 29, Альбидум 188	Прохоровка	10/1	10

Таблица 1. Окончание

Годы районирования	Названия сортов		Сортов всего/с <i>Gli-B1</i>	Частота сортов с 1RS.1BL (%)
	без транслокации	с транслокацией		
2000–2021	Саратовские 64, 66, 68, 70, 73, 74, 76, Альбидум 31, 33, Добрыня, Гречанка, Фаворит, Александрит	Юго-восточная 2, Юго-восточная 4	15/2	13.3
ОМСК				
1929–1957	Ноз сибирский, Смена, Цезиум 111, 94, Лютесценс 956, Тарская 4, Чеченевская, Альбидум 3700, Мильтурум 553	–	10/0	0
1976–1989	Сибирячка 4, Нива, Лютесценс 34, Омские 9, 12, 17, 19, Сibaковская 3, Иртышанка 10	–	9/0	0
1991–1999	Омские 18, 24, 28, 29, 30, Диас 2, Нива 2, Росинка, Росинка 2	Омская 20, Омская 29	11/2	18.2
2000–2021	Памяти Азиева, Омские 32, 33, 35, 36, Омская юбилейная, Соната, Страда Сибири, Славянка Сибири, Светлана, Катюша, Боевчанка, Омская краса, Волошинка, Серебристая, Силангий, Столыпинская, Столыпинская 2, Тарская 12	Омская 37, Омская 38, Геракл, Мелодия, ОмГАУ 100, Омская 44, Сигма, Элемент 22	27/8	29.6

Примечание. * – потомки сортов Княжна и Половчанка, полученные в результате межвидовой гибридизации пшеницы и тритикале – АД 206/Рубин//АД-17-Т-3; ** – потомки сорта Таня, отселектированного из гибридов межвидового скрещивания (*Тритикале* Grado/*T. aestivum*); *** – потомки сорта Яшкульянка, полученные в результате межвидовой гибридизации пшеницы и тритикале (*Тритикале* PRAG-48-4 × AD-196) × *T. aestivum* сорт Спартанка).

Таблица 2. Распределение краснодарских сортов по группам устойчивости к патогенам в зависимости от наличия/отсутствия транслокации 1RS.1BL на основе данных Каталогов сортов [28, 29]

Группы сортов	Число сортов	Средняя оценка устойчивости в группах сортов, балл					
		ржавчины			септориоз	мучн. роса	фузариоз
		бурая	желтая	стеблевая			
Сорта с 1RS.1BL	24	3.21	3.71	4.42**	3.17	4.25*	2.54
Без 1RS.1BL	35	3.31	3.09	2.87	2.83	3.49	2.09

Примечание. * – различие между группами сортов с 1RS.1BL и без транслокации достоверно на уровне значимости $0.01 < P < 0.05$; ** – различие между группами сортов с 1RS.1BL и без транслокации достоверно на уровне значимости $P < 0.01$.

Хлебопекарные свойства муки и хлеба формируются в значительной степени под воздействием генотипической среды [31, 37], на которую определяющее влияние оказывают множественные аллели шести глютенин- и шести глиадин-кодирующих локусов – *Glu-1*, *Glu-3*, *Gli-1*, *Gli-2*. [см. 14, 37]. Ухудшение качества сортов 1RS.1BL с локусами ржи на коротком плече 1R пытались исправить методами хромосомной инженерии. Для этого заменили ржаные локусы, контролируемые глиадины и глютенины, на пшеничные (*Gli-*

B1 и *Glu-B3*). Однако в ходе работы было выявлено неаллельное положение генов пшеницы и ржи и сложное их сцепление с набором генов устойчивости (*Pm8*, *Lr26*, *Yr9* и *Sr31*), расположенных между генами запасных белков, которые должны были быть заменены. Еще одна попытка, в результате которой была отобрана двойная рекомбинантная хромосома, несущая две небольшие интеркалярные вставки и все желаемые аллели как пшеницы, так и ржи, также оказалась неудачной, т.к. возникла неожиданная связь изъятых

Таблица 3. Распределение краснодарских сортов по группам качества в зависимости от присутствия/отсутствия в них транслокации 1RS.1BL на основе данных Госреестра [30]

Период селекции, годы	Группа сортов	Группы качества			
		сильные и ценные		филлеры	
		<i>N</i>	%	<i>N</i>	%
1958–2009*	Сорта с 1RS.1BL	10	38.5	16	61.5
	Сорта без 1RS.1BL	36	78.3	10	21.7
2010–2021	Сорта с 1RS.1BL	20	90.9	2	9.1
	Сорта без 1RS.1BL	14	93.3	1	6.7

Примечание. * – различие между группами сортов с 1RS.1BL и без транслокации достоверно на уровне значимости $0.01 < P < 0.05$, *N* – число сортов.

генетического материала ржи с генетическими факторами, контролирующими эффективность корневой системы [38, 39].

Активная работа по преодолению жесткой связи между присутствием 1RS/1BL и низким качеством велась в Краснодаре, начиная с 1990-х гг., с использованием традиционных методов. В результате появились сорта, которые наряду с присутствием TR:1RS.1BL, обладали высоким качеством зерна [20]. Некоторые сорта (Алексеич, Баграт, Безостая 100, Горлица, Дея, Стан, Жива) относились даже к разряду сильных. Используя данные Госреестра [30], мы сравнили сорта с TR:1RS.1BL и без TR:1RS.1BL по качеству. Сравнение проводилось в двух группах сортов: 1) сорта, созданные до 2010 г. и 2) сорта, созданные с 2010 по 2020 гг. (табл. 3). Установили достоверное различие в качестве муки, присущем сортам более ранней селекции, то есть наблюдается статистически достоверное преобладание сильных и ценных по качеству генотипов в группе сортов, свободных от 1RS.1BL. Сорта, созданные в последнее десятилетие, характеризуются близкими значениями показателей качества независимо от наличия или отсутствия в них TR:1RS.1BL.

Исследования интрогрессивных линий, созданных на разном пшеничном генетическом фоне, с одним и тем же генетическим материалом ржи, показали, что удачно подобранный генотип пшеницы способен нивелировать отрицательное влияние генетического материала ржи [31]. Эти данные позволяют предположить, что для современных краснодарских сортов подобран определенный генетический фон, в котором, в частности, совмещаются такие аллели *Glu*- и *Gli*-локусов, которые влияют на формирование высокого качества муки и хлеба.

Транслокация 1RS.1BL в яровых сортах Нижневолжского региона

Исследовано 63 сорта, созданных с 1924 по 2021 гг. в ФАНЦ Юго-Востока, а также в его фи-

лиалах: Краснокутской и Ершовской селекционных станциях (табл. 1). До 1990-х гг. в Саратове не зафиксировано сортов с TR:1RS/1BL, первым таким сортом стала Прохоровка (1995 г.). В 2000 г. и в 2002 г. были районированы еще два сорта с TR:1RS/1BL, маркируемой аллелем *Gli-B11* – Юго-Восточная 2 и ЮВ 4, унаследовавшие 1RS от сорта Кавказ. Среди сортов, созданных после 2000 г., 1RS.1BL встречается с частотой 13.3% (табл. 1). Низкая частота встречаемости 1RS.1BL может быть связана как со сравнительно недавним включением в селекционный процесс сортов-носителей транслокации, так и с недостаточной ценностью этого генетического материала в Нижневолжском регионе. В отношении хлебопекарного качества зерна сорта Юго-Восточная 2 и ЮВ 4 являются сильными пшеницами, т.е. это зерно самого высшего качества.

Транслокация 1RS.1BL в яровых сортах Сибирского ФНЦ агробиотехнологий (г. Омск)

Исследование 57 сортов мягкой пшеницы, созданных в период 1929–2021 гг. (табл. 1), позволило выявить, что первым сортом с TR:1RS.1BL от сорта Кавказ стал сорт Омская 20 (1993 г.). Сорт гетерогенен – только 25% зерен несет TR:1RS.1BL, представляя собой отдельный биотип сорта [15]. Необходимо отметить, что *in situ* гибридизация у сорта Омская 20 не выявила 1RS.1BL-транслокацию [26], что, вероятно связано с анализом небольшого количества зерен, в число которых не вошел биотип с TR:1RS.1BL. Позже был создан гетерогенный сорт Омская 29 (1999 г.), а в 2000-е гг. появилось сразу восемь сортов с TR:1RS.1BL (табл. 1); и только один из них гетерогенен (сорт Омская 38) по TR:1RS.1BL. Остальные же сорта гомогенны.

Судя по динамике частоты встречаемости TR:1RS.1BL у сортов 1990-х (18.2%) и 2000-х годов (29.6%), в Омске происходит накопление транслокации (табл. 1). Пять из восьми сортов с TR:1RS.1BL, созданных в 2000-х гг. – Омские 20,

29, 37, 38 и ОмГАУ 100 – обладают высоким качеством, относясь к группе сильных пшениц.

Помимо ржаных транслокаций, в селекционный процесс активно вовлекаются линии, полученные с участием генетического материала дикорастущих злаков. Так, сорта Омская 37 и Омская 38 кроме TR:1RS.1BL несут пшенично-пырейную транслокацию 7DL-7Ai, в которой сегмент хромосомы 7Ai принадлежит пырею *Agropyron elongatum* (Host) Beauv. Комплексная устойчивость этих сортов обусловлена влиянием кластеров генов, локализованных на сегментах хромосом ржи и пырея [40]. Показано, что, пирамидирование “ржаного” гена *Sr31* с другими эффективными генами *Sr* в одном генотипе обеспечивает защиту растений даже от наиболее вредоносной угандийской расы стеблевой ржавчины Ug99 [41].

Транслокация 1RS/1AL

В трех изученных нами в настоящей работе группах сортов транслокация 1RS.1AL не обнаружена. Ранее отмечалось активное привлечение исходного материала с TR:1RS.1AL в селекционные программы Краснодарского центра [20]. Однако работа с этим материалом началась только с 2010-х годов, и, вероятно, сорта с TR:1RS.1AL могут еще появиться.

В целом среди сортов, культивируемых в РФ, данная транслокация встречается редко. Она выявлена у сорта Богданка (Белгородского НИИСХ, 2009 г.) [18] и подтверждена нашими исследованиями (рис. 2) – сорт несет аллель *Gli-A1w*. Эту транслокацию включает также и одесский сорт Княгиня Ольга [42]. В ходе мониторинга и генотипирования современных сортов РФ из более чем 500 образцов разных сортов, поступивших нам из многочисленных регионов страны в последние несколько лет, обнаружено только два сорта с аллелем *Gli-A1w* – это озимая пшеница Чорнява украинской селекции и яровая пшеница Канюк французской селекции. Наличие у сорта Канюк TR:1RS/1AL подтверждено также цитологически, с использованием С-бэндинга (Е.Д. Бадаева, личное сообщение).

Таким образом, на данный момент TR:1RS.1AL идентифицирована только у четырех сортов, районированных в России. Как показано в ряде исследований [43], наличие транслокации 1RS.1AL, в отличие от 1RS.1BL, не приводит к резкому снижению хлебопекарного качества сортов. Вероятно, ограниченное распространение этой транслокации может быть связано как с поздним включением в селекционный процесс сортов-доноров, так и неэффективностью TR:1RS.1AL в генотипах сортов в РФ.

ЗАКЛЮЧЕНИЕ

Исследование 240 сортов из трех крупных селекционных центров позволило показать динамику распространения TR:1RS.1BL во времени. Выявлено широкое распространение TR:1RS.1BL среди сортов нашей страны, особенно в сортах, созданных в Краснодарском центре, и частота ее встречаемости продолжает там увеличиваться со временем. В сортах, выведенных до 2000-х гг., частота встречаемости транслокации составляла 20%, а после 2010-х гг. – 59.5%. Сравнение групп сортов с TR:1RS.1BL и без нее, районированных после 1990 г. подтвердило, что устойчивость к ряду болезней до сих пор обеспечивается комплексом генов, локализованных на 1RS-хромосоме. Наряду с этим преодоление отрицательной корреляции между присутствием в геноме TR:1RS.1BL и качеством зерна, позволяет материалу со ржаной транслокацией получить преимущество в селекционном процессе.

Частоты встречаемости 1RS.1BL у яровых сортов из саратовского и омского центров значительно ниже по сравнению с озимыми сортами Краснодара; это может быть связано с более поздним включением доноров транслокации в селекционный процесс, но динамика распространения TR:1RS.1BL и здесь имеет положительную тенденцию: в Саратове частота аллеля *Gli-B1l* среди сортов, созданных после 2000 г., достигает 13.3%, а в Омске повышается с 18.2 (у сортов 1990-х гг.) до 29.6% (у сортов 2000-х гг.). Полезной особенностью сортов из Омска и Саратова является то, что наличие у них TR:1RS.1BL практически не отражается на качестве зерна, что также отвечает интересам селекции.

Положительной динамике частоты встречаемости сортов с 1RS/1BL может способствовать также и потепление климата, т.к. TR:1RS.1BL способствует повышению урожайности. Это происходит не только за счет устойчивости к ряду заболеваний, но и повышенной устойчивости растений к засухе – за счет увеличения длины корня [44], позволяющего улучшать поглощение воды и минералов из глубоких слоев почвы, так и за счет увеличения содержания пролина. Последний играет ключевую роль в устойчивости растений к стрессу, удаляя активные формы кислорода и, как молекулярный шаперон, обеспечивая активность ферментов [45]. Более высокая урожайность сортов с TR:1RS.1BL может быть связана также с влиянием хроматина ржи на задержку старения листьев после цветения – растение долго остается зеленым и, следовательно, обеспечивается поступление ассимилированного углерода

к зерну, что также играет важную роль в повышении продуктивности [46].

Транслокация 1RS.1AL только в последние годы получила некоторое распространение среди российских сортов, и ее частота в наших сортах незначительна. Причиной этого может быть сравнительно недавнее включение TR:1RS.1AL в селекционный процесс, с одной стороны, или отсутствие преимущества таких сортов – с другой.

Финансирование работы осуществлялось за счет средств Федерального бюджета, номер государственного учета РИД и ЕГИСУ НИОКТР АААА-А19-119022790099-9.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Rabinovich S.V.* Importance of wheat-rye translocations for breeding modern cultivar of *Triticum aestivum* L. // *Euphytica*. 1998. V. 100. P. 323–340. <https://doi.org/10.1023/A:10183618192>
2. *Bauer E., Schmutzer T., Barilar I. et al.* Towards a whole-genome sequence for rye (*Secale cereale* L.) // *Plant J*. 2017. V. 89. P. 853–869. <https://doi.org/10.1111/tpj.13436>
3. *Дорофеев В.Ф., Якубцинер М.М., Руденко М.И. и др.* Пшеницы мира. Л.: 1976.
4. *Schlegel R.* Current list of wheats with rye and alien introgression. V. 01.22. 2022. <http://rye-gene-map.de/rye-introgression>.
5. *McIntosh R.A., Hart G., Gale M.* Catalogue of gene symbols for wheat // *Proc. 8th Intern. Wheat Gen. Symp.* / Eds Li Z., Xin Z.Y. Beijing, China: 1993. P. 1333–1500.
6. *Lee J.H., Graybosch R.A., Peterson C.J.* Quality and biochemical effects of a 1RS.1BL wheat-rye translocation in wheat // *Theor. Appl. Genet.* 1995. V. 90. P. 105–112. <https://doi.org/10.1007/BF00221002>
7. *Kumlay A.M., Baenziger P.S., Gill K.S. et al.* Understanding the effect of rye chromatin in bread wheat // *Crop Sci*. 2003. V. 43. № 5. P. 1643–1651. <https://doi.org/10.2135/cropsci2003.1643>
8. *Чуманова Е.В., Ефремова Т.Т., Трубочеева Н.В. и др.* Хромосомный состав пшенично-ржаных линий и влияние хромосом ржи на устойчивость к болезням и на хозяйственно ценные признаки // *Генетика*. 2014. Т. 50. № 11. С. 1319–1328. <https://doi.org/10.7868/S0016675814110034>
9. *Sebesta E.E., Wood E.F.* Transfer of greenbug resistance from rye to wheat with X-rays // *Agronomy Abstracts*. Madison, WI: 1978. P. 61–62.
10. *Burnett C.J., Lorenz K.J., Carver B.F.* Effects of the 1B/1R translocation in wheat on composition and properties of grain and flour // *Euphytica*. 1995. V. 86. P. 159–166. <https://doi.org/10.1007/BF00016353>
11. *Li Z., Ren T., Yan B. et al.* A mutant with expression deletion of gene Sec-1 in a 1RS.1BL line and its effect on production quality of wheat // *PLoS One*. 2016. V. 11(1). e0146943. <https://doi.org/10.1371/journal.pone.014694>
12. *Lundh G., MacRitchie F.* Size exclusion HPLC characterization of gluten protein fractions varying in bread-making potential // *J. Cereal Sci.* 1989. V. 10. P. 247–253.
13. *Liu C.-Y., Shepherd K.W.* Inheritance of B subunits of glutenin and ω - and γ -gliadins in tetraploid wheats // *Theor. Appl. Genet.* 1995. V. 90. P. 1149–1157. <https://doi.org/10.1007/BF00222936>
14. *Новосельская-Драгович А.Ю.* Генетика и геномика пшеницы: запасные белки, экологическая пластичность и иммунитет // *Генетика*. 2015. Т. 51. № 5. С. 1–16.
15. *Упельник В.П., Новосельская-Драгович А.Ю., Шишкина А.А. и др.* Лабораторный анализ белков семян пшеницы. М.: ООО “Ваш формат”, 2013. 173 с.
16. *Новосельская-Драгович А.Ю., Янковская А.А., Бадаева Е.Д.* Интрогрессии и хромосомные перестройки не влияют на активность глиадинокодирующих генов в линиях гибридов *Triticum aestivum* L. × *Aegilops columnaris* Zhuk. // *Вавилов. журн. генет. и селекции*. 2018. Т. 22. № 5. С. 507–514. <https://doi.org/18699/VJ18.388>
17. *Metakovsky E.V.* Gliadin allele identification on common wheat. II. Catalogue of gliadin allele in common wheat // *J. Genet. & Breed.* 1991. V. 45. P. 325–344.
18. *Козуб Н.А., Созинов И.А., Собко Т.А. и др.* Идентификация ржаных транслокаций у сортов озимой мягкой пшеницы Богданка и Синтетик // *Науч. ведомости. Серия Естественные науки*. 2010. Т. 15. № 12. С. 47–54.
19. *Singh N.K., Shepherd K.W., McIntosh R.A.* Linkage mapping of genes for resistance to leaf, stem and stripe rusts and ω -secalins on the short arm of rye chromosome 1R // *Theor. Appl. Genet.* 1990. V. 80. P. 609–616. <https://doi.org/10.1007/s00122-004-1807-5>
20. *Мельникова Е.Е., Букреева Г.И., Беспалова Л.А. и др.* Динамика генетического разнообразия сортов и линий мягкой пшеницы краснодарской селекции по аллелям глиадинокодирующих локусов // *Достижения науки и техники АПК*. 2016. Т. 30. № 3. С. 51–53.
21. *Ren T.H., Yang Z.J., Yan B.J. et al.* Development and characterization of a new 1BL.1RS translocation line

- with resistance to stripe rust and powdery mildew of wheat // *Euphytica*. 2009. V. 169. P. 207–213. <https://doi.org/10.1007/s10681-009-9924-5>
22. Qi W., Tang Y., Zhu W. et al. Molecular cytogenetic characterization of a new wheat-rye 1BL/1RS translocation line expressing superior stripe rust resistance and enhanced grain yield // *Planta*. 2016. V. 244. P. 405–416. <https://doi.org/10.1094/PDIS-93-2-0124>
 23. Maraci Ö., Özkan H., Bilgin R. Phylogeny and genetic structure in the genus *Secale* // *PLoS One*. 2018. V. 13(7). <https://doi.org/10.1371/journal.pone.0200825>
 24. Landjeva S., Korzun V., Tsanev V. et al. Distribution of the wheat-rye translocation 1RS.1BL among bread wheat varieties of Bulgaria // *Plant Breeding*. 2006. V. 125. P. 102–104. <https://doi.org/10.1111/j.1439-0523.2006.01142.x>
 25. Yediay F.E., Baloch F.S., Kilian B., Özkan H. Testing of rye-specific markers located on 1RS chromosome and distribution of 1AL.RS and 1BL.RS translocations in Turkish wheat (*Triticum aestivum* L., *T. durum* Desf.) varieties and landraces // *Genet. Res. Crop Evol.* 2010. V. 57. P. 119–129. <https://doi.org/10.1007/s10722-009-94569>
 26. Трубочеева Н.В., Россеева Л.П., Белан И.А. и др. Особенности сортов яровой мягкой пшеницы Западной Сибири, несущих пшенично-ржаную транслокацию 1RS.1BL // *Генетика*. 2011. Т. 47. № 1. С. 18–24.
 27. GRIS. Genetic Resources Information System for Wheat and Triticale. <http://wheatpedigree.net/> (Last update: 2017-01-09).
 28. Каталог. Сорты и гибриды КНИИСХ им. П.П. Лукьяненко. Краснодар: 2009. 67 с.
 29. Каталог. Сорты и гибриды НЦЗ им. П.П. Лукьяненко. Краснодар: 2021. 135 с.
 30. Государственный реестр селекционных достижений, допущенных к использованию. Т. 1. Сорты растений. М.: 2009 и 2021. <https://reestr.gossortrf.ru/>.
 31. Ren T.H., Chen F., Yan B.Ju. et al. Genetic diversity of wheat-rye 1BL.1RS translocation lines derived from different wheat and rye sources // *Euphytica*. 2012. V. 183. P. 133–146. <https://doi.org/10.1007/s10681-011-0412-3>
 32. Созинов А.А., Новосельская А.Ю., Лушникова А.А., Боданов Ю.Ф. Цитолого-биохимический анализ сортов мягкой пшеницы с замещениями и транслокациями 1В/1R в кариотипе // *Цитология и генетика*. 1987. Т. 21. № 4. С. 256–261.
 33. McIntosh R.A., Yamazaki Y., Dubcovsky J. Catalogue of gene symbols for wheat // 11th Int. Wheat Genet. Symp. 24–29 August 2008. Brisbane, Australia: 2008. 137 p.
 34. He Zhong-Hu. High molecular weight glutenin subunit composition of Chinese bread wheats // *Euphytica*. 1992. V. 64. № 1. P. 11–20. <https://doi.org/10.1002/cche.10290>
 35. Hazen S.P., Zhu L., Kim H.-S. et al. Genetic diversity of winter wheat in Shaanxi province, China, and other common wheat germplasm pools // *Genet. Resour. Crop Evol.* 2002. V. 49. № 4. P. 437–445. <https://doi.org/10.1023/A:1020670013249>
 36. Булойчик А.А., Долматович Т.В., Борзяк В.С., Волуевич Е.А. Молекулярная идентификация и эффективность генов *Lr26/Pm8* в сортообразцах мягкой пшеницы (*Triticum aestivum*) // *Весці НАН Беларусі*. 2014. Т. 2. С. 60–63.
 37. Козуб Н.А., Созинов И.А., Созинов А.А. Сопряженность 1BL/1RS-транслокации с качественными и количественными признаками у мягкой пшеницы *T. aestivum* // *Цитология и генетика*. 2001. Т. 5. С. 74–80.
 38. Lukaszewski A.J. Manipulation of the 1RS.1BL translocation in wheat by induced homoeologous recombination // *Crop Sci.* 2000. V. 40. P. 216–225.
 39. Lukaszewski A.J. Introgressions between wheat and rye // *Alien Introgression in Wheat*. Cham, Switzerland. Springer, 2015. P. 163–189. https://doi.org/10.1007/978-3-319-23494-6_7
 40. Белан И.А., Россеева Л.П., Россеев В.М. и др. Изучение хозяйственно ценных и адаптивных признаков у линий сорта яровой мягкой пшеницы Омская 37, несущих транслокации 1RS.1BL и 7DL-7Ai // *Вавилов. журн. генет. и селекции*. 2012. Т. 16. № 1. С. 178–186.
 41. Шаманин В.П., Морзунов А.И., Петуховский С.Л. и др. Селекция яровой мягкой пшеницы на устойчивость к стеблевой ржавчине в Западной Сибири. Омск: ОмГАУ им. П.А. Столыпина, 2015. 175 с.
 42. Kozub N.O., Sozinov I.O., Chaika V.M. et al. Changes in allele frequencies at storage protein loci of winter common wheat under climate change // *Cytol. and Genet.* 2020. V. 54. № 4. P. 305–317. <https://doi.org/10.3103/S0095452720040076>
 43. Kumlay A.M., Baenziger P.S., Gill K.S. et al. Understanding the effect of rye chromatin in bread wheat // *Crop Sci.* 2003. V. 43. № 5. P. 1643–1651. <https://doi.org/10.2135/cropsci2003.1643>
 44. Liu H., Tang H., Ding P. et al. Effects of the 1BL/1RS translocation on 24 traits in a recombinant inbred line population // *Cereal Res. Commun.* 2020. V. 48. P. 225–232. <https://doi.org/10.1007/s42976-020-00027-y>
 45. Jang J.H., Jung W.J., Kim D.Y., Seo Y.W. cDNA-AFLP analysis of 1BL.1RS under water-deficit stress and development of wheat-rye translocation-specific markers // *N. Zel. J. Crop Hort. Sci.* 2017. V. 45. P. 150–164. <https://doi.org/10.1080/01140671.2016.1269018>
 46. Moskal K., Kowalik S., Podyma W. et al. The pros and cons of rye chromatin introgression into wheat genome // *Agronomy*. 2021. V. 11. P. 456. <https://doi.org/10.3390/agronomy11030456>

Dynamics of Rye Translocations Frequency in Russian Common Wheat *Triticum aestivum* L. Cultivars

A. V. Fisenko^{a, *}, O. A. Lyapunova^b, E. V. Zuev^b, and A. Yu. Novoselskaya-Dragovich^{a, **}

^aVavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia

^bVavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, 190000 Russia

*e-mail: fisenko800@mail.ru

**e-mail: dragova@mail.ru

The widespread using 1RS/1BL and 1RS/1AL translocation cultivars in wheat breeding was due to their high complex resistance provided by genes on 1RS chromosomes. Numerous data on the emergence of new virulent pathogen races that caused the loss of protective properties by genes on 1RS make it relevant to analyse the prospects for using both rye translocations in modern breeding programs. The use of various genetic sources of the 1RS rye chromosome in introgressive hybridization poses the problem of studying the diversity of 1RS linkage groups. We have studied the rye translocations dynamics in 240 genotypes of the wheat cultivars from 3 breeding centers of the Russian Federation. Multiple alleles of storage protein genes specific for 1RS were used to mark 1RS. The effectiveness of 1RS resistance genes to a number of pathogens, in particular to powdery mildew and stem rust, has been established. The grain quality dynamics of the 1RS/1BL cultivars has been traced. A statistically significant grain quality improvement of winter wheat cultivars since the 2000s has been confirmed. The spring cultivars quality does not depend on the presence of rye translocations in the genotypes. Accumulation of cultivars with TR:1RS/1BL observed in all three centers. A high allelic diversity of the secalin coding locus on 1RS chromosome was shown. However, all cultivars with rye translocations, regardless of 1RS origin, carried the same allele. Possible reasons for this uniformity are discussed. Translocation 1RS/1AL was not detected in the studied cultivars. Possible reasons are discussed.

Keywords: common wheat, rye translocation, storage protein genes.

АНАЛИЗ ГЕНЕТИЧЕСКИХ ФАКТОРОВ СПОРАДИЧЕСКИХ СЛУЧАЕВ ШИЗОФРЕНИИ В СЕМЕЙНЫХ ТРИО С ИСПОЛЬЗОВАНИЕМ МЕТОДА ПОЛНОГЕНОМНОГО СЕКВЕНИРОВАНИЯ

© 2023 г. Т. В. Андреева^{1, 2, *}, Ф. А. Афанасьев², Ф. Е. Гусев^{2, 3}, А. Д. Патрикеев²,
С. С. Кунижева^{2, 3}, Е. И. Рогаев^{3, 4, 5, **}

¹Центр генетики и генетических технологий, Московский государственный университет
им. М.В. Ломоносова, Москва, 119234 Россия

²Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

³Центр генетики и наук о жизни, Научно-технологический университет “Сириус”,
Краснодарский край, пгт. Сириус, 354340 Россия

⁴Московский государственный университет им. М.В. Ломоносова, Москва, 119234 Россия

⁵Медицинская школа Чан Массачусетского университета, департамент психиатрии, Шрусбери, МА 01545 США

*e-mail: an_tati@vigg.ru

**e-mail: evgeny.rogaev@umassmed.edu

Поступила в редакцию 07.12.2022 г.

После доработки 03.02.2023 г.

Принята к публикации 06.02.2023 г.

Шизофрения – распространенное психическое заболевание, наследственная природа которого подтверждена многочисленными исследованиями. В настоящее время выявлено более сотни генетических локусов, ассоциированных с шизофренией, также идентифицированы редкие варианты в генах и хромосомные перестройки, связанные с семейными случаями заболевания. Однако не всегда удается определить наследственную природу патологии, многие случаи шизофрении являются спорадическими, а генетическая причина таких случаев остается неизвестна. С использованием данных полногеномного секвенирования трех семейных трио российского происхождения со спорадическими формами шизофрении мы провели поиск редких потенциально патогенных вариантов в кодирующих и регуляторных локусах генома, включая *de novo* и компаундные мутации. Также провели оценку полигенного риска развития шизофрении с использованием распространенных полиморфных маркеров. В результате проведенного анализа были показаны генетическая гетерогенность спорадических форм шизофрении, а также потенциальный вклад редких замен в генах, связанных с метаболизмом глутамата и инозитолфосфата, в развитие спорадических случаев шизофрении.

Ключевые слова: шизофрения, полногеномное секвенирование, полигенный риск, *de novo* варианты.

DOI: 10.31857/S0016675823060024, **EDN:** SRPBKR

Шизофрения является одним из наиболее распространенных в мире психических заболеваний, которым страдает до 1–2% всего населения [1]. Роль генетических факторов в развитии шизофрении давно известна, а многочисленные исследования свидетельствуют о высокой генетической гетерогенности заболевания.

Результаты полногеномного анализа генетических ассоциаций (GWAS) предполагают существование сотен и даже тысяч генетических маркеров в геноме человека, потенциально связанных с шизофренией, каждый из которых вносит незначительный индивидуальный вклад в развитие патологии, тогда как комбинация этих маркеров приводит к болезни [2–5]. Оценка предрасположенности к шизофрении на основе расчетов полигенного

риска позволяет объяснить до 7% случаев шизофрении [6, 7]. Около 5% случаев шизофрении связаны с хромосомными перестройками [8, 9]. Исследования последних лет свидетельствуют о существенной роли *de novo* мутаций в развитии шизофрении, что подтверждается более высокой частотой развития патологий у детей, родившихся у возрастных отцов [10, 11].

Высокая генетическая гетерогенность шизофрении, а также пересечение ее симптомов и уже выявленных генетических факторов с другими психическими патологиями затрудняют идентификацию первичных ключевых механизмов развития заболевания. Кроме того, большинство выявленных генетических маркеров шизофрении до сих пор еще не имеют прямого клинического приме-

Таблица 1. Характеристика образцов, использованных в работе

Семья (этническое происхождение)	Образец (член семьи)	Диагноз	Пол	Возраст на момент обследования, лет
Семья 1 (балкарцы)	SF1 (мать)	Здорова	Ж	55
	SF2 (отец)	Здоров	М	61
	SF3 (пробанд)	Шизофрения (F20.024)	Ж	26
Семья 2 (русские)	SF4 (отец)	Здоров	М	68
	SF5 (мать)	Здорова	Ж	61
	SF6 (пробанд)	Шизофрения (F20.014)	Ж	32
Семья 3 (русские)	SF7 (отец)	Здоров	М	55
	SF8 (мать)	Здорова	Ж	51
	SF9 (пробанд)	Шизофрения (F20.02)	Ж	23

нения. Важное значение для разработки эффективных методов терапии и профилактики данной болезни имеет оценка наиболее распространенных генетических факторов заболевания в той или иной популяционной группе. Целью данной работы является оценка вклада различных генетических факторов в развитие спорадических форм шизофрении в семейных трио из российских популяций.

МАТЕРИАЛЫ И МЕТОДЫ

В исследовании были использованы три ядерные семьи, включающие здоровых родителей и ребенка с диагнозом шизофрения (табл. 1). Исследования были одобрены этическим комитетом ИОГен РАН; от всех членов семей, включенных в исследование, было получено письменное согласие на участие в исследовании. Диагноз пациентов установлен врачами ГКУЗ ПНД МЗ КБР (семьи 1 и 2) и ФГБУ НЦПЗ (семья 3) в соответствии с критериями МКБ-10, оценка психического состояния здоровых родственников была проведена врачом, обследовавшим пробандов.

ДНК выделяли из образцов периферической венозной крови с использованием набора реагентов QIAamp DNA Blood (Qiagen). 1 мкг полученной ДНК использовали для приготовления фрагментных геномных библиотек с помощью набора реагентов TruSeq DNA Sample Preparation Kit (Illumina). Подготовленные геномные библиотеки секвенировали на платформе HiSeq2000/2500 (Illumina) в режиме парных прочтений. Полученные прочтения были картированы на референсный геном человека GRCh37/hg19 с помощью *bwa mem* [12]. Для поиска генетических вариантов использовали *HarplotypeCaller* из пакета программ *GATK v4* [13], с последующей фильтрацией выявленных вариантов с помощью *ApplyRecalibration*, аннота-

цию проводили с помощью *VEP* [14], предсказание патогенности выявленных миссенс-вариантов проводили с помощью *SIFT* и *PolyPhen2* [15, 16]. Для поиска структурных вариаций использовали *SpeedSeq* [17] и *forestSV* [18], генотипирование структурных вариантов проводили с помощью *SV2* [19].

Для проверки родственных связей в каждой семье использовали набор из 52 идентификационных маркеров и рассчитывали комбинированный индекс родства (CPI) [20].

Для идентификации вероятных генетических факторов, приводящих к развитию шизофрении у исследованных пациентов, проводили расчет полигенного риска, а также поиск редких потенциально патогенных вариантов, включая *de novo* варианты, гомозиготные и компаундные мутации.

Для оценки полигенного риска развития шизофрении у членов каждой семьи были использованы данные полногеномного анализа ассоциаций, полученные на основе анализа 46 независимых выборок образцов европейского происхождения [2]. При расчете использовали 93 однонуклеотидных полиморфных маркера, статистика по которым представлена в базе *GWAS Catalog* [21]. В качестве контрольной группы при анализе полигенного риска использовали генотипы 503 индивидов европейского происхождения из проекта “1000 геномов” [22]. Генотипы девяти исследуемых образцов были объединены с генотипами полученной референсной панели, после чего для каждого образца был рассчитан полигенный риск (PRS) по следующей формуле:

$$PRS_{\text{образца}} = \sum_{i=1}^{N_{\text{SNP}}} \beta_{\text{SNP}i} n_{\text{риск-аллель}}$$

где N_{SNP} — это количество анализируемых генетических вариантов; $\beta_{\text{SNP}i}$ — это размер эффекта ге-

нетического варианта i , полученный из статистики в GWAS Catalog; $n_{\text{риск-аллель}}$ — это количество риск-аллелей генетического варианта i в генотипе данного индивида. На основе полученных значений полигенного риска в популяционной референсной панели из проекта “1000 геномов” были рассчитаны стандартизированные оценки (Z -оценки) полигенного риска для каждого пациента с шизофренией и его родителей.

При поиске потенциально патогенных вариантов рассматривали только редкие, встречающиеся в популяциях с частотой менее 1% [23]. Для отбора *de novo* вариантов (однонуклеотидные замены и короткие инсерции/делеции) применяли следующие фильтры: референсный гомозиготный генотип у родителей и гетерозиготный у ребенка; количество прочтений, подтверждающих мутацию, у родителей должно было быть не больше 3%, у ребенка от 25 до 75%. Кроме того, при отборе *de novo* мутаций были подобраны такие значения параметров программы GATK ApplyRecalibration, чтобы число предсказанных вариантов для индивидуального генома было не менее 70, что соответствует среднему числу новых замен на геном человека [24]. Отбор структурных *de novo* вариантов проводили с применением следующих фильтров: гомозиготный референсный генотип у каждого родителя и гетерозиготный генотип у ребенка, а также отсутствие прочтений, содержащих структурный вариант у обоих родителей [25]. Все выявленные *de novo* варианты дополнительно были визуально проверены в IGV [26].

Поиск *de novo* вариантов в регуляторных участках генома проводили с использованием эпигенетических профилей, представленных в базе данных NIH Roadmap Epigenomics (<http://www.roadmapepigenomics.org>). Были использованы области генома, соответствующие промоторам и энхансерам в клетках тканей мозга. Для анализа коротких инсерционно-делеционных вариантов в регуляторных областях отбирали те, которые были предсказаны с помощью алгоритма DeepSea [27] как нарушающие регуляторные свойства ($p < 0.01$). Изменения в сайтах связывания в случае однонуклеотидных замен анализировали с помощью сервиса PERFECTOS-APE и базы данных HOC-OMOCO для человека [28]. Для идентификации генов, в энхансерах которых были выявлены замены, использовали базу данных GeneLoc [29].

При анализе потенциально патогенных рецессивных вариантов отбирали такие, которые присутствуют у родителей в гетерозиготном состоянии, а у ребенка — в гомозиготном. Для поиска компаундных вариантов был написан скрипт, который отбирает гетерозиготные варианты, присутствующие у одного из родителей и ребенка в гетерозиготном состоянии и отсутствующие у второго родителя, после пересечения двух таких

списков были получены компаундные варианты, которые ребенок унаследовал от разных родителей.

Дополнительно проводили поиск редких вариантов в кодирующих участках генов, для которых ранее была выявлена ассоциация с шизофренией [2, 30–32]. В анализ включали только варианты в генах, экспрессирующихся в тканях мозга [33].

Валидацию *de novo* вариантов в кодирующих участках генов проводили с помощью секвенирования методом Сэнгера, анализ продуктов амплификации локусов, содержащих короткие инсерции/делеции, проводили на приборе Bioanalyzer 2100 (Agilent) с использованием набора реагентов High Sensitive DNA Chip (Agilent). Олигонуклеотидные праймеры для ПЦР-амплификации (доступны по запросу) подбирали с использованием программы Primer3 [34]. Для амплификации использовали набор реагентов GenPak PCR Core (ООО “Лаборатория Изоген”, Россия), реакцию секвенирования проводили с помощью набора BigDye terminator v3.1 cycle sequencing kit (Thermo Fisher Scientific).

РЕЗУЛЬТАТЫ

В табл. 2 представлена общая статистика по результатам секвенирования геномов представителей трех семейных трио со спорадическими случаями шизофрении. В результате геномного секвенирования для каждого образца было получено 32–50× геномное покрытие, среднее покрытие 97.5% кодирующих последовательностей генома (экзонов) для всех образцов составило >5×.

Анализ родства

Вероятность отцовства и материнства по результатам анализа 52 несцепленных маркеров составила более 0.9996 для всех трех семей, что подтверждает родственные связи внутри семей.

Полигенный риск развития шизофрении

Результаты оценки полигенного риска развития шизофрении у исследованных пациентов и их родителей представлены на рис. 1. В двух семьях (семья 1 и семья 2) полигенный риск развития шизофрении у ребенка оказался выше, чем у его здоровых родителей. При этом в семье 1 полигенный риск ребенка выше, чем у 97.5% в европейской популяции (рис. 1,а). В семье 3 полигенный риск развития шизофрении ребенка был выше такового у его матери, но не отца (рис. 1,б).

De novo мутации в кодирующих и регуляторных последовательностях

В результате поиска *de novo* мутаций были выявлены 70, 103 и 102 однонуклеотидные замены в

Таблица 2. Общая статистика по полученным данным полногеномного секвенирования

Образец	Общее число полученных прочтений, млн пар	Прочтения, картированные на геном человека GRCh37, %	Среднее покрытие генома
SF1	719	99.7	44×
SF2	610	99.7	34×
SF3	630	99.6	36×
SF4	593	99.6	32×
SF5	843	99.7	50×
SF6	848	99.8	49×
SF7	646	99.6	37×
SF8	606	99.7	34×
SF9	747	99.5	37×

Таблица 3. Варианты *de novo*, выявленные у пациентов с шизофренией и отсутствующие у их здоровых родителей

Семья (пробанд)	Геномные координаты	Референсный аллель	Альтернативный аллель	Ген	Аминокислотная замена	Частота в европейских популяциях*
Семья 2 (SF6)	19:422205	G	A	<i>SHC2</i>	p.Leu521Phe	—
Семья 3 (SF9)	7:138602846	G	A	<i>KIAA1549</i>	p.Ala509Val	—
	19:48305654	C	T	<i>TPRX1</i>	p.Gly205Asp	0.1033
	17:77807917	T	TGCCGCC	<i>CBX4</i>	p.Ala509_Ala510dup	0.08477

* По базе данных GnomAD v2.1.1.

геномах пациентов с шизофренией из семьи 1, 2 и 3 соответственно; структурных вариаций, которые отсутствовали у родителей, но присутствуют у пробандов, не было обнаружено ни в одной семье. Новые мутации, найденные в кодирующих областях генома пациентов со спорадическими случаями шизофрении, представлены в табл. 3, две из них встречаются в европейских популяциях с высокой частотой, поэтому были исключены из дальнейшего рассмотрения.

Секвенирование по методу Сэнгера подтвердило присутствие мутации p.Ala509Val в гене *KIAA1549*, отсутствующей у здоровых родителей, у пробанда из семьи 3, а также мутации p.Leu521Phe в гене *SHC2* у пациента с шизофренией из семьи 2 (рис. 2, а, б).

Выявленная у пациента из семьи 2 однонуклеотидная замена в гене *SHC2* (19:422205 G>A, p.Leu521Phe) приводит к замене лейцина на фенилаланин в консервативной среди всех хордовых позиций белка. Мутация затрагивает функциональный домен SH2 продукта гена, отвечающий за его взаимодействие с другими белками и передачу им сигналов. Следует отметить, что аналогичная замена (p.Leu43Phe) в SH2-доме другого белка (продукт гена *PTPN11*) является пато-

генной и приводит к развитию синдрома Нунана [35], что свидетельствует о важной функциональной роли данной аминокислотной позиции в белке и всего белкового домена. Ген *SHC2* экспрессируется в коре больших полушарий мозга, а ранее были высказаны предположения о его возможной связи с шизофренией как белка, участвующего в процессах синаптической передачи сигналов [36].

De novo вариант, выявленный в семье 3 в гене *KIAA1549* (7:138602846 G>A, c.1526C>T, p.Ala509Val), отсутствует в базе данных gnomAD. Позиция аминокислоты не обладает консервативностью: замены в этой позиции по сравнению с белком человека встречаются у 23% млекопитающих. Ген *KIAA1549* экспрессируется в различных тканях, включая ткани мозга. Его функции до сих пор неизвестны, хотя мутации в этом гене описаны при аутосомно-рецессивных формах пигментного ретинита [37].

В результате проведенного поиска *de novo* вариантов в регуляторных участках генома, активных в нейрональных клетках мозга человека, были выявлены две замены, значимо изменяющие сайты связывания транскрипционных факторов согласно базе HOCOMOCO (<https://hocomoco11.autosome.org>). У пациентки из семьи 1 идентифици-

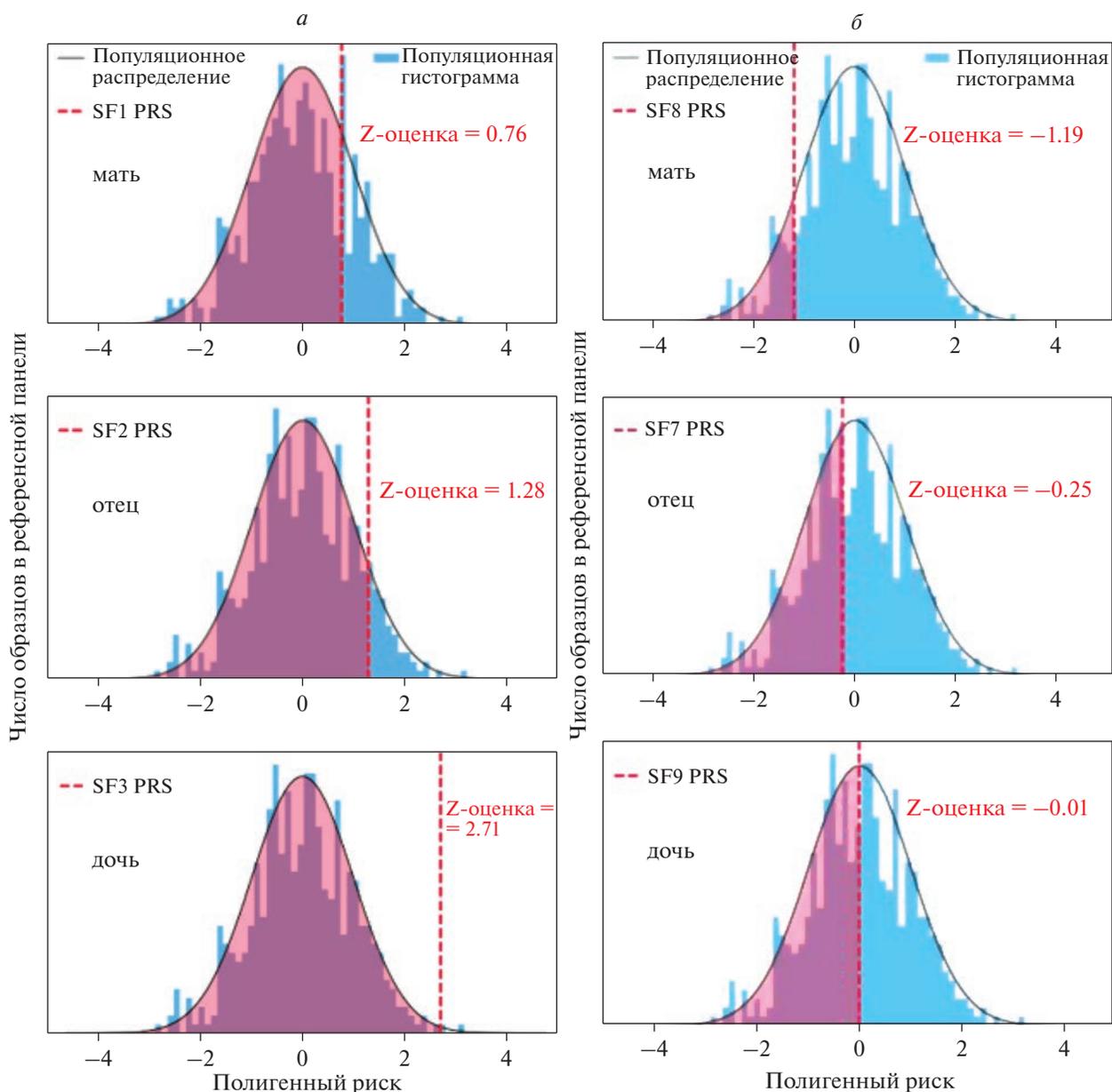


Рис. 1. Оценка полигенного риска развития шизофрении у пациентов с шизофренией и их здоровых родителей. *a* – семья 1, *б* – семья 2, *в* – семья 3. Z-оценка для исследованных членов семей обозначена красным пунктиром. Столбчатая диаграмма синего цвета соответствует индивидуальным оценкам полигенного риска у индивидов европейского происхождения из проекта “1000 геномов”.

рована мутация в энхансере гена *TSPAN9* (12:3327271, T>C), а у пациентки из семьи 3 была обнаружена новая мутация в промоторной области гена *CNTRL* (9:123835451, A>C). Обе нуклеотидные замены отсутствуют в базе данных GnomAD. В настоящее время в литературе отсутствует информация о возможном участии двух этих генов в функционировании нервной системы или развитии нейропсихических патологий, поэтому их возможный вклад в патогенез шизофрении требует дополнительных исследований.

Рецессивные и компаундные мутации

В результате поиска потенциально патогенных редких аутомно-рецессивных мутаций в кодирующих областях генома была выявлена единственная такая вариация – замена в гене *C1orf87* (1:60474407 G>A, p.Arg3Cys) в семье 3. Функции этого гена в настоящее время неизвестны, информация о его возможной ассоциации с шизофренией или другими психическими патологиями также отсутствует. Гомозиготных вариантов в кодирующих участках генома у пациентов с ши-

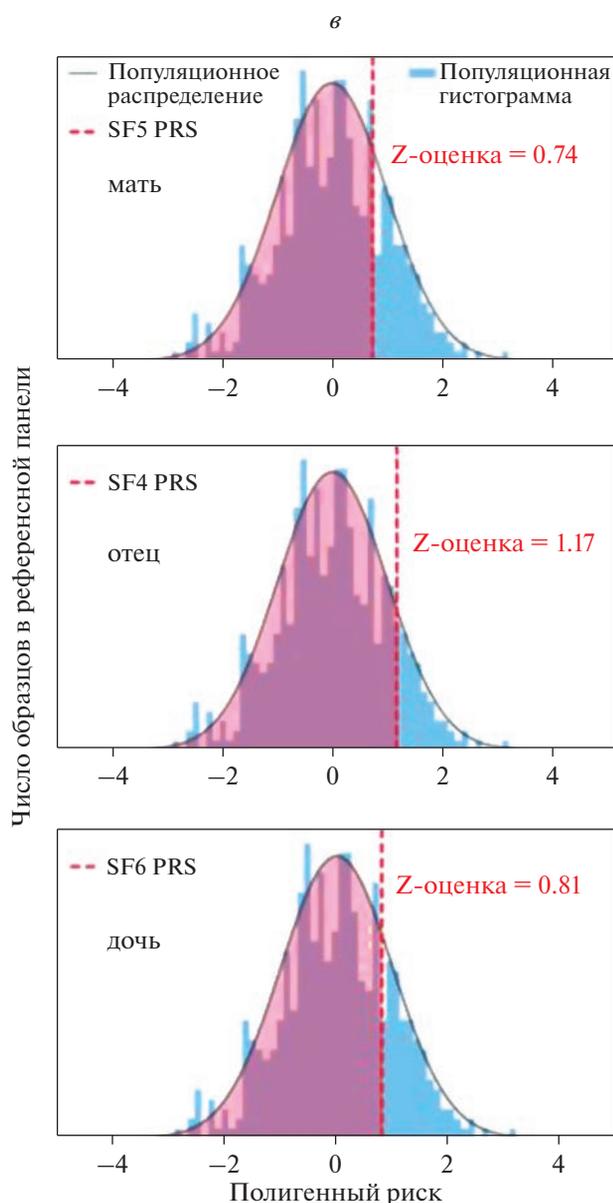


Рис. 1. Окончание.

зофренией из семей 1 и 2, которые присутствуют у здоровых родителей в гетерозиготном состоянии, выявлено не было.

У пробанда из семьи 2 были выявлены компаундные мутации в гене *INPP5J*. Замена 22:31522435 G>A (p.Gly449Ser) была унаследована от отца, а вариант 22:31521105 C>T (p.Ala127Val) был получен пациентом от матери. Продукт гена участвует в метаболизме инозитолтрифосфата. Обе мутации расположены в сайтах белка, которые не являются эволюционно-консервативными, а алгоритмы SIFT и PolyPhen2 предсказывают их как непатогенные.

Редкие варианты в генах, ассоциированных с шизофренией

В результате анализа генов, для которых ранее была показана ассоциация с шизофренией, в исследованных семьях были выявлены варианты в кодирующих областях известных генов, потенциально связанных с шизофренией (табл. 4).

В семье 1 было выявлено пять вариантов, приводящих к аминокислотной замене в генах или локусах, ассоциированных с шизофренией. Четыре из них пациентка получила от матери, один вариант (в гене *SLC18A1*) был унаследован от отца. Ген *SLC18A1* кодирует транспортер моноаминов (серотонина, дофамина и норэпинефрина) и расположен в локусе на хромосоме 18, ассоциированном с различными психическими патологиями, включая биполярные расстройства и шизофрению [38, 39]. Выявленная в семье 1 нуклеотидная замена расположена в сайте сплайсинга, изменения которого потенциально затрагивают почти все транскрипты гена и могут привести к снижению экспрессии гена. В недавних исследованиях было показано, что варианты в гене *SLC18A1* находились под действием положительного отбора в процессе эволюции человека, и высказано предположение о связанной с этим роли продукта этого гена в развитии специфических для человека психических заболеваний [40].

Вариант в гене *CIT* (p.Arg831Gln), унаследованный пациенткой с шизофренией от матери, расположен в консервативном среди млекопитающих и птиц сайте белка. Мутации в этом гене описаны при аутосомно-рецессивных формах микроцефалии, кроме того показано эпистатическое взаимодействие полиморфных маркеров в генах *CIT* и *DISC1* (ген, ассоциированный с шизофренией по результатам нескольких независимых исследований [41]) как при развитии шизофрении, так и при когнитивных процессах у психически здоровых людей [42].

Уникальная нуклеотидная замена в гене *WFS1* (p.Pro675Leu), выявленная у пациентки с шизофренией и ее здоровой матери, отсутствует в популяционной базе данных GnomAD. Аминокислотная замена расположена в консервативной среди млекопитающих области белка вольфрамина. Мутации в гене *WFS1* связаны с аутосомно-рецессивным нейро-дегенеративным синдромом Вольфрама, основными симптомами которого являются сахарный диабет, атрофия зрительного нерва и нарушения слуха. В целом ряде работ обсуждается связь мутаций в гене *WFS1* с различными психическими заболеваниями, включая шизофрению [43, 44].

По результатам проведенного ранее полногеномного анализа ассоциаций полиморфный вариант в гене *NOS1* (p.Asp394Ala), выявленный в семье 1, не ассоциирован с шизофренией [2].

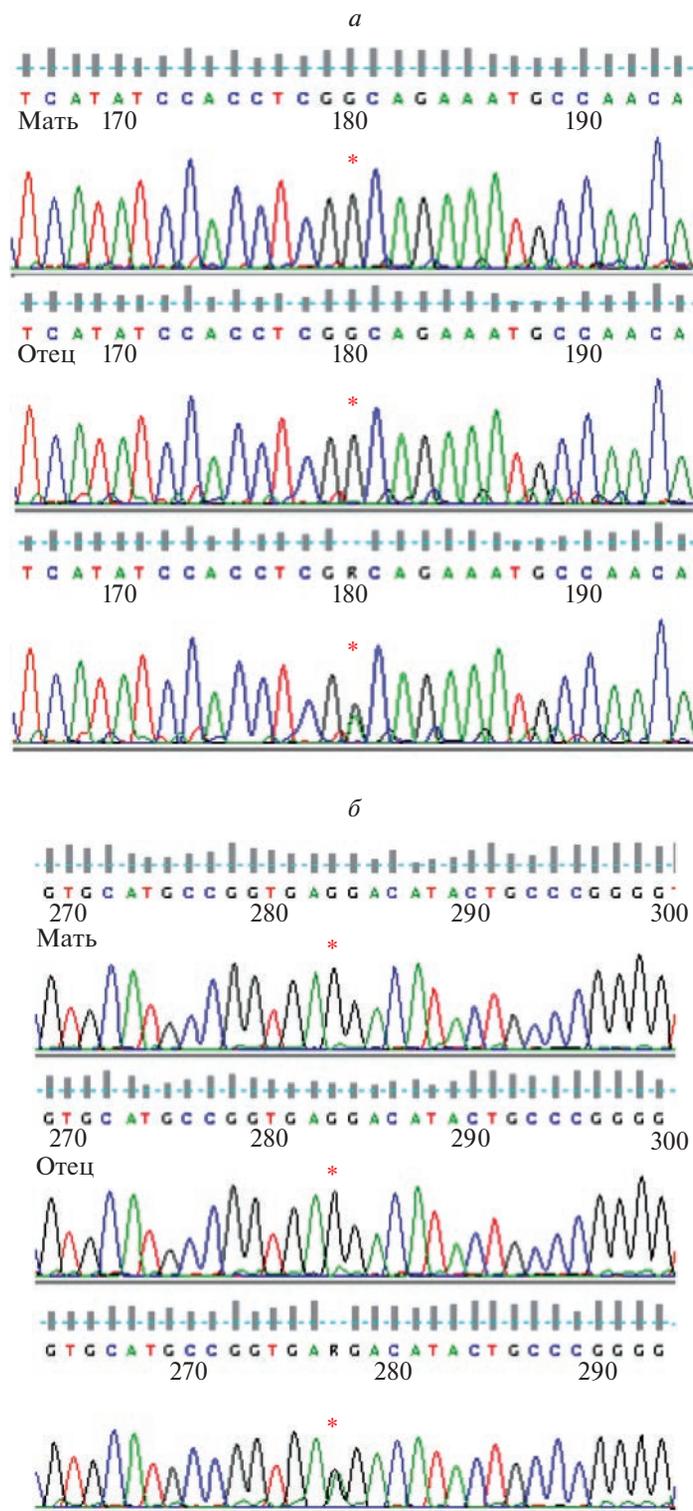


Рис. 2. Результаты секвенирования методом Сэнгера вариантов в генах *KIAA1549* (а) и *SHC2* (б) у пациентов с шизофренией (внизу) и их здоровых родителей. Звездочкой отмечено положение выявленной нуклеотидной замены.

Единственный вариант, выявленный в кодирующих участках генов, ассоциированных с шизофренией, в семье 2 у пациентки и ее матери приводит к аминокислотной замене в гене *CYFIP1*

(p.Glu469Asp). Данная замена расположена в высококонсервативном среди хордовых сайте белка, а в базе данных GnomAD присутствует только у одного индивида европейского происхождения.

Таблица 4. Редкие варианты в генах, связанных с шизофренией, выявленные у пациентов из исследованных семей

Семья (образцы)	Нуклеотидная замена	Ген	Аминокислотная замена	Частота в европейских популяциях*	SIFT	PolyPhen2
Семья 1 (SF1, SF3)	12:117724018 T>G	<i>NOS1</i>	p.Asp394Ala	0.007437	0.15	0.014
Семья 1 (SF1, SF3)	12:120195263 C>T	<i>CIT</i>	p.Arg831Gln	0	0.46	0.001
Семья 1 (SF2, SF3)	8:20038604 A>T	<i>SLC18A1</i>	Сайт сплайсинга	0.00006	—	—
Семья 1 (SF1, SF3)	4:6303546 C>T	<i>WFS1</i>	p.Pro675Leu	—	0*	0.996 [#]
Семья 2 (SF5, SF6)	15:22954257 G>T	<i>CYFIP1</i>	p.Glu469Asp	0.0000088	0.02*	0.247
Семья 3 (SF7, SF9)	6:146720043 T>C	<i>GRM1</i>	p.Val623Ala	0.000008791	0*	1 [#]
Семья 3 (SF8, SF9)	21:30968856 A>G	<i>GRIK1</i>	p.Val414Ala	0.000008842	0.02*	0.003
Семья 3 (SF7, SF9)	10:121581975 A>G	<i>INPP5F</i>	p.Tyr664Cys	—	0.02*	1 [#]

* SIFT. deleterious. [#] PolyPhen. probably damaging.

Ранее было показано, что полиморфизм в гене *CYFIP1* ассоциирован с шизофренией в китайской популяции [45], кроме того показан доза-зависимый вклад продукта гена в поддержание нормальных синаптических функций и развитие различных психических патологий [46, 47].

У пациентки из семьи 3 были обнаружены мутации в двух различных генах, кодирующих глутаматные рецепторы. Первый вариант — замена в гене *GRM1* (p.Val623Ala), кодирующем метаболический глутаматный рецептор mGluR1, был получен ею от здорового отца, вторая замена — в гене *GRIK1* (p.Val414Ala), кодирующем ионотропный глутаматный рецептор GluK1, — от здоровой матери. Обе аминокислотные замены в продуктах генов *GRM1* и *GRIK1* затрагивают эволюционно-консервативную среди позвоночных позицию соответствующего рецептора, в базе данных GnomAD присутствуют только по одному носителю каждого варианта. Ранее у пациентов с шизофренией были выявлены различные мутации в гене *GRM1* [48, 49], а потенциальный вклад изменений в гене *GRIK1* и уровня его экспрессии в развитие шизофрении был показан как на основе генетического анализа, так и в функциональных исследованиях [50–52].

Кроме того, у пациентки с шизофренией из семьи 3 и ее здорового отца была обнаружена уникальная замена в гене *INPP5F* (p.Tyr664Cys), отсутствующая в базе данных GnomAD и приводящая к аминокислотной замене в консервативном сайте функционального домена белка, участвующего в метаболизме инозитолтрифосфата.

ОБСУЖДЕНИЕ

В результате проведенного нами полногеномного анализа данных секвенирования трех семейных трио со спорадическими случаями шизофрении впервые был оценен вклад различных гене-

тических факторов в развитие спорадических форм шизофрении в семьях российского происхождения. Оценка полигенного риска развития шизофрении в последнее время рассматривается как одно из перспективных приложений для оценки риска развития заболевания, хотя обсуждаются значительные ограничения этого метода, связанные, например, с выбором подходящей контрольной группы [53], а также с отсутствием значимого вклада этого метода для диагностики [54]. В исследованных нами семьях вероятный вклад полигенного риска в развитие болезни был показан в двух семьях из трех (семья 1 и семья 2) — риск развития шизофрении у пробандов выше, чем у их здоровых родителей. Причем в семье 1 полигенный риск у пациентки оказался не только выше такового у обоих родителей, но и выше, чем значение риска у 97.5% индивидов европейского происхождения. Следует отметить, что члены семьи 1 относят себя к балкарцам, проживающим на ограниченной территории в ущельях и предгорьях Северного Кавказа, численность которых составляет около 125 тыс. человек (по данным переписи 2020 г.), поэтому нельзя исключить наличие дальних родственных связей между родителями пациентки, а следовательно гомозиготное состояние многих генетических маркеров, потенциально участвующих в развитии патологий.

Развитие шизофрении у пациентки из семьи 3 вероятно связано с комбинацией в ее геноме нескольких потенциально патогенных редких вариантов в генах *GRM1* и *GRIK1*, полученных от разных родителей и приводящих к нарушениям в сигнальных путях, связанных с глутаматом — основным нейромедиатором в центральной нервной системе, роль которого в развитии шизофрении давно известна.

В двух исследованных семьях нами были обнаружены замены в консервативных сайтах белков, участвующих в метаболизме инозитолтрифосфа-

та. В семье 3 у пациентки и ее отца выявлена замена в консервативном сайте гена *INPP5F*, в семье 2 пациентка является носителем компаундных вариантов в гене *INPP5J*.

Инозитолтрифосфат является вторичным мессенджером, производство которого запускается по сигналу глутаматных рецепторов, необходим для передачи сигнала по глутаматному метаболическому пути. У пациентов с шизофренией было показано снижение уровня инозитола в тканях мозга [55], а также выявлена связь между уровнем инозитолтрифосфата в крови и эффективностью терапии [56]. До настоящего времени мутации в генах *INPP5F* и *INPP5J* не были описаны в связи с шизофренией, однако их прямая связь с компонентами сигнального пути глутамата и выявленные нами нарушения в их генах у пациентов с шизофренией позволяют предположить их вероятный вклад в развитие патологии.

Работа выполнена при финансовой поддержке РНФ (номер гранта 19-75-30039).

Авторы работы выражают благодарность М.М. Шаваевой (ГКУЗ ПНД МЗ КБР) и д.м.н. Л.И. Абрамовой (ФГБНУ НЦПЗ) за предоставленные образцы биологического материала пациентов и членов их семей; Т.А. Бородиной и В.Н. Рыкалинной за помощь в проведении геномного секвенирования.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников было получено информированное добровольное согласие.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

- Owen M.J., Sawa A., Mortensen P.B. Schizophrenia // *Lancet* (London, England). 2016. V. 388. № 10039. P. 86. [https://doi.org/10.1016/S0140-6736\(15\)01121-6](https://doi.org/10.1016/S0140-6736(15)01121-6)
- Ripke S., Neale B.M., Corvin A. et al. Biological insights from 108 schizophrenia-associated genetic loci // *Nature*. 2014. V. 511. № 7510. P. 421–427. <https://doi.org/10.1038/nature13595>
- Goes F.S., McGrath J., Avramopoulos D. et al. Genome-wide association study of schizophrenia in Ashkenazi Jews // *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 2015. V. 168. № 8. P. 649–659. <https://doi.org/10.1002/AJMG.B.32349>
- Ikeda M., Takahashi A., Kamatani Y. et al. Genome-wide association study detected novel susceptibility genes for schizophrenia and shared trans-populations diseases genetic effect // *Schizophr. Bull.* 2019. V. 45. № 4. P. 824–834. <https://doi.org/10.1093/SCHBUL/SBY140>
- Li Z., Chen J., Yu H. et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia // *Nat. Genet.* 2017. V. 49. № 11. P. 1576–1583. <https://doi.org/10.1038/NG.3973>
- Martin A.R., Daly M.J., Robinson E.B. et al. Predicting polygenic risk of psychiatric disorders // *Biol. Psychiatry*. 2019. V. 86. № 2. P. 97–109. <https://doi.org/10.1016/J.BIOPSYCH.2018.12.015>
- Kendler K.S. The schizophrenia polygenic risk score: to what does it predispose in adolescence? // *JAMA Psychiatry*. 2016. V. 73. № 3. P. 193–194. <https://doi.org/10.1001/JAMAPSYCHIATRY.2015.2964>
- Kato H., Kimura H., Kushima I. et al. The genetic architecture of schizophrenia: Review of large-scale genetic studies // *J. Hum. Genet.* 2023. V. 68. P. 175–182. <https://doi.org/10.1038/S10038-022-01059-4>
- Farrell M., Dietterich T.E., Harner M.K. et al. Increased prevalence of rare copy number variants in treatment-resistant psychosis // *Schizophr. Bull.* 2022. <https://doi.org/10.1093/SCHBUL/SBAC175>
- Wu Y., Liu X., Luo H. et al. Advanced paternal age increases the risk of schizophrenia and obsessive-compulsive disorder in a Chinese Han population // *Psychiatry Res.* 2012. V. 198. № 3. P. 353. <https://doi.org/10.1016/J.PSYCHRES.2012.01.020>
- Khachadourian V., Zaks N., Lin E. et al. Advanced paternal age and risk of schizophrenia in offspring—review of epidemiological findings and potential mechanisms // *Schizophr. Res.* 2021. V. 233. P. 72. <https://doi.org/10.1016/J.SCHRES.2021.06.016>
- Li H., Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform // *Bioinformatics*. 2009. V. 25. № 14. P. 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- McKenna A., Hanna M., Banks E. et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data // *Genome Res.* 2010. V. 20. № 9. P. 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McLaren W., Gil L., Hunt S.E. et al. The ensembl variant effect predictor // *Genome Biol.* 2016. V. 17. № 1. P. 1–14. <https://doi.org/10.1186/S13059-016-0974-4>
- Adzhubei I.A., Schmidt S., Peshkin L. et al. A method and server for predicting damaging missense mutations // *Nat. Methods*. 2010. V. 7. № 4. P. 248–249. <https://doi.org/10.1038/nmeth0410-248>
- Kumar P., Henikoff S., Ng P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm // *Nat. Protoc.* 2009. V. 4. № 7. P. 1073–1081. <https://doi.org/10.1038/nprot.2009.86>
- Chiang C., Leyer R.M., Faust G.G. et al. SpeedSeq: Ultra-fast personal genome analysis and interpretation // *Nat. Methods*. 2015. V. 12. № 10. P. 966–968. <https://doi.org/10.1038/nmeth.3505>
- Michaelson J.J., Sebat J. ForestSV: Structural variant discovery through statistical learning // *Nat. Methods*. 2012. V. 9. № 8. P. 819–821. <https://doi.org/10.1038/nmeth.2085>

19. *Antaki D., Brandler W.M., Sebat J.* SV2: Accurate structural variation genotyping and de novo mutation detection from whole genomes // *Bioinformatics*. 2018. V. 34. № 10. P. 1774–1777.
<https://doi.org/10.1093/BIOINFORMATICS/BTX813>
20. *Sanchez J.J., Phillips C., Børsting C. et al.* A multiplex assay with 52 single nucleotide polymorphisms for human identification // *Electrophoresis*. 2006. V. 27. № 9. P. 1713–1724.
<https://doi.org/10.1002/elps.200500671>
21. *Buniello A., MacArthur J.A.L., Cerezo M. et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019 // *Nucl. Acids Res.* 2019. V. 47. № D1. P. D1005–D1012.
<https://doi.org/10.1093/NAR/GKY1120>
22. *Sudmant P.H., Rausch T., Gardner E.J. et al.* An integrated map of structural variation in 2,504 human genomes // *Nature*. 2015. V. 526. № 7571. P. 75–81.
<https://doi.org/10.1038/nature15394>
23. *Purcell S.M., Moran J.L., Fromer M. et al.* A polygenic burden of rare disruptive mutations in schizophrenia // *Nature*. 2014. V. 506. № 7487. P. 185–190.
<https://doi.org/10.1038/nature12975>
24. *Roach J.C., Glusman G., Smit A.F.A. et al.* Analysis of genetic inheritance in a family quartet by whole genome sequencing // *Science*. 2010. V. 328. № 5978. P. 636.
<https://doi.org/10.1126/SCIENCE.1186802>
25. *Brandler W.M., Antaki D., Gujral M. et al.* Frequency and complexity of de novo structural mutation in autism // *Am. J. Hum. Genet.* 2016. V. 98. № 4. P. 667–679.
<https://doi.org/10.1016/J.AJHG.2016.02.018>
26. *Robinson P., Zemo jtel T.* Integrative genomics viewer (IGV): Visualizing alignments and variants // *Computational Exome and Genome Analysis*. 2018. P. 233–245.
<https://doi.org/10.1201/9781315154770-17>
27. *Zhou J., Troyanskaya O.G.* Predicting effects of non-coding variants with deep learning-based sequence model // *Nat. Methods*. 2015. V. 12. № 10. P. 931–934.
<https://doi.org/10.1038/nmeth.3547>
28. *Kulakovskiy I.V., Vorontsov I.E., Yevshin I.S. et al.* HOCOMOCCO: Expansion and enhancement of the collection of transcription factor binding sites models // *Nucl. Acids Res.* 2016. V. 44. № D1. P. D116–D125.
<https://doi.org/10.1093/NAR/GKV1249>
29. *Rosen N., Chalifa-Caspi V., Shmueli O. et al.* GeneLoc: Exon-based integration of human genome maps // *Bioinformatics*. 2003. V. 19. Suppl. 1.
<https://doi.org/10.1093/BIOINFORMATICS/BTG1030>
30. *Carbon S., Dietze H., Lewis S.E. et al.* Expansion of the gene ontology knowledgebase and resources // *Nucl. Acids Res.* 2017. V. 45. № D1. P. D331–D338.
<https://doi.org/10.1093/NAR/GKW1108>
31. *Rappaport N., Twik M., Plaschkes I. et al.* MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search // *Nucl. Acids Res.* 2017. V. 45. № D1. P. D877–D887.
<https://doi.org/10.1093/NAR/GKW1012>
32. *Ashburner M., Ball C.A., Blake J.A. et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium // *Nat. Genet.* 2000. V. 25. № 1. P. 25–29.
<https://doi.org/10.1038/75556>
33. *Lonsdale J., Thomas J., Salvatore M. et al.* The Genotype-Tissue Expression (GTEx) project // *Nat. Genet.* 2013. V. 45. № 6. P. 580–585.
<https://doi.org/10.1038/NG.2653>
34. *Untergasser A., Cutcutache I., Koressaar T. et al.* Primer3—new capabilities and interfaces // *Nucl. Acids Res.* 2012. V. 40. № 15. P. e115.
<https://doi.org/10.1093/nar/gks596>
35. *Lappalainen I., Thusberg J., Shen B., Vihinen M.* Genome wide analysis of pathogenic SH2 domain mutations // *Proteins*. 2008. V. 72. № 2. P. 779–792.
<https://doi.org/10.1002/PROT.21970>
36. *Glessner J.T., Reilly M.P., Kim C.E. et al.* Strong synaptic transmission impact by copy number variations in schizophrenia // *Proc. Natl Acad. Sci. USA*. 2010. V. 107. № 23. P. 10584–10589.
https://doi.org/10.1073/PNAS.1000274107/SUPPL_FILE/PNAS.201000274SI.PDF
37. *De Bruijn S.E., Verbakel S.K., De Vrieze E. et al.* Homozygous variants in *KIAA1549*, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa // *J. Med. Genet.* 2018. V. 55. № 10. P. 705–712.
<https://doi.org/10.1136/JMEDGENET-2018-105364>
38. *Greenwood T.A., Lazzeroni L.C., Murray S.S. et al.* Analysis of 94 candidate genes and 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia // *Am. J. Psychiatry*. 2011. V. 168. № 9. P. 930–946.
<https://doi.org/10.1176/APPI.AJP.2011.10050723>
39. *Lohoff F.W.* Genetic variants in the vesicular monoamine transporter 1 (*VMAT1/SLC18A1*) and neuropsychiatric disorders // *Methods Mol. Biol.* 2010. V. 637. P. 165–180.
https://doi.org/10.1007/978-1-60761-700-6_9
40. *Sato D.X., Kawata M.* Positive and balancing selection on *SLC18A1* gene associated with psychiatric disorders and human-unique personality traits // *EV. Lett.* 2018. V. 2. № 5. P. 499–510.
<https://doi.org/10.1002/EVL3.81>
41. *Schumacher J., Laje G., Jamra R.A. et al.* The DISC locus and schizophrenia: Evidence from an association study in a central European sample and from a meta-analysis across different European populations // *Hum. Mol. Genet.* 2009. V. 18. № 14. P. 2719–2727.
<https://doi.org/10.1093/HMG/DDP204>
42. *Nicodemus K.K., Callicott J.H., Higer R.G. et al.* Evidence of statistical epistasis between *DISC1*, *CIT* and *NDEL1* impacting risk for schizophrenia: Biological validation with functional neuroimaging // *Hum. Genet.* 2010. V. 127. № 4. P. 441–452.
<https://doi.org/10.1007/S00439-009-0782-Y/FIGURES/5>
43. *Cryns K., Sivakumaran T.A., Van den Ouweland J.M.W. et al.* Mutational spectrum of the *WFS1* gene in Wolfram syndrome, nonsyndromic hearing impairment, diabetes mellitus, and psychiatric disease // *Hum. Mutat.* 2003. V. 22. № 4. P. 275–287.
<https://doi.org/10.1002/HUMU.10258>
44. *Munshani S., Ibrahim E.Y., Domenicano I., Ehrlich B.E.* The impact of mutations in wolframin on psychiatric

- disorders // *Front. Pediatr.* 2021. V. 9.
<https://doi.org/10.3389/FPED.2021.718132>
45. *Zhao Q., Li T., Zhao X. et al.* Rare CNVs and Tag SNPs at 15q11.2 are associated with schizophrenia in the Han Chinese population // *Schizophr. Bull.* 2013. V. 39. № 3. P. 712.
<https://doi.org/10.1093/SCHBUL/SBR197>
46. *Kim N.S., Ringeling F.R., Zhou Y. et al.* *CYFIP1* dosages exhibit divergent behavioral impact via diametric regulation of NMDA receptor complex translation in mouse models of psychiatric disorders // *Biol. Psychiatry.* 2022. V. 92. № 10. P. 815–826.
<https://doi.org/10.1016/J.BIOPSYCH.2021.04.023>
47. *Davenport E.C., Szulc B.R., Drew J. et al.* Autism and schizophrenia-associated *CYFIP1* regulates the balance of synaptic excitation and inhibition // *Cell Rep.* 2019. V. 26. № 8. P. 2037–2051. e6.
<https://doi.org/10.1016/J.CELREP.2019.01.092>
48. *Cho H.P., Garcia-Barrantes P.M., Brogan J.T. et al.* Chemical modulation of mutant mGlu1 receptors derived from deleterious *GRM1* mutations found in schizophrenics // *ACS Chem. Biol.* 2014. V. 9. № 10. P. 2334–2346.
<https://doi.org/10.1021/CB500560H>
49. *Ayoub M.A., Angelicheva D., Vile D. et al.* Deleterious *GRM1* mutations in schizophrenia // *PLoS One.* 2012. V. 7. № 3. P. e32849.
<https://doi.org/10.1371/JOURNAL.PONE.0032849>
50. *Hirata Y., Zai C.C., Souza R.P. et al.* Association study of *GRIK1* gene polymorphisms in schizophrenia: case-control and family-based studies // *Hum. Psychopharmacol.* 2012. V. 27. № 4. P. 345–351.
<https://doi.org/10.1002/HUP.2233>
51. *Costain G., Lionel A.C., Merico D. et al.* Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays // *Hum. Mol. Genet.* 2013. V. 22. № 22. P. 4485–4501.
<https://doi.org/10.1093/HMG/DDT297>
52. *Hu W., Macdonald M.L., Elswick D.E., Sweet R.A.* The glutamate hypothesis of schizophrenia: Evidence from human brain tissue studies // *Ann. N. Y. Acad. Sci.* 2015. V. 1338. № 1. P. 38–57.
<https://doi.org/10.1111/NYAS.12547>
53. *Curtis D.* Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia // *Psychiatr. Genet.* 2018. V. 28. № 5. P. 85–89.
<https://doi.org/10.1097/YPG.0000000000000206>
54. *Landi I., Kaji D.A., Cotter L. et al.* Prognostic value of polygenic risk scores for adults with psychosis // *Nat. Med.* 2021. V. 27. № 9. P. 1576–1581.
<https://doi.org/10.1038/s41591-021-01475-7>
55. *Shimon H., Sobolev Y., Davidson M. et al.* Inositol levels are decreased in postmortem brain of schizophrenic patients // *Biol. Psychiatry.* 1998. V. 44. № 6. P. 428–432.
[https://doi.org/10.1016/S0006-3223\(98\)00071-7](https://doi.org/10.1016/S0006-3223(98)00071-7)
56. *Arranz B., Rosel P., San L. et al.* Low baseline serotonin-2A receptors predict clinical response to olanzapine in first-episode schizophrenia patients // *Psychiatry Res.* 2007. V. 153. № 2. P. 103–109.
<https://doi.org/10.1016/J.PSYCHRES.2006.12.015>

Analysis of Genetic Factors of Sporadic Schizophrenia in Family Trios Using Whole Genome Sequencing

T. V. Andreeva^{a, b, *}, Ph. A. Afanasiev^b, F. E. Gusev^{b, c}, A. D. Patrikeev^b,
 S. S. Kunizheva^{b, c}, and E. I. Rogaev^{c, d, e, **}

^aCenter for Genetics and Genetic Technologies, Faculty of Biology,
 Moscow State University, Moscow, 119234 Russia

^bVavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia

^cCenter for Genetics and Life Science, Sirius University of Science and Technology, Krasnodarski Krai, p. Sirius, 354340 Russia

^dMoscow State University, Moscow, 119234 Russia

^eDepartment of Psychiatry, UMass Chan Medical School, Worcester, MA 01545 USA

*e-mail: an_tati@vigg.ru

**e-mail: evgeny.rogaev@umassmed.edu

Schizophrenia is a mental disorder, the hereditary nature of which has been confirmed by numerous studies. Currently, more than a hundred genetic loci associated with schizophrenia have been described, and rare variants in genes and chromosomal rearrangements associated with familial cases of the disease have also been identified. However, it is not always possible to determine the hereditary nature of the pathology, many cases of schizophrenia are sporadic, and the genetic cause of such cases remains unknown. Using whole genome sequencing data for three family trios from Russia with sporadic cases of schizophrenia, we searched for rare potentially pathogenic variants in the coding and regulatory loci of the genome, including *de novo* and compound mutations. The polygenic risk of schizophrenia was also assessed using common polymorphic markers. As a result of the analysis, the genetic heterogeneity of sporadic forms of schizophrenia was shown, as well as the potential role of rare substitutions in genes associated with the metabolism of glutamate and inositol phosphate in sporadic cases of schizophrenia.

Keywords: schizophrenia, whole genome sequencing, polygenic risk score, *de novo* variants.

ПОВЫШЕННЫЙ УРОВЕНЬ ЛИЗОСФИНГОЛИПИДОВ КРОВИ У ПАЦИЕНТОВ С ШИЗОФРЕНИЕЙ

© 2023 г. А. Э. Копытова^{1, 2, *}, Т. С. Усенко^{1, 2}, А. И. Безрукова², К. С. Башарова²,
Т. В. Андреева^{3, 4, 5}, Е. В. Волкова¹, А. Д. Манахов^{3, 4, 5}, Г. В. Байдакова⁶, Е. И. Пальчикова⁷,
Е. Ю. Захарова⁶, Н. М. Залуцкая⁷, Н. Г. Незнанов^{1, 7}, С. Н. Пчелина^{1, 2}

¹Первый Санкт-Петербургский государственный медицинский университет
им. И.П. Павлова, Санкт-Петербург, 197022 Россия

²Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра
“Курчатовский институт”, Санкт-Петербург, Гатчина, 188300 Россия

³Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

⁴Центр генетики и генетических технологий, Московский государственный университет
им. М.В. Ломоносова, Москва, 119234 Россия

⁵Центр генетики и наук о жизни, Научно-технологический университет “Сириус”,
Краснодарский край, пгт. Сириус, 354340 Россия

⁶Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, 115478 Россия

⁷Национальный медицинский исследовательский центр психиатрии и неврологии
им. В.М. Бехтерева, Санкт-Петербург, 192019 Россия

*e-mail: kopytovaalena@mail.ru

Поступила в редакцию 29.12.2022 г.

После доработки 23.01.2023 г.

Принята к публикации 24.01.2023 г.

Шизофрения представляет собой психическое расстройство и характеризуется нарушением передачи дофамина в нейронах. Последние данные показали, что лизосомные болезни накопления (ЛБН), сопровождающиеся снижением активности ферментов и соответствующим накоплением субстрата в лизосомах из-за мутаций в лизосомных генах, могут проявляться широким спектром клинических проявлений, включая психозы, аффективные расстройства, деменцию с ранним началом и шизофренией. Цель настоящего исследования заключалась в оценке уровня лизосфинголипидов у пациентов с шизофренией, болезнью Паркинсона (БП) и контроле. В исследование включены 52 пациента с шизофренией, 170 пациентов с БП и 166 неврологически здоровых лиц (контрольная группа). Концентрацию лизосомных субстратов (гексозилсфингозин (HexSph), глоботриаозилсфингозин (LysoGb3), лизосфингомиелин (LysoSM)) определяли высокоэффективной жидкостной хроматографией в сочетании с тандемной масс-спектрометрией (ВЭЖХ-МС/МС) в крови. Выявлено повышение концентрации LysoSM, LysoGb3, HexSph у пациентов с шизофренией по сравнению с контролем ($p < 0.0001$, $p < 0.0001$, $p < 0.0001$ соответственно). Наши результаты подтверждают нарушение лизосфинголипидного состава крови у пациентов с шизофренией.

Ключевые слова: шизофрения, лизосфинголипиды, ВЭЖХ-МС/МС, кровь.

DOI: 10.31857/S0016675823060085, **EDN:** SSPQEG

Шизофрения, а точнее расстройства шизофренического спектра — одно из наиболее тяжелых по своим проявлениям и последствиям психических расстройств, характеризующееся дезинтеграцией психической деятельности, сочетанием позитивной (галлюцинации, бред, аффект и др.) и негативной (апатия, эмоциональная и социальная отгороженность и др.) симптоматики, поведенческих и когнитивных нарушений [1, 2]. В зависимости от принятых критериев и принципов регистрации распространенность шизофрении определяют в

пределах от 0.5 до 1.6% во всем мире (20–45 млн больных). На сегодняшний день расстройства шизофренического спектра являются болезнью с неизвестной этиологией, что стимулирует исследователей к созданию единой теории. В генетике психических заболеваний прочно утвердилось представление о полигенной и многофакторной природе шизофрении со сложным взаимодействием полиморфных вариантов генов и условий среды, включающих как пре-, так и постнатальные особенности развития организма.

Таблица 1. Демографические характеристики сравниваемых групп

Группа	Возраст \pm SD, лет	Возраст начала \pm SD, лет	Пол мужчины : женщины
Шизофреники ($N = 52$)	61.0 \pm 11.1	51.1 \pm 11.5	20 : 32
Пациенты с БП ($N = 170$)	63.5 \pm 9.2	55.6 \pm 8.1	73 : 97
Контроль ($N = 166$)	61.2 \pm 6.9	—	70 : 96

Ряд исследований показывает ассоциацию шизофрении с метаболическим синдромом. Так, метаболический синдром встречается среди пациентов с шизофренией в 2 раза чаще, чем в популяции [3]. Сфинголипиды являются основным компонентом клеточных мембран и преобладают в нервных клетках. Последние данные связывают уровень сфинголипидов с воспалением. Обсуждается, что воспаление, индуцированное повышенным уровнем церамида, может играть роль в патогенезе шизофрении [4]. Важно также отметить, что некоторые из врожденных заболеваний обмена веществ могут проявляться в виде острых приступов психоза или характеризоваться хроническим психозом. На основании обзора литературы известно не менее 40 таких заболеваний [5]. Несколько генов, ассоциированных с врожденными заболеваниями обмена веществ, были недавно связаны с шизофренией на основании сравнения данных экзозом между 2545 пациентами с шизофренией и 2545 контроля [5]. Среди 13 генов, ассоциированных с шизофренией, Y. Trakadis и соавт. выявили пять, мутации в которых приводят к развитию лизосомных болезней накопления (ЛБН), связаны с лизосомной дисфункцией и накоплением лизосфинголипидов [6].

Настоящее исследование посвящено оценке лизосфинголипидов крови, накапливающихся при наиболее распространенных сфинголипидозах, а именно, болезни Гоше, болезни Фабри и болезни Нимана-Пика типа А/В. Проведенная, соответственно, оценка концентрации лизосфинголипидов (гексозилсфингозина (HexSph) (глюкозилсфингозин (GlcSph) + галактозилсфингозин (GalSph)), глоботриаозилсфингозина (LysoGb3), лизосфингомиелина (LysoSM)) в крови пациентов с шизофренией методом высокоэффективной жидкостной хроматографии в сочетании с тандемной масс-спектрометрией (ВЭЖХ-МС/МС) показала ассоциацию повышенного уровня лизосфинголипидов с развитием заболевания.

МАТЕРИАЛЫ И МЕТОДЫ

Характеристика групп, включенных в исследование

Для данного исследования была сформирована группа из 52 больных шизофренией с поздним началом. Также в настоящее исследование в качестве групп сравнения были включены группа из 170 пациентов с болезнью Паркинсона (БП) и 166

человек без неврологических и психических расстройств (контроль). У всех пациентов, включенных в исследование, были исключены мажорные мутации в гене *GBA*. Скрининг мажорных мутаций в гене *GBA* (N370S, L444P) проводился согласно методике, описанной ранее [7]. Демографические характеристики исследуемых групп представлены в табл. 1.

Исследование одобрено локальными этическими комитетами указанных медицинских учреждений и проводилось при подписании информированного согласия пациента. Включенные в исследование группы пациентов с шизофренией и БП не отличались от контрольной группы по полу и возрасту ($p > 0.05$).

Определение концентрации лизосфинголипидов в крови

У каждого из участников исследования были взяты образцы венозной крови в пробирки с ЭДТА. Карточки сухих пятен крови готовили путем пиетирования 40 мкл цельной крови в каждое пятно. Пятна крови высушивали при комнатной температуре в течение двух часов, а затем хранили при 4°C до экстракции. Концентрацию субстратов (гексозилсфингозин (HexSph) (глюкозилсфингозин (GlcSph) + галактозилсфингозин (GalSph)), глоботриаозилсфингозин (LysoGb3), лизосфингомиелин (LysoSM)) оценивали методом ВЭЖХ-МС/МС в сухих пятнах крови, как описано ранее [8].

Статистическая обработка результатов

Полученные данные проверяли на соответствие нормальному распределению методом Шапиро–Уилка, сравнение вариационных рядов между исследуемыми группами – методом Крускала–Уоллиса. Сравнение вариационных рядов между двумя группами пациентов проводили с использованием критерия Манна–Уитни. Все расчеты были проведены с использованием встроенных пакетов R (версия 4.2.0). Использовалась поправка на множественную проверку гипотез – поправка Бонферрони. Значимыми считали различия при $p < 0.05$. Клинические характеристики представлены в виде среднего значения \pm стандартное отклонение (SD), экспериментальные данные – медиана (мин–макс).

Таблица 2. Концентрация лизосфинголипидов в крови во всех исследуемых группах

Группа	Концентрация (нг/мл) лизосфинголипидов, медиана (мин–макс)		
	HexSph	LysoGb3	LysoSM
Пациенты с шизофренией ($N = 52$)	6.04 (1.68–15.40) $p = 4.8e-09^*$ $p = 1.7e-10^{**}$	1.21 (0.38–2.42) $p = 4.2e-10^*$ $p = 4.2e-08^{**}$	5.55 (2.35–24.29) $p = 3.9e-07^*$ $p = 3.3e-12^{**}$
Пациенты с БП ($N = 170$)	2.49 (0.86–3.23)	0.78 (0.04–3.73)	3.47 (0.72–7.91) $p = 0.00063^*$
Контроль ($N = 166$)	2.97 (0.57–15.36)	0.78 (0.03–2.31)	3.97 (0.59–11.60)

Примечание. * – по сравнению с контролем, ** – по сравнению с пациентами с БП.

РЕЗУЛЬТАТЫ

Концентрацию лизосфинголипидов оценивали в сухих пятнах крови в группе пациентов с шизофренией, пациентов с БП и контроля (табл. 2). Выявлено повышение концентрации всех исследуемых лизосфинголипидов: LysoSM, LysoGb3, HexSph у пациентов с шизофренией по сравнению как с контролем, так и с пациентами с БП ($p < 0.0001$) (табл. 2, рис. 1).

ОБСУЖДЕНИЕ

В настоящее время широко обсуждается роль нарушения метаболизма сфинголипидов в нейродегенерации [9]. Хорошо известно, что липиды (в том числе керамиды) содержатся в нервных тканях и важны для функционирования клеток головного мозга [10].

Предыдущие данные показали, что ЛБН, такие как болезнь Санфилиппо подтипов А, В, С, D, мукополисахаридоз (МПС типа 3 А–D), болезнь Хантера (МПС типа 2), альфа-маннозидоз и β -маннозидоз и болезнь Ниманна–Пика типа С1, С2, болезнь Фабри, метахроматическая лейкодистрофия, ганглиозидоз GM2 (болезнь Тея-Сакса и Сандхоффа) характеризуются рядом различных психических проявлений [11]. Следует отметить, что все обсуждаемые ЛБН являются аутомно-рецессивными заболеваниями. Мы можем предположить, что в сочетании с другими неизвестными факторами эти варианты могут нарушать работу лизосом и таким образом влиять на функцию нейронов. Это предположение может быть подтверждено примером роли дисфункции лизосомного фермента глюкоцереброзидазы (GCase) в патогенезе БП. Мутации в гене *GBA*, кодирующем этот фермент, приводят к наиболее частой форме ЛБН – болезни Гоше и в то же время мутации в гене *GBA* увеличивают риск БП в 7–8 раз во всех популяциях [7, 12]. Носительство мутаций в гене *GBA* приводит к снижению активности GCase и накоплению субстрата HexSph при БП [13]. В настоящем исследовании нами показано повышение концентрации HexSph в крови

только у пациентов с шизофренией, но не у пациентов с БП. Следует отметить, что в настоящее исследование была включена группа пациентов с БП, с отсутствием мажорных мутаций (N370S, L444P) в гене *GBA*.

Следует отметить, что ранее нарушения активности лизосомных ферментов, участвующих в метаболизме керамидов, были продемонстрированы при спорадической форме БП, а также у пациентов с другими синуклеинопатиями (деменция с тельцами Леви, мультисистемная атрофия) [14, 15]. Интересно отметить, что синуклеинопатии в некоторых случаях могут рассматриваться как сопутствующее заболевание у пациентов с шизофренией [2, 16–18]. Исследования аутоптов мозга пациентов с шизофренией продемонстрировали увеличение содержания керамидов в виде гидроксидцерамидов, фитощерамидов, гексозилцерамидов, лактозилцерамидов и керамидфосфоэтаноламинов и других в белом веществе и лобной коре мозга [19, 20]. Кроме того, недавние исследования показали, что в плазме крови были повышены концентрации Cer (d18 : 1/16 : 0), Cer (d18 : 1/18 : 0) и Cer (d18 : 1/24 : 1) у пациентов с шизофренией по сравнению с контрольной группой [9]. Ранее сообщалось, что нарушения метаболизма керамидов могут быть связаны с когнитивными нарушениями у пациентов с шизофренией [21]. Керамиды участвуют в различных физиологических и патологических клеточных процессах [22] и могут быть частью патогенеза шизофрении, в частности из-за нарушения миелинизации и дисфункции олигодендроцитов и нейровоспаления [23–25].

Болезнь Нимана–Пика является редкой аутомно-рецессивной ЛБН, характеризующейся дефицитом сфингомиелиназы (ASMase), приводящим к накоплению сфингомиелина (LysoSM), может характеризоваться психотическим поведением, соматическими галлюцинациями, приводящими к диагнозу шизофрениеподобного психоза [26]. С. Zhuo с соавт. предположили, что нормализация aberrантной системы ASMase/церамид или снижение уровня керамида путем использования

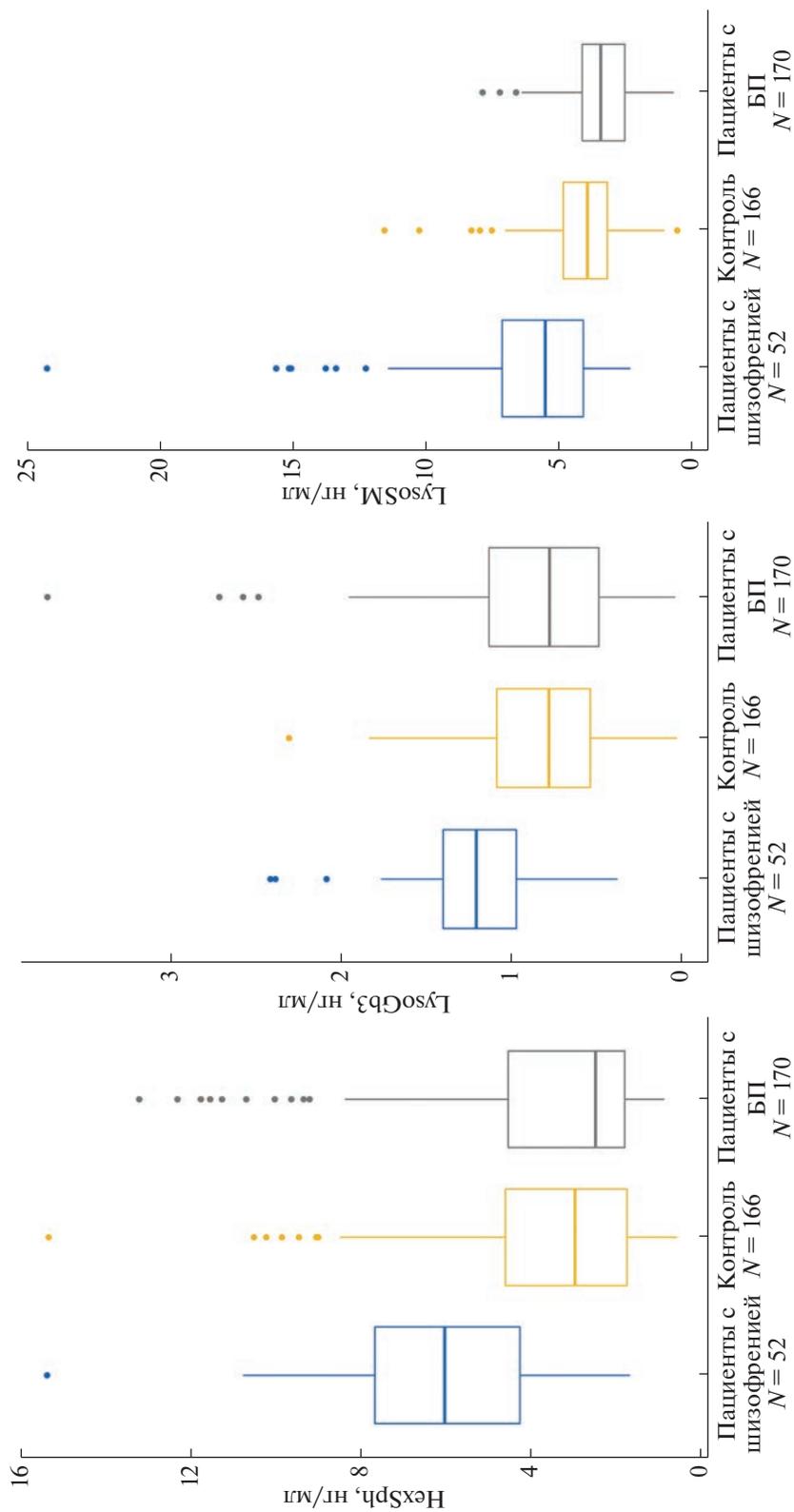


Рис. 1. Концентрация лизосфинголипидов в исследуемых группах в крови.

одобренных функциональных ингибиторов ASMase, таких как флувоксамин и розувастатин, могут улучшить клиническое течение у пациентов с шизофренией [27]. Дефицит другого лизосомного фермента альфа-галактозидазы (GLA) из-за мутации в гене *GLA* вызывает болезнь Фабри и приводит к накоплению нейтральных гликофинголипидов (LysoGb3) в лизосомах в различных тканях, включая нервную систему [28]. В предыдущем исследовании нами впервые показано снижение активности ASMase в крови пациентов с шизофренией [29]. В настоящем исследовании мы наблюдали повышение концентрации субстрата данного фермента LysoSM в крови пациентов с шизофренией, но не при БП.

Полученные в настоящем исследовании данные подтверждают селективное накопление лизосфинголипидов в крови пациентов с шизофренией, предполагая вовлечение в патогенез шизофрении как сигнального пути ASMase-церамид, так и GLA-церамид, и находятся в соответствии с гипотезой церамид-опосредованного воспаления в патогенезе шизофрении [4].

Авторы выражают благодарность за руководство и критику рукописи Евгению Ивановичу Роговцу.

Работа выполнена при финансовой поддержке Российского научного фонда в рамках научного проекта № 19-75-30039.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 года и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников было получено информированное добровольное согласие.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Howrigan D.P., Rose S.A., Samocha K.E. et al. Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations // *Nat. Neurosci.* 2020. V. 23. P. 185–193. <https://doi.org/10.1038/s41593-019-0564-3>
2. Oh J., Shen G., Nan G. et al. Comorbid schizophrenia and Parkinson's disease: A case series and brief review // *Neurol. Asia.* 2017. V. 22. P. 139–142.
3. Papanastasiou E. The prevalence and mechanisms of metabolic syndrome in schizophrenia: a review // *Theor. Adv. Psychopharmacol.* 2013. V. 3. № 1. P. 33–51. <https://doi.org/10.1177/2045125312464385>
4. Castillo R.I., Rojo L.E., Henriquez-Henriquez M. et al. From molecules to the clinic: Linking schizophrenia and metabolic syndrome through sphingolipids metabolism // *Front. Neurosci.* 2016. V. 10. P. 488–503. <https://doi.org/10.3389/fnins.2016.00488>
5. Pardiñas A.F., Holmans P., Pocklington A.J. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection // *Nat. Genet.* 2018. V. 50. P. 381–389.
6. Trakadis Y.J., Fulginiti V., Walterfang M. Inborn errors of metabolism associated with psychosis: Literature review and case-control study using exome data from 5090 adult individuals // *J. Inher. Metab. Dis.* 2018. V. 41. № 4. P. 613–621. <https://doi.org/10.1007/s10545-017-0023-9>
7. Emelyanov A.K., Usenko T.S., Tesson C. et al. Mutation analysis of Parkinson's disease genes in a Russian data set // *Neurobiol. Aging.* 2018. V. 71. P. 267.e7–267.e10. <https://doi.org/10.1016/j.neurobiolaging.2018.06.027>
8. Polo G., Burlina A.P., Ranieri E. et al. Plasma and dried blood spot lysosphingolipids for the diagnosis of different sphingolipidoses: a comparative study // *Clin. Chem. Lab. Med.* 2019. V. 57. № 12. P. 1863–1874.
9. Czubowicz K., Jęško H., Wencel P. et al. The role of ceramide and sphingosine-1-phosphate in Alzheimer's disease and other neurodegenerative disorders // *Mol. Neurobiol.* 2019. V. 56. P. 5436–5455.
10. Tkachev A.I., Stekolshchikova E.A., Morozova A.Y. et al. Ceramides: Shared lipid biomarkers of cardiovascular disease and schizophrenia // *Consort. Psychiatr.* 2021. V. 2. P. 35–43. <https://doi.org/10.17816/CP101>
11. Cox T.M. Lysosomal diseases and neuropsychiatry: Opportunities to rebalance the mind // *Front. Mol. Biosci.* 2020. V. 7. P. 177–185. <https://doi.org/10.3389/fmolb.2020.00177>
12. Sidransky E., Nalls M.A., Aasly J.O. et al. Multi-center analysis of glucocerebrosidase mutations in Parkinson disease // *N. Engl. J. Med.* 2009. V. 361. P. 1651–1661. <https://doi.org/10.1056/NEJM0A0901281>
13. Kopytova A.E., Usenko T.S., Baydakova G.V. et al. Could blood hexosylsphingosine be a marker for Parkinson's disease linked with *GBA1* mutations? // *Mov. Disord.* 2022. V. 37. № 8. P. 1779–1781. <https://doi.org/10.1002/mds.29132>
14. Moors T.E., Paciotti S., Ingrassia A. et al. Characterization of brain lysosomal activities in GBA-related and sporadic Parkinson's disease and dementia with Lewy bodies // *Mol. Neurobiol.* 2019. V. 56. P. 1344–1355.
15. Usenko T.S., Senkevich K.A., Bezrukova A.I. et al. Impaired sphingolipid hydrolase activities in dementia with Lewy bodies and multiple system atrophy // *Mol. Neurobiol.* 2022. V. 59. P. 2277–2787.
16. Kuusimäki T., Al-Abdulrasul H., Kurki S. et al. Increased risk of Parkinson's disease in patients with schizophrenia spectrum disorders // *Mov. Disord.* 2021. V. 36. P. 1353–1361.
17. De Vries P.J., Honer W.G., Kemp P.M., McKenna P.J. Dementia as a complication of schizophrenia // *J. Neurol. Neurosurg. Psychiatr.* 2001. V. 70. P. 588–596.
18. Komatsu H., Kato M., Kinpara T. et al. Possible multiple system atrophy with predominant parkinsonism in a patient with chronic schizophrenia: A case report // *BMC Psychiatry.* 2018. V. 18. P. 1–9.

19. Schwarz E., Prabakaran S., Whitfield P. et al. High throughput lipidomic profiling of schizophrenia and bipolar disorder brain tissue reveals alterations of free fatty acids, phosphatidylcholines, and ceramides // *J. Proteome Res.* 2008. V. 7. P. 4266–4277.
20. Wood P.L. Targeted lipidomics and metabolomics evaluations of cortical neuronal stress in schizophrenia // *Schizophr. Res.* 2019. V. 212. P. 107–112.
21. Tessier C., Sweers K., Frajerman A. et al. Membrane lipidomics in schizophrenia patients: A correlational study with clinical and cognitive manifestations // *Transl. Psychiatry.* 2016. V. 6. P. e906–e914.
22. Paciotti S., Albi E., Parnetti L., Beccari T. Lysosomal ceramide metabolism disorders: Implications in Parkinson's disease // *J. Clin. Med.* 2020. V. 9. P. 594–614.
23. Takahashi N., Sakurai T., Davis K.L., Buxbaum J.D. Linking oligodendrocyte and myelin dysfunction to neurocircuitry abnormalities in schizophrenia // *Prog. Neurobiol.* 2011. V. 93. P. 13–24.
24. Vallée A. Neuroinflammation in schizophrenia: The key role of the WNT/ β -catenin pathway // *Int. J. Mol. Sci.* 2022. V. 23. P. 2810–2825.
25. Gouvêa-Junqueira D., Falvella A.C.B., Antunes A.S.L.M. et al. Novel treatment strategies targeting myelin and oligodendrocyte dysfunction in schizophrenia // *Front. Psychiatry.* 2020. V. 11. P. 379–395.
26. Mihaylova V., Hantke J., Sinigerska I. et al. Highly variable neural involvement in sphingomyelinase-deficient Niemann–Pick disease caused by an ancestral Gypsy mutation // *Brain.* 2007. V. 130. P. 1050–1061. <https://doi.org/10.1093/brain/awm026>
27. Zhuo C., Zhao F., Tian H. et al. Acid sphingomyelinase/ceramide system in schizophrenia: Implications for therapeutic intervention as a potential novel target // *Transl. Psychiatry.* 2022. V. 12. P. 260–266. <https://doi.org/10.1038/s41398-022-01999-7>
28. Akhtar M.M., Elliott P.M. Anderson–Fabry disease in heart failure // *Biophys. Rev.* 2018. V. 10. P. 1107–1119.
29. Usenko T.S., Bezrukova A.I., Basharova K.S. et al. Link between schizophrenia and lysosomal storage disorders based on NGS and LC-MS/MS analyses // *Am. J. Geriatr. Psychiatry.* 2023. Submitted to journal.

Elevated Level of Blood Lysosphingolipids in Patients with Schizophrenia

A. E. Kopytova^{a, b, *}, T. S. Usenko^{a, b}, A. I. Bezrukova^b, K. S. Basharova^b, T. V. Andreeva^{c, d, e},
E. V. Volkova^a, A. D. Manakhov^{c, d, e}, G. V. Baydakova^f, E. I. Palchikova^g, E. Yu. Zakharova^f,
N. M. Zalutskaya^g, N. G. Neznanov^{a, g}, and S. N. Pchelina^{a, b}

^a*Pavlov First Saint-Petersburg State Medical University, Saint-Petersburg, 197022 Russia*

^b*Konstantinov Petersburg Nuclear Physics Institute of National Research Centre “Kurchatov Institute”, Saint-Petersburg, Gatchina, 188300 Russia*

^c*Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, 119991 Russia*

^d*Center of Genetics and Genetic Technologies, Lomonosov Moscow State University, Moscow, 119234 Russia*

^e*Sirius University of Science and Technology, Krasnodarski Krai, Sirius, 354340 Russia*

^f*Research Center for Medical Genetics, Moscow, 115478 Russia*

^g*Bekhterev National Medical Research Center Psychiatry and Neurology, Saint-Petersburg, 192019 Russia*

*e-mail: kopytovaalena@mail.ru

Schizophrenia is a mental disorder with a prevalence of 0.7–1% of the general population and is characterized by impaired dopamine transmission in neurons. Recent data have shown that lysosomal storage disorders (LSD), characterized by a decrease in enzyme activity and a corresponding accumulation of substrate in lysosomes due to mutations in lysosomal genes, can manifest a wide range of clinical symptoms including psychosis, affective disorders, early onset dementia and schizophrenia. The purpose of this study was to assess the level of lysosphingolipids in patients with schizophrenia, Parkinson's disease (PD) and control. The study included 52 patients with schizophrenia, 170 patients with PD patients and 166 neurologically healthy individuals (control group). Concentration of lysosomal substrates (hexosylsphingosine (HexSph), globotriaosylsphingosine (LysoGb3), lysosphingomyelin (LysoSM)) were measured by high-liquid chromatography with tandem mass spectrometry (HPLC-MS/MS) in blood. An increase in the concentration of LysoSM, LysoGb3, HexSph was detected in patients with schizophrenia compared to control ($p < 0.0001$, $p < 0.0001$, $p < 0.0001$, respectively). Our results confirm a violation of the lysosphingolipid composition of the blood in patients with schizophrenia.

Keywords: schizophrenia, lysosphingolipids, HPLC-MS/MS, blood.

ВЛИЯНИЕ 3Д-ОРГАНИЗАЦИИ ХРОМАТИНА НА СУПЕРЭНХАНСЕР–ПРОМОТОРНЫЕ ВЗАИМОДЕЙСТВИЯ В ЭМБРИОНАЛЬНЫХ СТВОЛОВЫХ И ЗЛОКАЧЕСТВЕННЫХ КЛЕТКАХ

© 2023 г. Ю. А. Эйдельман^{1, 2, *}, С. Г. Андреев^{1, 2, **}

¹Институт биохимической физики им. Н.М. Эмануэля Российской академии наук, Москва, 119334 Россия

²Национальный исследовательский ядерный университет “МИФИ”, Москва, 115409 Россия

*e-mail: eidel@mail.ru

**e-mail: andreev_sg@mail.ru

Поступила в редакцию 11.10.2022 г.

После доработки 05.12.2022 г.

Принята к публикации 08.12.2022 г.

Взаимодействие энхансеров и суперэнхансеров (СЭ) с промотором функционально значимо для регуляции экспрессии генов. Паттерны этих взаимодействий играют ключевую роль в регуляции таких процессов, как дифференцировка, злокачественная трансформация и др. Чтобы получить количественную информацию о взаимосвязи структуры хроматина и экспрессии генов, в настоящей работе был проведен компьютерный анализ конформаций хроматина вблизи гена плюрипотентности *Nanog* мыши, включая контакты СЭ с промотором, в нормальных эмбриональных стволовых (mESC) и в клетках лимфомы (CH12LX). С помощью биофизических методов моделирования идентифицированы параметры суперэнхансер-промоторных взаимодействий – распределения расстояний между промотором *Nanog* и СЭ, частоты контактов с одним и несколькими СЭ одновременно. В нормальных mESC частота контактов промоторов с СЭ выше, чем в злокачественных, и чаще образуются сложные контакты, с двумя и более СЭ. В злокачественных клетках обнаружена малая субпопуляция клеток, где промотор контактирует одновременно с тремя СЭ. Субпопуляции злокачественных клеток с множественными контактами промотор–СЭ могут предрасполагать к проявлению свойств стволовых клеток и гипотетически быть кандидатом на опухолевые стволовые клетки.

Ключевые слова: *Nanog*, промотор-энхансерные взаимодействия, Hi-C, полимерное моделирование, молекулярная динамика, структура хроматина.

DOI: 10.31857/S0016675823060036, **EDN:** SRUZYX

Суперэнхансерами (СЭ) принято считать кластеры энхансеров, характеризующиеся повышенным уровнем связывания транскрипционных факторов, они играют важную роль в регуляции экспрессии генов [1]. Связывание промоторов не с одним, а с несколькими энхансерами или СЭ приводит к повышению экспрессии соответствующего гена. В [1] показано, что в эмбриональных стволовых клетках повышена частота взаимодействий промотора с СЭ генов плюрипотентности *Oct4*, *Sox2* и *Nanog*. В 2000 г. были созданы эффективные методы количественного исследования геномных контактов: “захват хромосомных конформаций” (Chromosome Conformation Capture, 3C) [2] и его производные, включая полногеномный метод Hi-C [3]. Методы основаны на лигировании ДНК вблизи контактов хромосомных локусов и последующем определении частот и расположения точек лигирования на геноме. Эти подходы стимулировали развитие структурных исследований,

взаимосвязи экспрессии генов и 3Д-контактов промоторов с энхансерами и СЭ. Однако стандартные методы захвата конформаций не дают прямой информации о 3Д-структуре генома [4, 5] и, в частности, о пространственной организации областей, где расположены промоторы и суперэнхансеры. Кроме того, они предназначены для измерения парных контактов (например, данного промотора с данным энхансером) в популяции клеток, но не могут детектировать более сложные непарные контакты. Недавно появившийся метод “картирование архитектуры генома” (Genome Architecture Mapping, GAM) позволяет детектировать и сложные контакты, но такие работы пока единичны [6]. Для восстановления информации о 3Д-структуре участков генома и контактах локусов по данным Hi-C применяются методы биоинформатики и полимерного моделирования, например [4, 5, 7, 8].

Таблица 1. Положение гена *Nanog* и суперэнхансеров (СЭ) в мышинной хромосоме 6 (геном mm10). Данные по СЭ в mESC из [13]

	Начало, млн пн	Конец, млн пн	Размер СЭ, тыс. пн
<i>Nanog</i>	122.708	122.715	
СЭ1	122.337	122.341	4
СЭ2	122.664	122.666	2
СЭ3	122.684	122.709	25
СЭ4	122.761	122.772	11

Одна из центральных проблем генетики и эпигенетики развития — как гены плюрипотентности вовлечены в регуляцию свойств нормальных клеток и могут ли они участвовать в регуляции свойств клеток со злокачественным фенотипом — остается слабоизученной. В настоящей работе мы провели сравнительный анализ структурной организации хроматина в окрестности гена плюрипотентности *Nanog*, включающей в себя промотор-суперэнхансерные контакты в эмбриональных стволовых (mESC) и в лимфоидных злокачественных (CH12LX) клетках мыши. С помощью разработанного метода полимерного моделирования были получены детальные количественные характеристики промотор-СЭ контактов, распределения расстояний между промотором и СЭ, частоты контактов с одним и несколькими СЭ одновременно. Мы демонстрируем, что в mESC средняя частота контактов промоторов с СЭ выше, чем в злокачественных, и в них намного чаще образуются сложные контакты, с двумя и более СЭ. В злокачественных CH12LX клетках расчеты обнаруживают наличие малой субпопуляции клеток, в которых промотор контактирует одновременно с тремя СЭ. Субпопуляция злокачественных клеток со структурой хроматина, характеризующаяся множественными промотор-СЭ контактами, не детектируемыми классическим методом Hi-C, может означать наличие субпопуляции опухолевых стволовых клеток (ОСК).

МАТЕРИАЛЫ И МЕТОДЫ

Метод полимерного моделирования структуры хромосом и описания карт Hi-C контактов изложен в предыдущей статье [5]. В данной работе фрагмент мышинной хромосомы 6 с геномными координатами 120–123.5 млн пн моделировался полимерной цепью размером 3.5 млн пн с субъединицами, содержащими 10 тыс. пн ДНК, диаметр субъединицы был выбран 88 нм.

Полимерное моделирование проводилось следующим образом. Конформация полимерной цепи формируется за счет потенциалов объемных взаимодействий и исключенного объема. Потен-

циалы объемных взаимодействий гетерогенны, т.е. зависят от конкретной пары локусов (i, j). Их величины определяются с помощью итеративного алгоритма по критерию наилучшего согласия расчетной и экспериментальной карт контактов. На каждой итерации для данного набора потенциалов моделируется ансамбль конформаций, определяется расчетная карта контактов, потенциал для каждой пары (i, j) меняется в зависимости от того, расчетная частота контактов больше или меньше экспериментальной.

Расчетная карта контактов вычисляется для каждой конформации: пара локусов (i, j) находится в контакте, если расстояние между ними меньше заданного, $r_{ij} < R_{\text{cont}}$; затем суммированием по всем конформациям получается средняя по ансамблю карта. Из готовых конформаций цепи 3.5 млн пн вычлняются субцепи с ~ 1 млн пн окрестностью гена *Nanog*, и дальнейшие действия производятся уже на ансамбле этих субцепей.

Одной из характеристик ансамбля конформаций является функция $\psi_{i,j}(r)$ — распределение расстояний между интересующими нас парами локусов (i — промотор, j — один из СЭ). В случае самого крупного из суперэнхансеров, СЭ3, занимающего три субъединицы (табл. 1), распределение считалось между промотором и геометрическим центром этих трех субъединиц.

Данные по экспрессии гена *Nanog* в клетках mESC и CH12LX взяты из ENCODE [9] по адресам ENCSR000CGU и ENCSR000AJV соответственно. Обработка данных по экспрессии осуществлялась с помощью HTSeq [10], и определялся уровень экспрессии в единицах FPKM (Fragments Per Kilobase Million). FPKM в клетках mESC равен 0.6, в клетках CH12LX равен нулю.

Hi-C карты рассматриваемого сегмента мышинной хромосомы 6 в разрешении 10 и 40 тыс. пн строились по базам данных из Gene Expression Omnibus: GSE35156 [11] для mESC, GSE63525 [12] для CH12LX клеток.

Исследование ансамбля конформаций на уровне 1 млн пн окрестности *Nanog* осуществлялось путем иерархического кластерного анализа [7]. Для каждой пары конформаций (p, q) вычислялось “среднеквадратичное расстояние” (root mean square distance, dRMSD) по формуле

$$\text{dRMSD}(p, q) = \sqrt{\frac{\sum_{i,j} (d_{ij}^p - d_{ij}^q)^2}{N(N-1)/2}},$$

где d_{ij}^p и d_{ij}^q — расстояния между i -м и j -м элементами в конформациях p и q соответственно [7], N — число элементов цепи. Этот показатель является количественным критерием степени сходства данных двух конформаций из ансамбля. Кластеризация осуществлялась методом невзвешенного

группового среднего. Для различного числа кластеров определялось среднее расстояние от всех конформаций до центра кластеров, к которым они принадлежат, и оптимальное число кластеров определялось как такое число, после которого среднее расстояние переставало существенно уменьшаться [7]. Для клеток mESC оно оказалось равным 7, для клеток CH12LX — 6. В дальнейшем мы рассматривали только кластеры, содержащие более 1% конформаций; таких кластеров по три для каждой линии клеток.

РЕЗУЛЬТАТЫ

Для выяснения роли пространственной структуры хромосомного локуса в окрестности гена *Nanog* в суперэнхансер-промоторных взаимодействиях нами проведено количественное описание Hi-C частот контактов и получены ансамбли конформаций 3.5-млн пн участка мышины хромосомы 6, включающего *Nanog*, для клеток mESC и CH12LX, по 4000 конформаций для каждой линии (рис. 1). Ансамбль представляет собой набор 3D-конформаций хроматина, которые дают информацию о флуктуациях положения локусов в единичных клетках (хромосомах), о распределении расстояний между локусами, флуктуациях числа контактов и др. При этом средние по ансамблю карты контактов описываются полимерной моделью с высокими коэффициентами корреляции. На рис. 1 представлено сравнение экспериментальных карт контактов [11, 12] с расчетными, полученными в настоящей работе. Приведены примеры 3D-структур хроматина в окрестности гена *Nanog* для нормальных и злокачественных клеток, восстановленных на основе полимерной модели 3.5-млн пн участка хромосомы 6 клеток мыши с разрешением 10 тыс. пн. Укрупнение расчетных и экспериментальных карт контактов на рис. 1, а, д выбрано 40 тыс. пн, как в экспериментальной работе [11]. Чтобы получить представление о конформациях хроматина с более высоким разрешением, построены экспериментальные и расчетные карты с разрешением 10 тыс. пн, соответствующим размеру субъединицы полимерной цепи, в непосредственной (~1 млн пн) окрестности *Nanog* (рис. 1, в, ж). Показаны несколько конформаций участка хромосомы, на котором расположен ген *Nanog* и четыре ближайших к нему СЭ. Данные по положению СЭ в геноме взяты из [13]. Координаты гена и суперэнхансеров на хромосоме 6 приведены в табл. 1. Размеры СЭ варьируются от двух до 25 тыс. пн, т.е. они располагаются на одной—трех субъединицах полимерной цепи. Помимо четырех СЭ, на рассматриваемом участке хромосомы имеются также четыре отдельных энхансера [14], но для них расчеты контактов и расстояний не проводились.

Таблица 2. Средние расстояния между промотором гена *Nanog* и четырьмя суперэнхансерами в клетках mESC и CH12LX

Суперэнхансер	Среднее расстояние, нм		<i>p</i>
	mESC	CH12LX	
СЭ1	338	363	3e-18*
СЭ2	160	170	1e-18*
СЭ3	97	101	1e-29*
СЭ4	176	196	2e-51*

Примечание. В правой колонке — оценка значимости различий по *t*-тесту; звездочками помечены различия, значимые с $p < 0.01$.

Чтобы определить, как близость СЭ и промотора *Nanog* зависит от типа клеток, на основе полученных ансамблей конформаций для мышинных эмбриональных стволовых mESC и злокачественных CH12LX клеток были построены распределения расстояний между промотором и всеми четырьмя СЭ (рис. 2). Средние расстояния и статистическая значимость различий между двумя типами клеток приведены в табл. 2.

По данным на рис. 2 и в табл. 2 средние расстояния между промотором и всеми четырьмя СЭ различаются в двух типах клеток в пределах 10% (различия статистически значимые, по *t*-тесту $p < 1e-19$ для всех четырех СЭ). При этом на близких расстояниях, меньше 100 нм, промотор и три из четырех СЭ колокализуются чаще в mESC, чем в CH12LX (врезки на рис. 2).

Знание ансамбля 3D-конформаций, соответствующего популяционной Hi-C карте контактов, позволило получить частоты парных контактов промотора с конкретным СЭ, а также частоты образования контактов промотора с двумя, тремя или четырьмя СЭ одновременно. В табл. 3 приведены частоты конформаций в клетках mESC и CH12LX со всеми комбинациями контактов промотора с каждым конкретным СЭ или набором СЭ. Такая информация представляет собой новые данные, не извлекаемые из Hi-C экспериментов. На рис. 3 показаны распределения по числу СЭ, с которыми одновременно контактирует промотор. Распределения для нормальных mESC и злокачественных CH12LX клеток отличаются статистически значимо (критерий Колмогорова—Смирнова, $p = 1e-58$).

Из анализа рис. 3 и табл. 3 можно вывести следующие характерные особенности организации контактов промотор—СЭ в эмбриональных стволовых и злокачественных клетках. Частота сложных контактов промотора одновременно более чем с одним СЭ чувствительна к статусу клетки. Доля промоторов, контактирующих одновременно более чем с одним СЭ, составляет ~40% в

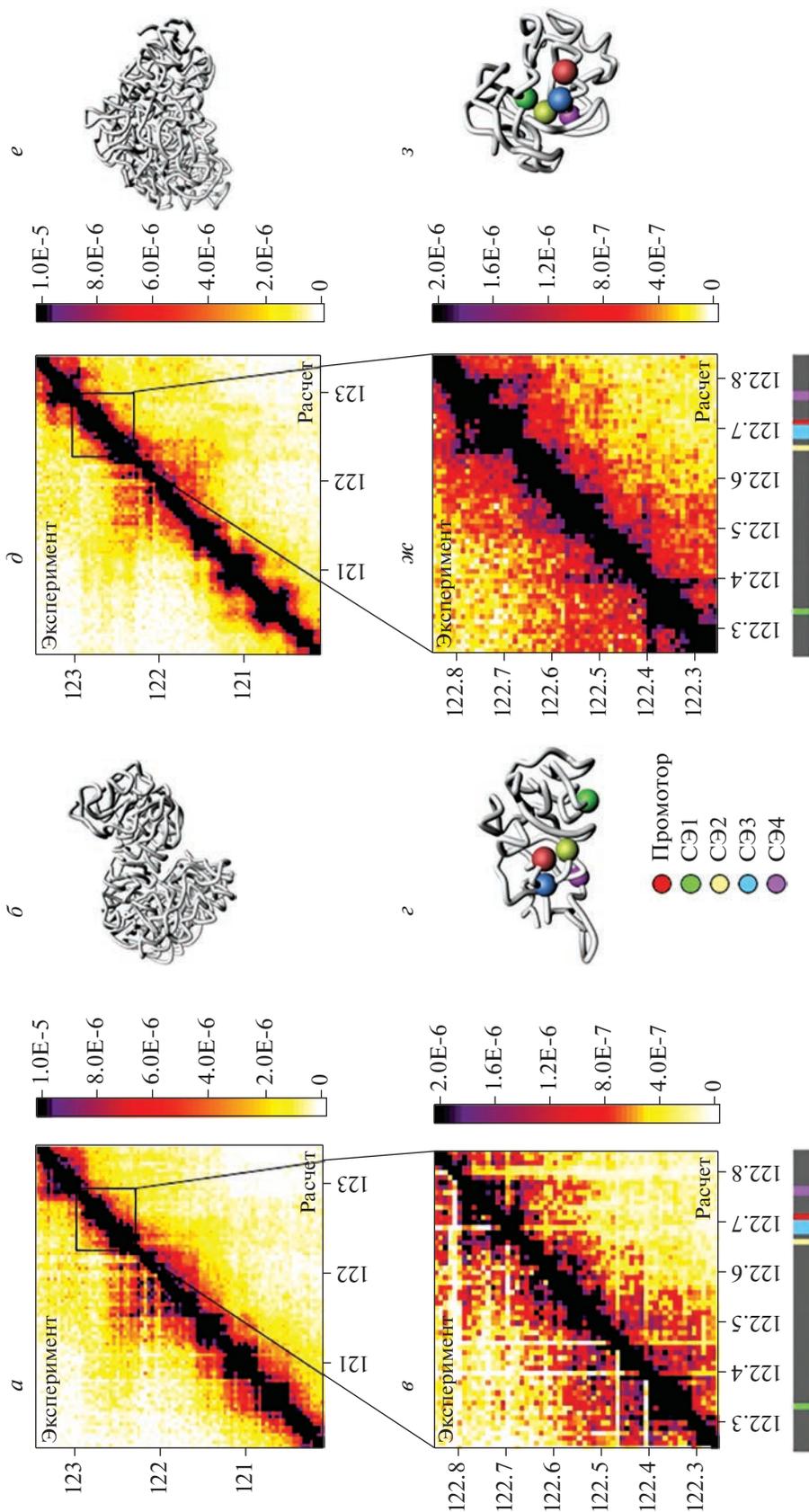


Рис. 1. Экспериментальные и расчетные карты контактов геномных локусов и 3D-конформации 3.5-млн пн участка хроматина, содержащего ген *Nanog*, и ближайших к нему в пределах ~0.4 млн пн в обе стороны суперэнхансеров для мышинной хромосомы 6 в эмбриональных стволовых и злокачественных клетках. *а-г* – нормальные эмбриональные стволовые клетки mESC; *д-з* – злокачественные клетки CH12LX, В-клеточная лимфома; *а, д* – карты контактов участка 120–123.5 млн пн для клеток mESC и CH12LX, расчет в сравнении с экспериментами [11] и [12] соответственно. Как расчетная, так и экспериментальная карта с разрешением 10 тыс. пн укрупнена до 40 тыс. пн. Корреляция Пирсона $R = 0.919$ и 0.955 соответственно; *б, е* – примеры 3D-конформаций из ансамбля структур; *в, ж* – карты контактов ближайшей окрестности гена *Nanog*, где расположены суперэнхансеры, разрешение 10 тыс. пн. Экспериментальные данные [11] и [12]. Под картами приведено положение промотора *Nanog* и суперэнхансеров 1–4 (СЭ1–СЭ4); *г, з* – примеры конформаций. Положение субъединиц, содержащих промотор и СЭ, изображено шарами, размер увеличен для наглядности. Для mESC в среднем характерно больше контактов промотора с СЭ, в том числе множественные контакты в индивидуальных конформациях, чем для CH12LX.

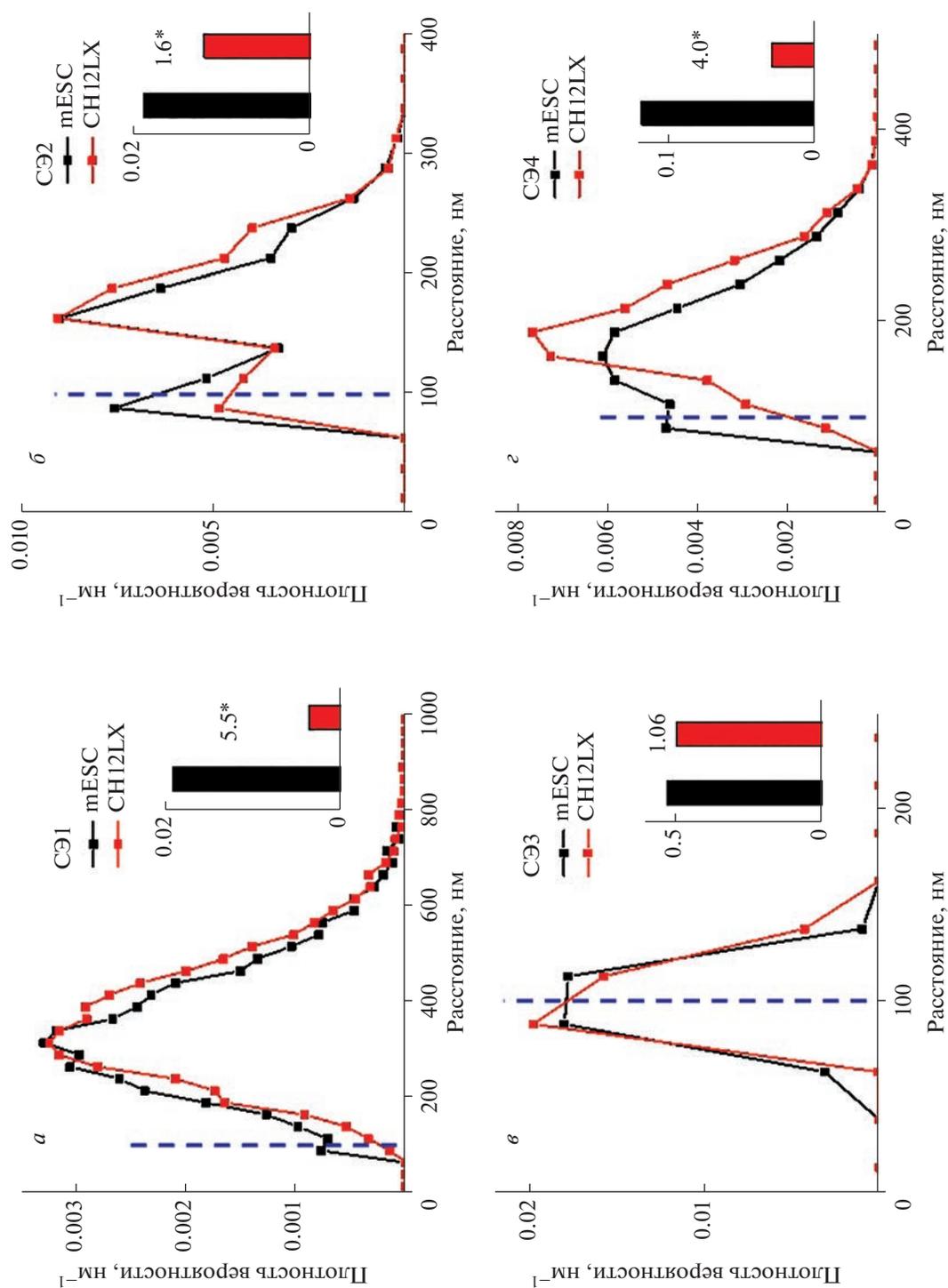


Рис. 2. Ген *Nanog* колокализован с суперэнхансерами в mESC чаще, чем в CH12LX. *а* – распределение расстояний между промотором и СЭ1; *б* – между промотором и СЭ2; *в* – между промотором и СЭ3; *г* – между промотором и СЭ4. Во врезках на всех панелях показана частота конформаций в двух типах клеток, где промотор находится в контакте с СЭ или на расстоянии меньше 100 нм (область левее синего пунктира на графиках). Числа во врезках – соотношение частот в mESC и в CH12LX; звездочками помечены статистически значимые различия по *t*-тесту ($p < 1e-4$). Распределения построены по ансамблям из 4000 конформаций.

Таблица 3. Рассчитанные частоты конформаций с контактами между промотором *Nanog* и суперэнхансерами, одним, двумя и тремя, в ансамблях из 4000 конформаций для нормальных (mESC) и злокачественных (CH12LX) клеток

	mESC	CH12LX	<i>p</i>
Нет контактов	0.072 ± 0.004	0.182 ± 0.006	<1e-5*
СЭ1	0.0035 ± 9e-4	0.0018 ± 7e-4	0.13
СЭ2	0.022 ± 0.002	0.029 ± 0.003	0.06
СЭ3	0.474 ± 0.008	0.547 ± 0.008	<1e-5*
СЭ4	0.015 ± 0.002	0.012 ± 0.002	0.20
СЭ1 + СЭ2	2.5e-4 ± 2.5e-4	2.5e-4 ± 2.5e-4	1
СЭ1 + СЭ3	0.016 ± 0.002	0.0035 ± 9e-4	<1e-5*
СЭ1 + СЭ4	5e-4 ± 4e-4	7.5e-4 ± 4e-4	0.65
СЭ2 + СЭ3	0.229 ± 0.007	0.167 ± 0.006	<1e-5*
СЭ2 + СЭ4	0.005 ± 0.001	0.0015 ± 6e-4	0.009*
СЭ3 + СЭ4	0.111 ± 0.005	0.042 ± 0.003	<1e-5*
СЭ1 + СЭ2 + СЭ3	0.006 ± 0.001	0.0025 ± 8e-4	0.016
СЭ1 + СЭ2 + СЭ4	2.5e-4 ± 2.5e-4	2.5e-4 ± 2.5e-4	1
СЭ1 + СЭ3 + СЭ4	0.0018 ± 7e-4	н/д	0.008*
СЭ2 + СЭ3 + СЭ4	0.044 ± 0.003	0.012 ± 0.002	<1e-5*
СЭ1 + СЭ2 + СЭ3 + СЭ4	7.5e-4 ± 4e-4	н/д	0.08
Всего			
0 СЭ	0.072 ± 0.004	0.182 ± 0.006	<1e-5*
1 СЭ	0.515 ± 0.008	0.589 ± 0.008	<1e-5*
2 СЭ	0.361 ± 0.008	0.215 ± 0.006	<1e-5*
3 СЭ	0.052 ± 0.003	0.015 ± 0.002	<1e-5*
4 СЭ	7.5e-4 ± 4e-4	н/д	0.08

Примечание. Приведены средние значения и стандартные отклонения. н/д – не детектировано. В правой колонке – оценка значимости различий по двухвыборочному *z*-тесту; звездочками помечены различия, значимые с *p* < 0.01.

mESC, где *Nanog* экспрессируется, и ~20% (рис. 3) в CH12LX, где *Nanog* не экспрессируется. В хромосомах злокачественных клеток обнаруживается малая субпопуляция, 1.5% (табл. 3), в которой конформация хроматина такова, что промотор контактирует с тремя СЭ. Для сравнения, в mESC частота этой субпопуляции 5.2% (табл. 3 и столбец “3 СЭ” на рис. 3).

Если сравнивать две клеточные линии по специфическим множественным контактам промотора с конкретными СЭ, то в первую очередь эмбриональные стволовые клетки характеризуются большей, по сравнению со злокачественными клетками, частотой контактов с СЭ1 + СЭ3 (доля субпопуляции хромосом с таким контактом 1.58% vs 0.35%, 4.5-кратная разница), СЭ3 + СЭ4 (11.1% vs 4.2%, 2.6-кратная разница), СЭ2 + СЭ3 + СЭ4 (4.4% vs 1.2%, 3.6-кратная разница), см. табл. 3. В злокачественных клетках больше в 1.5 раза субпопуляция с двойным контактом с СЭ1 + СЭ4, хотя частоты этой субпопуляции малы в обоих типах клеток (0.075% vs 0.05%) и разница статистически не значима.

Различия в частотах контактов на средних картах (рис. 1) предполагают отличие статистических ансамблей конформаций рассматриваемых участков хромосом. Чтобы количественно охарактеризовать гетерогенность структуры хроматина в окрестности гена *Nanog* в популяции хромосом (клеток), был проведен иерархический кластерный анализ ансамблей конформаций в области ~1 млн пн в нормальных и злокачественных клетках (рис. 4).

Кластерный анализ показывает, что поклеточная вариабельность конформаций хроматина в окрестности *Nanog* проявляется в существовании трех основных классов конформаций, различающихся в нормальных и злокачественных клетках структурными свойствами – степенью конденсированности и вовлечением локусов хроматина в дальние–ближние контакты. Доминирующий кластер в обоих типах клеток (72.5% всех конформаций в mESC, 85.9% в CH12LX) содержит более компактные конформации, чем остальные два кластера, представленные структурами в большей степени деконденсированными (рис. 4, а, в). Мы

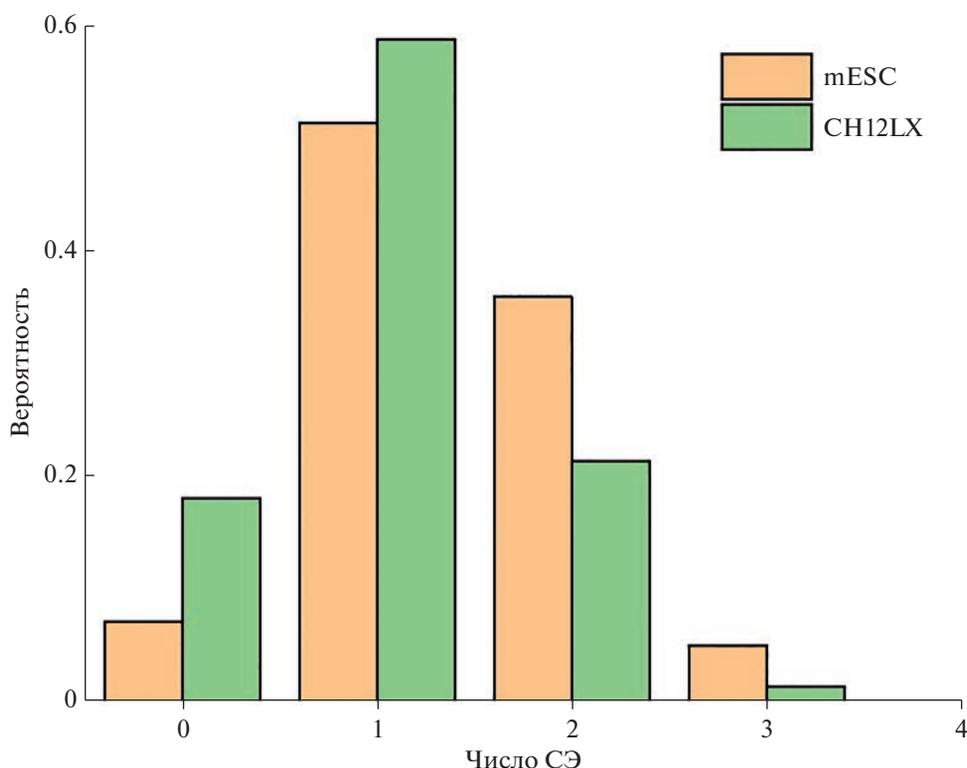


Рис. 3. Вероятность образования простых (1) и множественных (≥ 2) контактов промотора *Nanog* с СЭ в эмбриональных стволовых и злокачественных клетках.

также определили, как по этим кластерам распределены конформации с множественными (>2) контактами промотор–СЭ (рис. 4, б, г). В обоих типах клеток большинство таких конформаций принадлежит к компактным кластерам.

ОБСУЖДЕНИЕ

Суперэнхансеры как элементы регуляции экспрессии генов привлекают внимание с 2013 г., когда была введена сама концепция СЭ [1]. Будучи кластерами энхансеров они обладают повышенной способностью связывать транскрипционные факторы (Oct4, Sox2, Nanog, Klf4, Esrrb) и комплекс Mediator и тем отличаются от обычных энхансеров [14]. Критерием принадлежности локусов к обычным энхансерам или к СЭ в [14] служил уровень сигнала комплекса Mediator, конкретно Med1, хотя в этой работе конкретный порог не был указан.

Мы обнаружили различия в частотах разных комбинаций промотор–СЭ взаимодействий между mESC и злокачественными клетками в ~ 1 млн пн окрестности плюрипотентного гена *Nanog*. Частота взаимодействий промотор–СЭ определяется в модели распределением расстояний между соответствующими локусами в популяции клеток или

статистическим ансамблем конформаций хроматина. Кластерный анализ конформаций в окрестности гена *Nanog* показал больше компактных структур хроматина в злокачественных клетках по сравнению с mESC. С другой стороны, прямой корреляции между множественными контактами промотор–СЭ и компактностью структуры не обнаруживается. Несмотря на более высокую частоту множественных (≥ 3) контактов промотора *Nanog* с СЭ в mESC по сравнению со злокачественными клетками, $0.053 : 0.015$ (табл. 3), в этой окрестности гена *Nanog* доли компактных конформаций в mESC и в CH12LX соотносятся как $72.5 : 85.9$ (рис. 4).

Вероятностные характеристики взаимодействий промотор–СЭ были получены из имеющихся популяционных Hi-C данных путем генерации статистического ансамбля конформаций хроматина, на основе которого были рассчитаны поклеточные (похромосомные) характеристики, распределения расстояний и числа контактов между промоторами и СЭ. Полученные данные демонстрируют сложные взаимоотношения между экспрессией гена, частотой СЭ–промоторных контактов и компактизацией хроматина в обла-

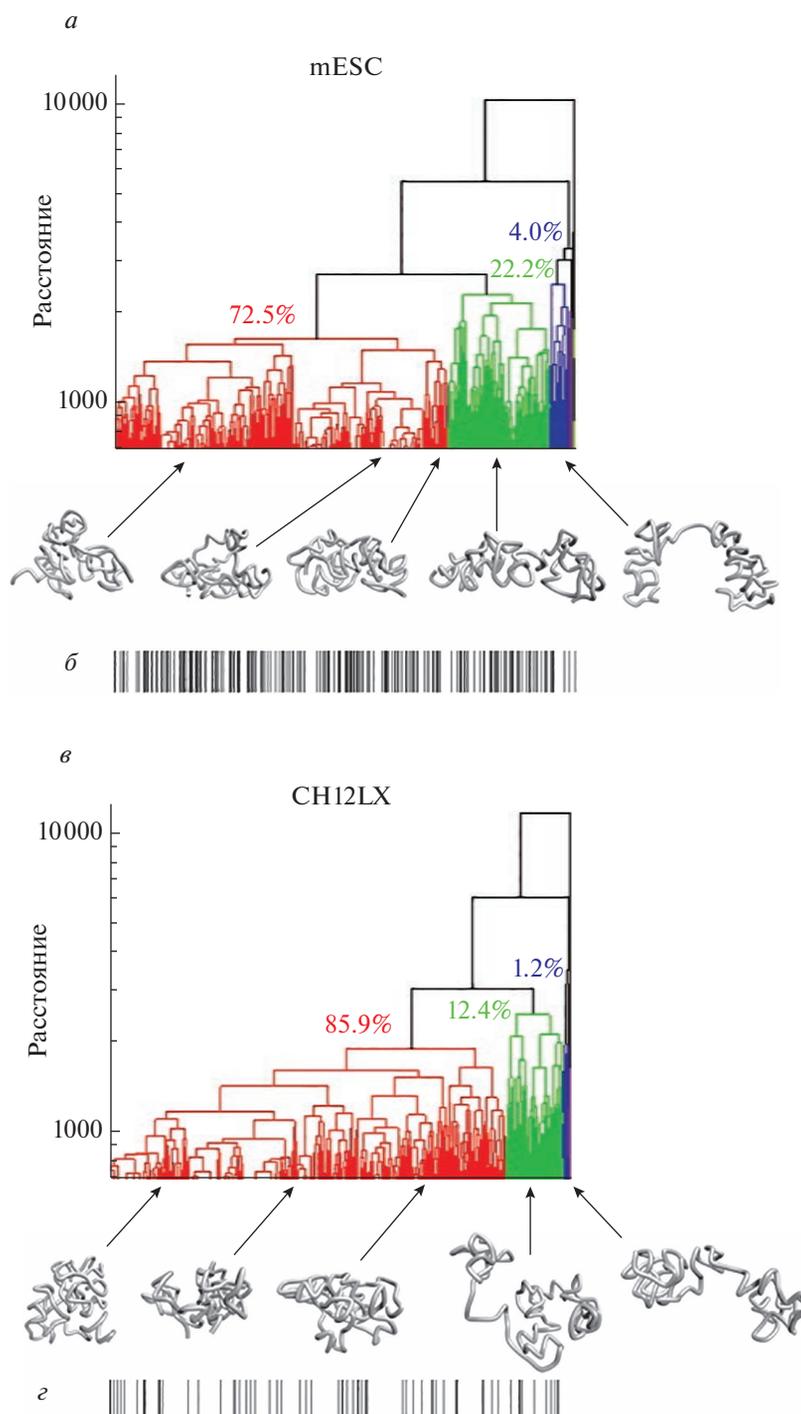


Рис. 4. Анализ поклеточной вариабельности 3Д-структуры хроматина в окрестности гена *Nanog* в клетках mESC и CH12LX. *a, б* – клетки mESC; *в, г* – клетки CH12LX; *a, в* – дендрограммы иерархической кластеризации конформаций хроматина. Найдено оптимальное число кластеров – 7 для mESC и 6 для CH12LX. Мелкие кластеры, включающие в себя менее 1% конформаций, далее не учитывали. На рисунке изображены по три основных кластера. Указан процент конформаций в них и показаны характерные конформации из разных кластеров и субкластеров; *б, г* – расположение на дендрограмме конформаций с множественными (>2) контактами промотора *Nanog* с СЭ.

стях хромосом, содержащих ген и далеко отстоящие энхансеры и СЭ.

Сравнивая с другими модельными подходами к получению информации о промотор-энхансерных и промотор-суперэнхансерных взаимодействиях, можно отметить ряд моментов, по которым наша модель отличается от других. В работе [15] представлен компьютерный подход для анализа Hi-C данных и идентификации энхансер-промоторных взаимодействий. Использовалась обратная, по сравнению с нашей, идеология: искали локусы, с повышенной частотой взаимодействующие с промоторами, и предсказывалось, что в этих локусах находятся энхансеры, которые затем проверялись на известные эпигенетические метки энхансеров, такие как DNКаза I-гиперчувствительные сайты. Такой подход позволяет определять только энхансеры, с которыми контактируют промоторы, но не те, с которыми они не контактируют. В этом принципиальное отличие от настоящей работы, где мы находим субпопуляции хромосом с промоторами как контактирующими, так и не контактирующими с каждым из СЭ.

Взаимодействие промотора с энхансерами изучалось путем моделирования структуры участков хромосом в разных типах клеток в [16]. В фокусе внимания был ген *Pax6*, участвующий в развитии нервной ткани, и два энхансера. Авторы показали взаимосвязь между экспрессией *Pax6* (транскрипционного фактора, участвующего в развитии нервной ткани и глаз) и контактами его промотора с одним из двух энхансеров и отсутствие такой взаимосвязи с другим. Отличие подхода [16] от нашего в том, что там не рассматривались множественные контакты, более чем с одним энхансером одновременно. Структурная организация хромосом на уровне топологически ассоциированных доменов (ТАДов) в клетках mESC моделировалась в [7]. Авторы основывались на 5С-данных, теоретическое описание более полных Hi-C карт контактов, которое доказывало бы справедливость модели, отсутствует. Преимущество нашей модели реконструкции 3Д-структуры хроматина в том, что она основана на количественном описании Hi-C карт контактов с высоким коэффициентом корреляции.

В настоящей работе в обоих типах клеток обнаружена субпопуляция хромосом, в которой промотор *Nanog* взаимодействует более чем с двумя СЭ одновременно. Предсказание субпопуляции злокачественных клеток с множественными контактами промотор-СЭ ставит вопрос, не являются ли они ОСК. *Nanog* активнее экспрессируется в ряде злокачественных опухолей человека, чем в нестволовых клетках [17]. Рассчитанные

нами по данным [9] величины ФРКМ в стволовых и злокачественных клетках (0.6 и 0) предполагают отсутствие экспрессии *Nanog* в клетках CH12LX. Однако информация об уровне экспрессии [9] была получена на большой популяции без клеточного анализа и детектирования субпопуляций со стволовыми маркерами. Поэтому теоретически не исключена возможность того, что линия клеток мышины лимфомы может содержать субпопуляции ОСК с активным геном *Nanog*. Другая возможная интерпретация субпопуляции с множественными контактами промотор-СЭ заключается в том, что предсуществующие контакты являются необходимым, но не достаточным условием экспрессии гена. Пространственная близость энхансер-промотор автоматически не индуцирует экспрессию гена. Контакты энхансер-промотор могут запускать некоторые гены для активации, но для запуска транскрипции требуются дополнительные триггеры или факторы [18]. Существует также возможность бесконтактного механизма действия энхансеров, когда транскрипция гена не требует физического контакта промотора с энхансером, или образования промотор-энхансерных петель [18]. Эти аргументы свидетельствуют в пользу того, что отсутствие прямой корреляции частоты контактов промотор-СЭ с уровнем экспрессии гена, определяемым в среднем по большой популяции клеток, отражает скорее вероятностные взаимоотношения между стохастической экспрессией генов и стохастической организацией хроматина и генома [18–20]. Представленная модель СЭ-промоторных взаимодействий допускает обобщение на случай неконтактного механизма. Для этого условие экспрессии гена через нахождение промотор-СЭ на расстоянии, не превышающем расстояние контакта, достаточно заменить на вероятность экспрессии гена, спадающую с увеличением расстояния между СЭ и промотором. Распределение расстояний между СЭ и промотором определяется 3Д-организацией хроматина и находится предложенным в работе методом.

Говоря о пределах применимости полученных результатов, следует отметить, что мы использовали данные по СЭ для mESC и предполагали, что в злокачественных мышинных клетках СЭ расположены в тех же локусах, что и в нормальных эмбриональных клетках.

Описание геномных данных Hi-C и полимерная реконструкция ансамбля конформаций хроматина в области гена *Nanog* в клетках разного статуса, нормальных и лимфомных, позволяют получить не наблюдаемую в популяционном анализе хроматиновых контактов информацию о частотах субпопуляций клеток с множественными

контактами СЭ—промотор. Продемонстрированы возможности полимерного моделирования применительно к анализу эпигенетической регуляции, СЭ—промоторных взаимодействий, на основе геномных Hi-C данных.

Расчеты были проведены с использованием вычислительных мощностей Межведомственно-го суперкомпьютерного центра РАН.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Hnisz D., Abraham B.J., Lee T.I. et al.* Super-enhancers in the control of cell identity and disease // *Cell*. 2013. V. 155. № 4. P. 934–947. <https://doi.org/10.1016/j.cell.2013.09.053>
2. *Dekker J., Rippe K., Dekker M., Kleckner N.* Capturing chromosome conformation // *Science*. 2002. V. 295. № 5558. P. 1306–1311. <https://doi.org/10.1126/science.1067799>
3. *Lieberman-Aiden E., van Berkum N.L., Williams L. et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome // *Science*. 2009. V. 326. № 5950. P. 289–293. <https://doi.org/10.1126/science.1181369>
4. *Zhang B., Wolynes P.G.* Topology, structures, and energy landscapes of human chromosomes // *Proc. Natl Acad. Sci. USA*. 2015. V. 112. № 19. P. 6062–6067. <https://doi.org/10.1073/pnas.1506257112>
5. *Eidelman Y., Salnikov I., Slanina S., Andreev S.* Chromosome folding promotes intrachromosomal aberrations under radiation- and nuclease-induced DNA breakage // *Int. J. Mol. Sci*. 2021. V. 22. № 22. P. 12186. <https://doi.org/10.3390/ijms222212186>
6. *Beagrie R.A., Scialdone A., Schueler M. et al.* Complex multi-enhancer contacts captured by genome architecture mapping // *Nature*. 2017. V. 543. № 7646. P. 519–524. <https://doi.org/10.1038/nature21411>
7. *Giorgetti L., Galupa R., Nora E.P. et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription // *Cell*. 2014. V. 157. № 4. P. 950–963. <https://doi.org/10.1016/j.cell.2014.03.025>
8. *Annunziatella C., Chiariello A.M., Bianco S., Nicodemi M.* Polymer models of the hierarchical folding of the Hox-B chromosomal locus // *Phys. Rev. E*. 2016. V. 94. № 4-1. P. 042402. <https://doi.org/10.1103/PhysRevE.94.042402>
9. *The ENCODE Project Consortium.* An integrated encyclopedia of DNA elements in the human genome // *Nature*. 2012. V. 489. № 7414. P. 57–74. <https://doi.org/10.1038/nature11247>
10. *Anders S., Pyl P.T., Huber W.* HTSeq – a Python framework to work with high-throughput sequencing data // *Bioinformatics*. 2015. V. 31. № 2. P. 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
11. *Dixon J.R., Selvaraj S., Yue F. et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions // *Nature*. 2012. V. 485. № 7398. P. 376–380. <https://doi.org/10.1038/nature11082>
12. *Rao S.S.P., Huntley M.H., Durand N.C. et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping // *Cell*. 2014. V. 159. № 7. P. 1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021>
13. *Novo C.L., Javierre B.-M., Cairns J. et al.* Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition // *Cell Rep*. 2018. V. 22. № 10. P. 2615–2627. <https://doi.org/10.1016/j.celrep.2018.02.040>
14. *Whyte W.A., Orlando D.A., Hnisz D. et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes // *Cell*. 2013. V. 153. № 2. P. 307–319. <https://doi.org/10.1016/j.cell.2013.03.035>
15. *Ron G., Globerson Y., Moran D., Kaplan T.* Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains // *Nat. Commun*. 2017. V. 8. № 1. P. 2237. <https://doi.org/10.1038/s41467-017-02386-3>
16. *Liu L., Kim M.H., Hyeon C.* Heterogeneous loop model to infer 3D chromosome structures from Hi-C // *Biophys. J*. 2019. V. 117. № 3. P. 613–625. <https://doi.org/10.1016/j.bpj.2019.06.032>
17. *Wang M.-L., Chiou S.-H., Wu C.-W.* Targeting cancer stem cells: emerging role of Nanog transcription factor // *Onco. Targets Ther*. 2013. V. 6. P. 1207–1220. <https://doi.org/10.2147/OTT.S38114>
18. *Schoenfelder S., Fraser P.* Long-range enhancer-promoter contacts in gene expression control // *Nat. Rev. Genet*. 2019. V. 20. № 8. P. 437–455. <https://doi.org/10.1038/s41576-019-0128-0>
19. *Finn E.H., Misteli T.* Molecular basis and biological function of variability in spatial genome organization // *Science*. 2019. V. 365. № 6457. P. eaaw9498. <https://doi.org/10.1126/science.aaw9498>
20. *Sood V., Misteli T.* The stochastic nature of genome organization and function // *Curr. Opin. Genet. Dev*. 2022. V. 72. P. 45–52. <https://doi.org/10.1016/j.gde.2021.10.004>

Impact of Chromatin 3D-Organization on Promoter–Superenhancer Interactions in Embryonic Stem vs Cancer Cells

Yu. A. Eidelman^{a, b, *} and S. G. Andreev^{a, b, **}

^a*Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, 119334 Russia*

^b*National Research Nuclear University “MEPHI”, Moscow, 115409 Russia*

**e-mail: eidel@mail.ru*

***e-mail: andreev_sg@mail.ru*

The interaction of enhancers and superenhancers (SE) with promoters is functionally significant for the regulation of gene expression. Pattern of these interactions plays a key role in various processes, such as differentiation, malignant transformation, etc. In order to quantify the relationship between 3D chromatin organization and promoter–SE contacts, a computational analysis of chromatin conformations near the murine *Nanog* pluripotency gene was performed for normal embryonic stem (mESC) and lymphoma (CH12LX) cells. Using biophysical modeling approach, the following parameters of the promoter–SE interactions were identified: the distribution of distances between the *Nanog* promoter and the SEs, the frequency of contacts with one and several SEs simultaneously. In normal mESC expressing *Nanog*, the frequency of contacts of promoters with SEs is higher than in cancer cells, and complex contacts with two or more SEs are more frequent. The modelling reveals a small subpopulation of cancer cells, where the promoter contacts simultaneously three SEs. The predicted subpopulation of cancer cells with multiple promoter–SE contacts may be predisposed to increased stemness and hypothetically be considered as a reservoir for generation of cancer stem cells.

Keywords: *Nanog*, promoter–enhancer interactions, Hi-C, polymer modeling, molecular dynamics, chromatin structure.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНСТРУМЕНТОВ ДЛЯ ПОИСКА СТРУКТУРНЫХ ВАРИАНТОВ НА ДАННЫХ КОРОТКОРИДОВОГО ПОЛНОГЕНОМНОГО СЕКВЕНИРОВАНИЯ

© 2023 г. А. А. Мкртчян^{1, *, #}, К. С. Грамматикати^{1, #}, П. Г. Казакова^{1, #}, С. И. Митрофанов^{1, #}, П. Ю. Земский¹, А. А. Ивашечкин¹, М. Н. Пилипенко¹, Д. В. Светличный¹, А. П. Сергеев¹, Е. А. Снигирь¹, Л. В. Фролова¹, Т. А. Шпакова¹, В. С. Юдин¹, А. А. Кескинов¹, С. М. Юдин¹, В. И. Скворцова²

¹Федеральное государственное бюджетное учреждение Центр стратегического планирования и управления медико-биологическими рисками здоровью

Федерального медико-биологического агентства России, Москва, 119121 Россия

²Федеральное медико-биологическое агентство России, Москва, 123182 Россия

*e-mail: AMkrtyan@cspsfmba.ru

Поступила в редакцию 25.10.2022 г.

После доработки 16.01.2023 г.

Принята к публикации 19.01.2023 г.

Проанализирована работа трех биоинформатических инструментов (Manta, Smoove, Delly) для поиска структурных вариантов на данных полногеномного секвенирования с использованием четырех разных алгоритмов выравнивания: DRAGEN, GDC DNA-SeqAlignment Workflow, GDC DNA-SeqAlignmentWorkflow + GDC DNA-SeqCo-cleaning Workflow, NovoAlign, разных длин “сырых” прочтений: 2×150 и 2×250 пн, разных показателей среднего покрытия генома. Полученные показания соотнесены с эталонными результатами команды GIAB. Проведена проверка выявленных структурных вариантов с помощью секвенирования по методу Сэнгера. Выявили, что структурные варианты типа делеция и инсерция лучше всего определяются с помощью инструмента Manta. Мы получили 89–96% точности и 59–70% чувствительности для проанализированных делеций, а также 96–99% точности и 15–36% чувствительности для инсерций. Smoove и Delly продемонстрировали менее точные и чувствительные результаты (Smoove: 91–95% точности и 8–54% чувствительности для делеций; Delly: 78–87% точности и 31–66% чувствительности для делеций, 99–100% точности и 1–13% чувствительности для инсерций). Показали, что использование комбинации из двух и даже трех инструментов не повышает процента подтвержденных структурных вариантов для типов делеция, инсерция, дупликация. Проведенный анализ выявил, что с увеличением среднего значения покрытия генома увеличиваются показатели точности и чувствительности работы биоинформатических инструментов для поиска структурных вариантов, а увеличение длины прочтений с 150 до 250 пн влияет в разной степени на точность и чувствительность работы отдельных инструментов. Также сделали вывод, что точность работы инструментов для поиска структурных вариантов варьируется в зависимости от диапазона размеров структурных вариантов. Например, Manta лучше находит делеции в диапазоне от 200 и более пн, Delly – от 1000 до 10000 пн, Smoove – от 200 до 10000 пн.

Ключевые слова: структурные варианты, полногеномное секвенирование, секвенирование на коротких прочтениях, выравнивание, GIAB.

DOI: 10.31857/S0016675823060115, **EDN:** SUEMFU

Однонуклеотидные полиморфизмы (SNP) являются наиболее изученными генетическими вариантами. Клиническая значимость именно однонуклеотидных генетических вариантов чаще всего интерпретируется в рамках работы различных герминальных и соматических пайплайнов. Первоначально считалось, что именно SNP вносят наиболее существенный вклад в генетическое

разнообразие людей, обеспечивая геномную вариативность между двумя индивидуумами около 0.1%. Однако сейчас установлено, что структурные варианты в геноме человека также дают значительный вклад в генетическое разнообразие, увеличивая геномную вариативность до 1.5% [1, 2].

Структурными вариантами называются генетические перестройки длиной не менее 50 пн [3]. Структурные варианты делятся на простые (делеции, дупликации, инсерции, инверсии, трансло-

[#] Вклад этих авторов в работу равнозначный.

кации) и комплексные, к которым относятся, например, хромоплексия и хромотриписис.

Выделяют отдельный подтип структурных вариантов – вариации числа копий (CNV), при котором количество копий некоторого участка ДНК варьирует в геномах разных индивидуумов [4]. CNV образуются в основном за счет делеций и дупликаций участков ДНК размером более 1000 пн [5–8].

По другому основанию структурные варианты делятся на сбалансированные и несбалансированные. Несбалансированные представляют собой увеличение или уменьшение количества фрагментов ДНК. Сбалансированные структурные варианты возникают в результате генетической перестройки без изменения количества копий участков ДНК, что характерно для инверсий и транслокаций [9].

Научные исследования в области молекулярной биологии и медицинской генетики показали, что структурные варианты играют значимую роль в развитии ряда фенотипических нарушений [10, 11] и таких тяжелых заболеваний, как синдром Паркинсона [12], склерозирующая пневмоцитоза [13], гематологические опухоли [14], рак груди и яичников [15]. Кроме того, многие заболевания, в том числе шизофрения [16] и спорадический рак простаты [17], ассоциированы с наличием в геноме CNV [18]: некоторые нейродегенеративные заболевания, например повышенный риск развития аутоиммунных заболеваний (системная красная волчанка [19], болезнь Крона [20], псориаз [21] и др.) также связан с наличием вариаций числа копий в геноме человека.

Установление перечисленных выше ассоциаций и зависимостей привело исследователей к выводу о том, что понимание патогенеза многих трудноизлечимых заболеваний невозможно без изучения и генотипирования структурных вариантов.

В настоящее время исследования генома ведутся в основном на базе более доступного секвенирования с помощью коротких прочтений, хотя технологии секвенирования с помощью длинных прочтений позволяют более детально изучать структурные варианты.

Данные секвенирования используются в первую очередь для выявления новых и идентификации существующих однонуклеотидных и структурных вариантов биоинформатическими методами. Этот процесс осложняется тем, что многие структурные варианты встречаются в тех регионах генома, которые богаты повторами. При этом точность идентификации вариантов во многом зависит от выравнивания прочтений, качество которого может снижаться в таких сложных регионах, как участки с tandemными повторами, теломерные участки хромосом и др. [5]. Соответственно глубокое изучение структурных вариантов невоз-

можно без совершенствования биоинформатических инструментов, предназначенных для поиска структурных вариантов и внедрения их в пайплайны.

Команда GIAB опубликовала эталонный файл для выявления структурных вариантов [22]. Этот файл в формате VCF был получен для образца HG002 на основе данных секвенирования с помощью четырех технологий при использовании более 30 биоинформатических инструментов. Используя этот материал, мы смогли оценить полученные нами данные о качестве работы трех биоинформатических инструментов: Manta, Smoove и Delly. Кроме того, мы провели проверку с помощью секвенирования по методу Сэнгера нескольких структурных вариантов (разных типов и длин) на биоматериале образцов от двух доноров.

МАТЕРИАЛЫ И МЕТОДЫ

Сбор биоматериала

Для данного исследования взяты образцы цельной крови от двух доноров (не HG002), подписавших добровольное информированное согласие. Забор биоматериала осуществлялся в соответствии с ГОСТ P53079.4-2008. Транспортировка биоматериала велась при постоянном контроле температурного режима. Образцы прошли проверки качества, включая проверку на отсутствие признаков гемолиза и хилеза. Биоматериал подвергался заморозке не более одного раза.

Выделение ДНК, подготовка геномных библиотек и секвенирование образцов с применением технологии Illumina

Выделение ДНК из образцов цельной крови проводили с помощью набора MagAttract HMW DNA Kit (Qiagen, Германия). Протокол выделения ДНК автоматизирован на станции Tecan Freedom EVO (Tecan, Швейцария). Измерение концентрации и чистоты выделенной ДНК осуществляли с помощью микропланшетного ридера Tecan Infinite 200 PRO F Nano+ (Tecan).

Геномные библиотеки для секвенирования готовились с использованием набора Nextera DNA Flex (Illumina, США) в соответствии с рекомендациями производителя. Каждый образец в проточной кювете был помечен с помощью индексов IDT-ILMN Nextera DNA UD (Illumina) для исключения возможности перекрестной контаминации.

Измерение концентрации геномных библиотек проводили на микропланшетном ридере Tecan Infinite 200 PRO F Nano+. Размер вставки геномных библиотек измеряли на системе Agilent 4200 TapeStation System (Agilent, США) с использованием набора Agilent DNA 1000 Kit (Agilent).

Библиотечные пулы включали помимо двух целевых образцов 22 образца из других исследований, так как одна ячейка S4 используемого секвенатора вмещает 24 образца, объединение которых осуществлялось с использованием автоматизированной станции Tecan Freedom EVO (Tecan). Размер вставки для каждого образца составил 323 и 333 пн.

Полногеномное секвенирование проводилось с использованием секвенатора Illumina NovaSeq 6000 (Illumina) и комплекта реагентов NovaSeq 6000 S4 (300 циклов) (Illumina) для парноконцевых прочтений 2×150 пн.

Секвенирование образцов по методу Сэнгера

Выделение геномной ДНК проводили из 200 мкл замороженной цельной крови с использованием набора MagAttract HMW DNA Kit в соответствии с протоколом производителя. Элюция проводилась в 100 мкл воды, свободной от ДНКаз.

Концентрацию образцов измеряли с помощью набора QuantiFluor dsDNA System (Promega, США) на флуориметре Quantus (Promega). Измерение чистоты выделенной ДНК проводилось с использованием спектрофотометра NanoDrop8000 (Thermo Scientific, США). Отбирались образцы ДНК, имеющие соотношения длин волны A260/280 в пределах 1.7–1.9 и A230/260 в пределах 1.8–2.2.

Для секвенирования по методу Сэнгера была сформирована выборка из 120 локусов с различными типами структурных вариантов. В выборку вошли 60 делеций (50%), 28 инсерций (23.3%), 23 дубликации (19.2%) и 9 инверсий (7.5%).

С помощью программного обеспечения Primer-BLAST [23] удалось подобрать праймеры к 17 делециям, 12 инсерциям, 12 дубликациям и двум инверсиям. Праймеры подобраны отдельно к прямой и обратной цепи ДНК.

Для синтеза праймеров использовались олигонуклеотидные синтезаторы OligoMaker X12 (OligoMaker ApS, Дания) и Polygen (Polygen GmbH, Германия).

Для наработки ампликонов проводили ПЦР со следующим температурным циклом: начальная денатурация $95^{\circ}\text{C} - 5$ мин; 35 циклов $95^{\circ}\text{C} - 15$ с, $60^{\circ}\text{C} - 30$ с, $72^{\circ}\text{C} - 30$ с; финальная элонгация $72^{\circ}\text{C} - 5$ мин; хранение при $4^{\circ}\text{C} -$ неограниченное время.

Для проведения ПЦР отбиралось 50–100 нг ДНК. После завершения ПЦР проводилась очистка ампликонов с помощью магнитных частиц Sam-

ple Purification Beads (Beckman Coulter, США) по следующей схеме:

- к 25 мкл образца добавляется 25 мкл частиц;
- смесь перемешивается и инкубируется в течение 10 мин при комнатной температуре;
- смесь ставится на магнитный штатив в ожидании просветления;
- после просветления супернатант удаляется;
- осадок дважды промывается 200 мкл 80%-ного этилового спирта;
- промытый осадок сушится на воздухе в течение 3–5 мин и элюируется 25 мкл Resuspension Buffer (Illumina);
- элюированный осадок инкубируется при комнатной температуре 2 мин, ставится на магнитный штатив;
- очищенные ампликоны переносятся в новые пробирки в объеме 20 мкл.

Контроль очистки проводили с помощью Agilent 4200 TapeStation System и набора реагентов D1000 Reagents (Agilent, США) согласно инструкции производителя.

Секвенирующая и терминирующая реакции проводились с использованием набора реагентов BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, США) согласно рекомендациям производителя. В каждую реакционную пробирку добавлялось 50–100 нг ДНК-матрицы образца.

Продукты реакции анализировали методом капиллярного электрофореза в автоматическом секвенаторе SeqStudio Genetic Analyzer (Applied Biosystems). При последующем анализе учитывали только те варианты, для которых были получены прямые и обратные прочтения.

Биоинформатическая обработка данных полногеномного секвенирования

Для контроля качества секвенирования в целом использовалась программа Illumina Sequencing Analysis Viewer v2.4.7 [24].

На первом этапе обработки данных полногеномного секвенирования осуществлялась демультипликация, при которой исходная выдача секвенатора Illumina NovaSeq 6000 конвертировалась из формата BCL в формат FASTQ с помощью программного обеспечения bcl2fastq v2.20 [25]. Для контроля качества прочтений применялся биоинформатический инструмент FastQC v0.11.9 [26]. Качество анализируемых образцов провере-

но по следующим метрикам: показатель равномерности распределения нуклеотидов в прочтениях; G/C состав; показатель Q30.

На втором этапе биоинформатической обработки проводилось выравнивание прочтений на референсный геном с помощью программно-аппаратной платформы Illumina DRAGEN Bio-IT Platform v07.021.510.3.5.7 [27]. В качестве референсного генома человека использовалась последовательность GRCh38.d1.vd1 [28]. После выравнивания оценивалось среднее значение покрытия. У всех анализируемых образцов среднее покрытие по геному составило не менее 30×.

Затем в файлах выравнивания осуществляли поиск структурных вариантов с помощью трех различных инструментов: Manta v1.6.0 [29], Smoove v0.2.6 [30] и Delly v0.9.1 [31].

Так как эталонные VCF- и BED-файлы GIAB сделаны для версии генома GRCh37 [32], а данный анализ проводился с выравниванием на версию генома GRCh38, то для того, чтобы отфильтровать полученные VCF-файлы по подтвержденным регионам эталонного BED-файла, его координаты конвертировались с GRCh37 на GRCh38 с помощью инструмента CrossMap [33].

Для проверки генетических вариантов секвенированием по методу Сэнгера отбирались делеции разной длины, а также дубликации и инверсии длиной до 500 пн, которые были найдены всеми тремя инструментами или двумя из них. Кроме того, отбирались инсерции длиной до 500 пн, которые были найдены хотя бы одним из инструментов Manta и Delly.

Биоинформатическая обработка данных HG002 и сравнение с эталоном

Для валидации качества работы инструментов, предназначенных для поиска структурных вариантов, с сайта NCBI были скачаны эталонные файлы: HG002_SVs_Tier1_v0.6.VCF.GZ, HG002_SVs_Tier1_v0.6.VCF.GZ.TBI и HG002_SVs_Tier1_v0.6.BED (https://ftp.ncbi.nih.gov/giab/ftp/data/AshkenazimTrio/analysis/NIST_SVs_Integration_v0.6/), а также прочтения образца HG002, полученные с помощью секвенатора Illumina HiSeq (2 × 250 пн) в формате FASTQ.GZ: https://ftp.ncbi.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG002_NA24385_son/NIST_Illumina_2x250bps/reads/, 68 файлов общим объемом 163 Гб, предполагаемое среднее покрытие – 68×. Размер вставки составил 350 пн.

Также были скачаны часть прочтений образца HG002, полученная с помощью секвенатора Illumina HiSeq (2 × 150 пн) в формате FASTQ.GZ: https://ftp.ncbi.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG002_NA24385_son/NIST_HiSeq_HG002_Homogeneity-10953946/HG002_HiSeq300x_fastq/, три первые по алфавиту директории, которые далее по тексту будем обозначать d1, d2, d3, общим объемом 174 Гб, предполагаемое среднее покрытие для каждой директории – 25×.

Затем были подготовлены директории с объединенными прочтениями: d1 + d2 с предполагаемым средним покрытием 50× и d1 + d2 + d3 с предполагаемым средним покрытием 75×.

Дополнительно был скачан BAM-файл с выравниванием прочтений образца HG002, полученных с секвенатора Illumina HiSeq (2 × 250 пн), инструментом NovoAlign на 37-ю версию референсного генома человека (hs37d5): https://ftp.ncbi.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG002_NA24385_son/NIST_Illumina_2x250bps/novoalign_bams/HG002.hs37d5.2x250.bam, 122 Гб, среднее покрытие 72×.

Размер вставки для данных прочтений составил 550 пн.

Для скачанных прочтений проведено выравнивание на референсный геном GRCh37 тремя способами: с помощью программно-аппаратной платформы DRAGEN, с помощью пайплайна GDC DNA-Seq Alignment Workflow (DNA-Seq A), а также с помощью пайплайна GDC DNA-Seq Alignment + Co-cleaning Workflow (DNA-Seq A + C) [34], в котором для выравнивания используется программа BWA [35].

Для скачанного BAM-файла выравнивания, выполненного инструментом NovoAlign, проведена сортировка и индексация с помощью инструмента samtools [36].

Во всех файлах с выравниваниями осуществлен поиск структурных вариантов с помощью инструментов Manta v1.6.0, Smoove v0.2.6 и Delly v0.9.1. Затем проведена дополнительная обработка результирующего VCF-файла. Для файлов Manta это была конвертация структурных вариантов типа BND в INV с помощью скрипта разработчиков данного инструмента, а также дополнительная аннотация инструментом Duphold v0.2.3 [37] с последующей фильтрацией в соответствии с рекомендациями разработчика (Приложение 1). Кроме того, для всех файлов проведена фильтрация с помощью bcftools 1.15 [36] (Приложение 1).

Таблица 1. Структурные варианты в эталонном файле HG002_SVs_Tier1_v0.6.vcf.gz

Количество вариантов										
до фильтрации, ед.	после фильтрации PASS									
	всего		в том числе типа:							
	ед.	%	DEL		INS		DUP		INV	
			ед.	%	ед.	%	ед.	%	ед.	%
74012	12745	100	5464	43	4462	35	2819	22	0	0

Примечание. DEL – делеции, INS – инсерции, DUP – дубликации, INV – инверсии.

При этом для Smoove по умолчанию брались варианты без фильтрации по полю FILTER, так как инструмент не проводит соответствующей фильтрации.

Сводные VCF-файлы готовили с помощью инструмента SURVIVOR v1.0.7 [38]. Для сравнения всех полученных VCF-файлов с эталонным VCF-файлом использовался инструмент witty.er v0.3.4 [39], конфигурационный файл для которого скачан из репозитория инструмента на ресурсе GitHub – <https://github.com/Illumina/witty.er/blob/master/docs/examples/SV/config.json>.

Команды запуска инструментов bcftools, SURVIVOR и witty.er представлены в Приложении 1.

Биоинформатическая обработка данных секвенирования по методу Сэнгера

Файлы прочтений прямой и обратной последовательностей в формате AB1 сконвертированы в формат FASTQ с помощью программы Unipro UGENE [40]. Далее эти прочтения выравнены на референсный геном GRCh38.d1.vd1 с помощью инструмента BWA, а полученные выравнения проанализированы в программе IGV [41], предназначенной для просмотра генетических данных. К выравнению применялось окрашивание soft-clipped-прочтений, т.е. таких, у которых часть прочтения выравняется на референс, а часть нет, и при этом прочтение остается в выравнении. Кроме того, использовалось окрашивание прочтений по последовательностям. Визуализация варианта помогала сделать вывод о наличии или отсутствии структурного варианта в образце.

РЕЗУЛЬТАТЫ

Валидация работы инструментов поиска структурных вариантов

Командой GIAB опубликован эталонный набор данных структурных вариантов в формате

VCF для версии генома GRCh37, а также BED-файл для образца HG002 на основе данных секвенирования с помощью четырех разных технологий. Для оценки качества работы биоинформатических инструментов, предназначенных для поиска структурных вариантов, с сайта NCBI скачаны указанные выше эталонные файлы. Эти файлы использовались для сравнения с полученными данными. В табл. 1 представлена статистика по вариантам эталонного VCF-файла Tier1_v0.6.

Как следует из табл. 1, в эталонном файле GIAB после фильтрации по фильтру “PASS” и по типу вариантов “не равно BND” (breakends) остается немногим более 17% структурных вариантов, большинство из которых (около 43%) составляют делеции, тогда как инверсии отсутствуют.

Эталонный BED-файл Tier1_v0.6 включает в себя 34830 регионов на всех хромосомах, за исключением митохондриальной.

Для прочтений образца HG002 нами были запущены три пайплайна выравнивания: DRAGEN, GDC’s DNA-Seq Alignment Workflow и GDC’s DNA-Seq Alignment + Co-cleaning Workflow. В качестве входных данных для пайплайнов выступали прочтения разной длины и глубины секвенирования, скачанные с сайта NCBI по протоколу FTP. Файлы с выравниваниями сортировались и индексировались с помощью инструмента samtools.

На сортированных BAM-файлах были запущены три биоинформатических инструмента: Manta, Delly и Smoove. В случае с Manta для VCF-файлов были применены аннотация программой Duphold и фильтрация на основе аннотированных метрик.

В табл. 2 представлена характеристика трех программ, детектирующих различные типы структурных вариантов, на основе данных секвенирования короткими прочтениями (2×150).

Таблица 2. Программы, используемые для поиска структурных вариантов

Название, версия и дата последнего обновления программы	Используемые программой методы выявления структурных вариантов	Определяемые типы структурных вариантов				
		DEL	INS	DUP	INV	TRA
Manta v1.6.0, 07.2019	read pairs + split reads + assembly	+	+	+	+	+
Delly v0.9.1, 12.2021	read pairs + split reads	+	+	+	+	+
Smoove v0.2.6, 10.2021	read pairs + split reads + read depth	+	–	+	+	+

Примечание. TRA – транслокации.

На рис. 1 представлены 16 диаграмм, на которых изображены результаты работы биоинформатических инструментов Manta, Smoove, Delly, а также результаты совместной работы инструментов Manta и Smoove, Manta и Delly, Delly и Smoove, Manta, Smoove и Delly. Каждому из инструментов, а также их парам и тройке присвоен свой цвет. Результаты работы каждого из инструментов по отдельности, а также их совместной работы обозначены прямоугольниками разных цветов. Высота каждого из цветных прямоугольников соответствует количеству найденных структурных вариантов, которые показаны в следующей разбивке: до фильтрации (Total) и после фильтрации (PASS), в том числе типов DEL (делеции), INS (инсерции), DUP (дубликации) и INV (инверсии). В надписях над каждой диаграммой указан использованный способ выравнивания, длина прочтений (в пн) и среднее покрытие. Рис. 1 показывает, что Delly нашел больше вариантов во всех случаях, но более половины не прошли фильтрации (PASS) при выборе пайплайнов DNA-Seq A и DNA-Seq A + C. При этом количество вариантов, найденных Delly, сильно росло с ростом покрытия (в 1.5 раза), а также при выборе пайплайнов выравнивания DNA-Seq A и DNA-Seq A + C (более чем в 2 раза). Smoove нашел меньше вариантов в сравнении с Delly и Manta во всех случаях. Количество вариантов, найденных Smoove и Manta, также увеличивалось с увеличением покрытия, но не так сильно, как для Delly (в 1.1 раз). При выборе пайплайнов выравнивания DNA-Seq A и DNA-Seq A + C количество вариантов незначительно увеличивалось для Smoove (в 1.3 раза), а для Manta почти не изменялось. Использование пайплайна выравнивания NovoAlign практически сводит к нулю число структурных вариантов типа DEL, находимых инструментом Smoove.

На рис. 2 и 3 представлены сопоставимые количественные оценки качества работы разных биоинформатических инструментов по выявлению структурных вариантов типа DEL и INS соответственно. Показано влияние выравнивания разными пайплайнами с входными данными с разным предполагаемым покрытием. Оценка качества работы инструментов проводилась посредством расчета и анализа трех показателей: точности, чувствительности и Fscore (суммарный показатель точности и чувствительности).

Как видно из рис. 2 и табл. 3, показатель чувствительности возрастает на 3–7% для инструментов Manta, Delly и Smoove при увеличении покрытия с 25x до 50x, а при увеличении покрытия с 50x до 75x почти не меняется. Показатель точности для инструментов Manta, Smoove и Delly падает при изменении покрытия с 25x до 75x. Таким образом, увеличение покрытия с 25x до 75x приводит к некоторому повышению качества работы двух из трех биоинформатических инструментов, предназначенных для поиска структурных вариантов типа DEL, по показателю Fscore. Так, показатель Fscore для инструмента Smoove увеличивается с 65–66 до 67–68%, а для инструмента Manta – с 73–74 до 78–79%. Для инструмента Delly такой зависимости не выявлено.

Увеличение длины прочтений со 150 до 250 пн по-разному влияет на качество работы разных инструментов при разных пайплайнах выравнивания. Для инструмента Delly показатель Fscore для вариантов типа DEL увеличивается с 64.5 (в среднем) до 67 при пайплайне DRAGEN (точность и чувствительность увеличиваются на 2–3% и до 72 при пайплайне DNA-Seq Alignment (точность увеличивается на 3%, чувствительность на 11%). А при пайплайне DNA-Seq Alignment + Co-cleaning этот показатель повышается с 65.8 (в среднем) до 74% (точность повышается на 4%,

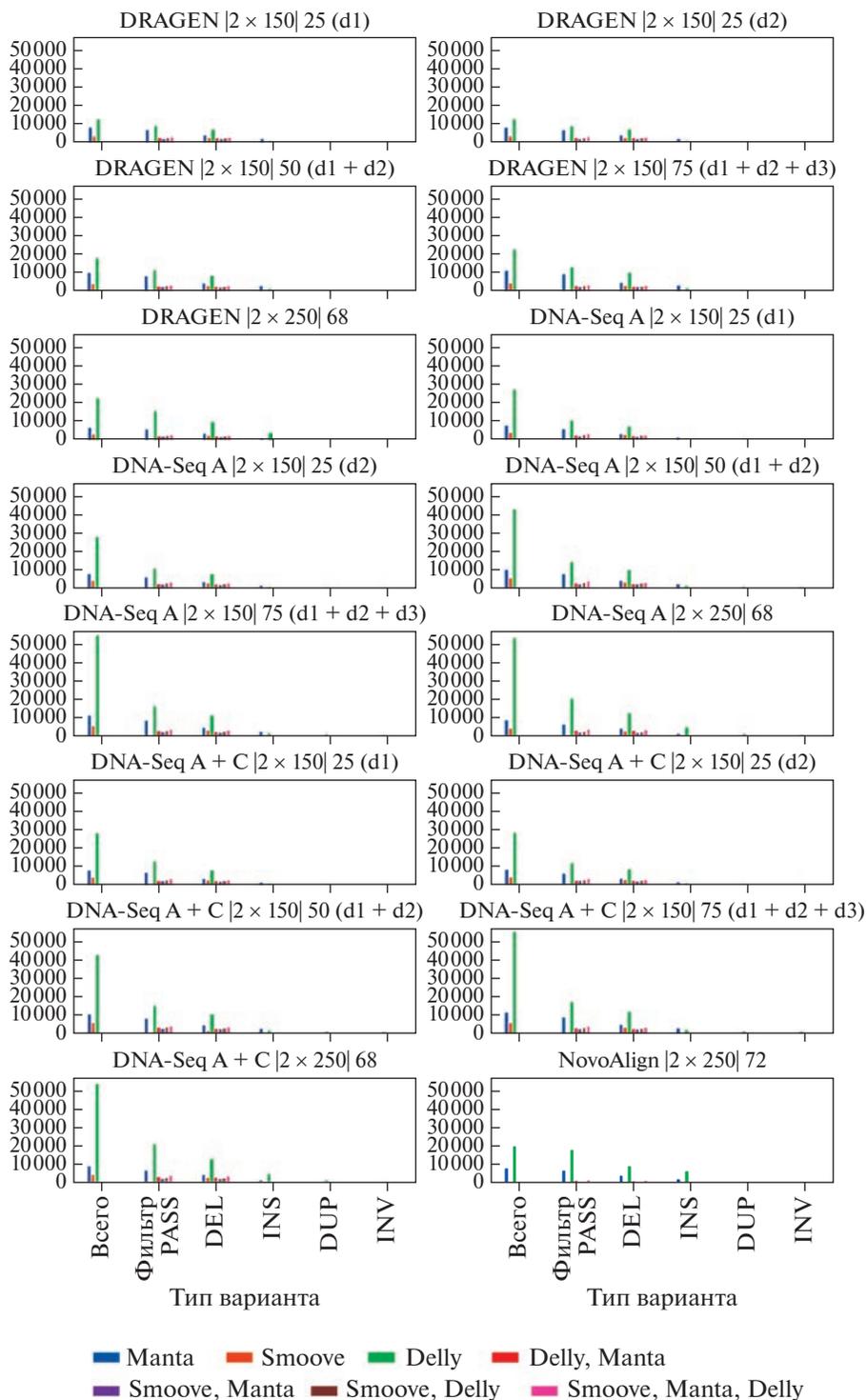


Рис. 1. Структурные варианты, найденные с помощью разных биоинформатических инструментов на результатах выравнивания разными пайплайнами.

чувствительность на 10%). Для инструмента Smoove показатель Fscore при увеличении длины прочтений на 100 пн практически не меняется при всех трех пайплайнах (соответственно точ-

ность и чувствительность также практически не меняются). В то же время для инструмента Manta этот показатель при пайплайнах DRAGEN и DNA-Seq Alignment + Co-cleaning изменяется не-

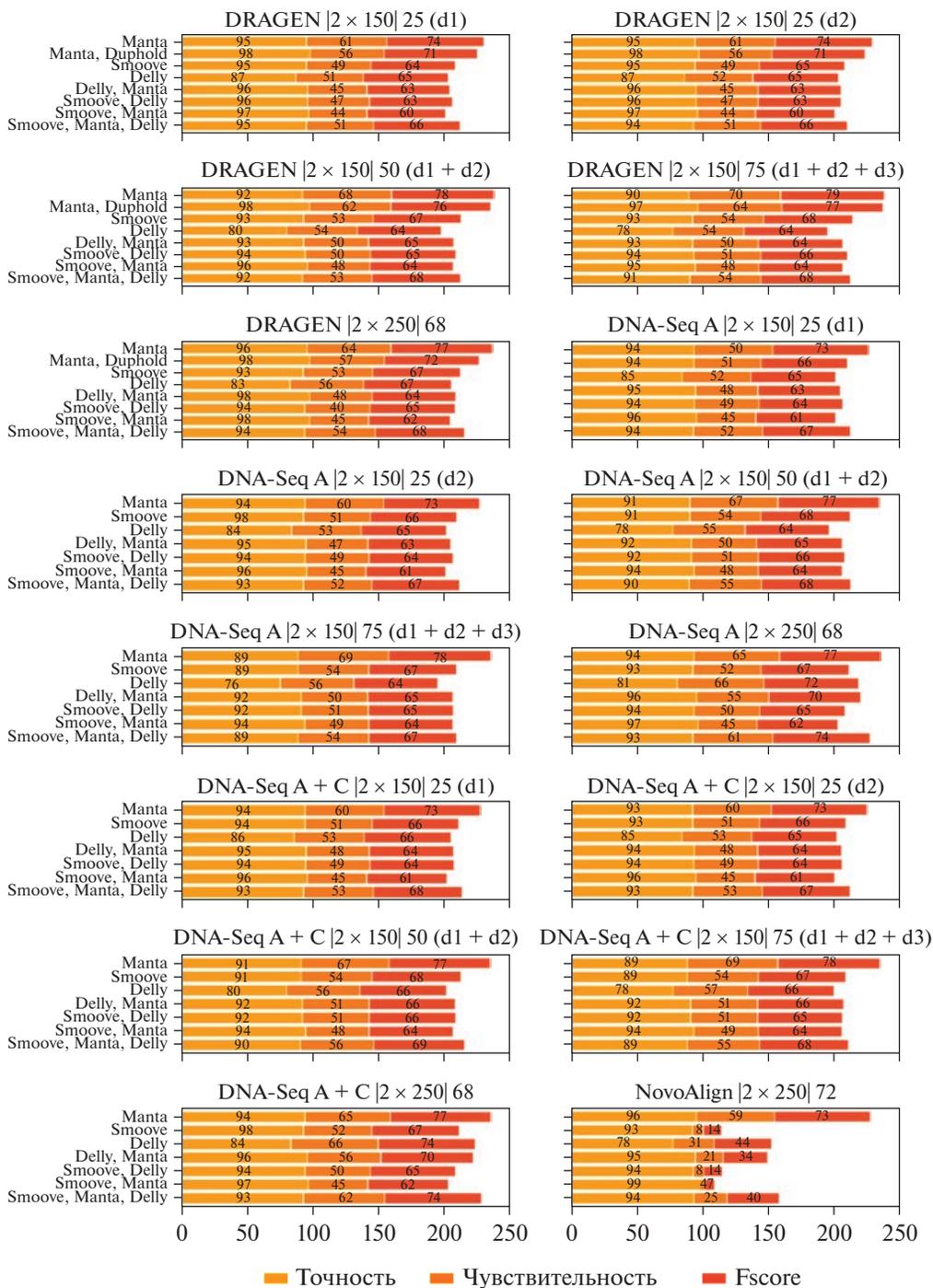


Рис. 2. Оценка качества работы разных биоинформатических инструментов по выявлению структурных вариантов типа DEL.

значительно, а при пайплайне DNA-Seq Alignment он увеличивается с 72.2 (в среднем) до 77% (точность увеличивается на 3–4%, а чувствительность уменьшается на 2–4% при всех трех пайплайнах). Фактически ни длина прочтений, ни глубина секвенирования, ни используемый для выравнивания пайплайн не дают существенного

вклада в изменение показателя Fscore, дающего сводную оценку качества детектирования структурных вариантов типа DEL.

Что касается самих биоинформатических инструментов для выявления структурных вариантов, то, судя по всем трем показателям: точность,

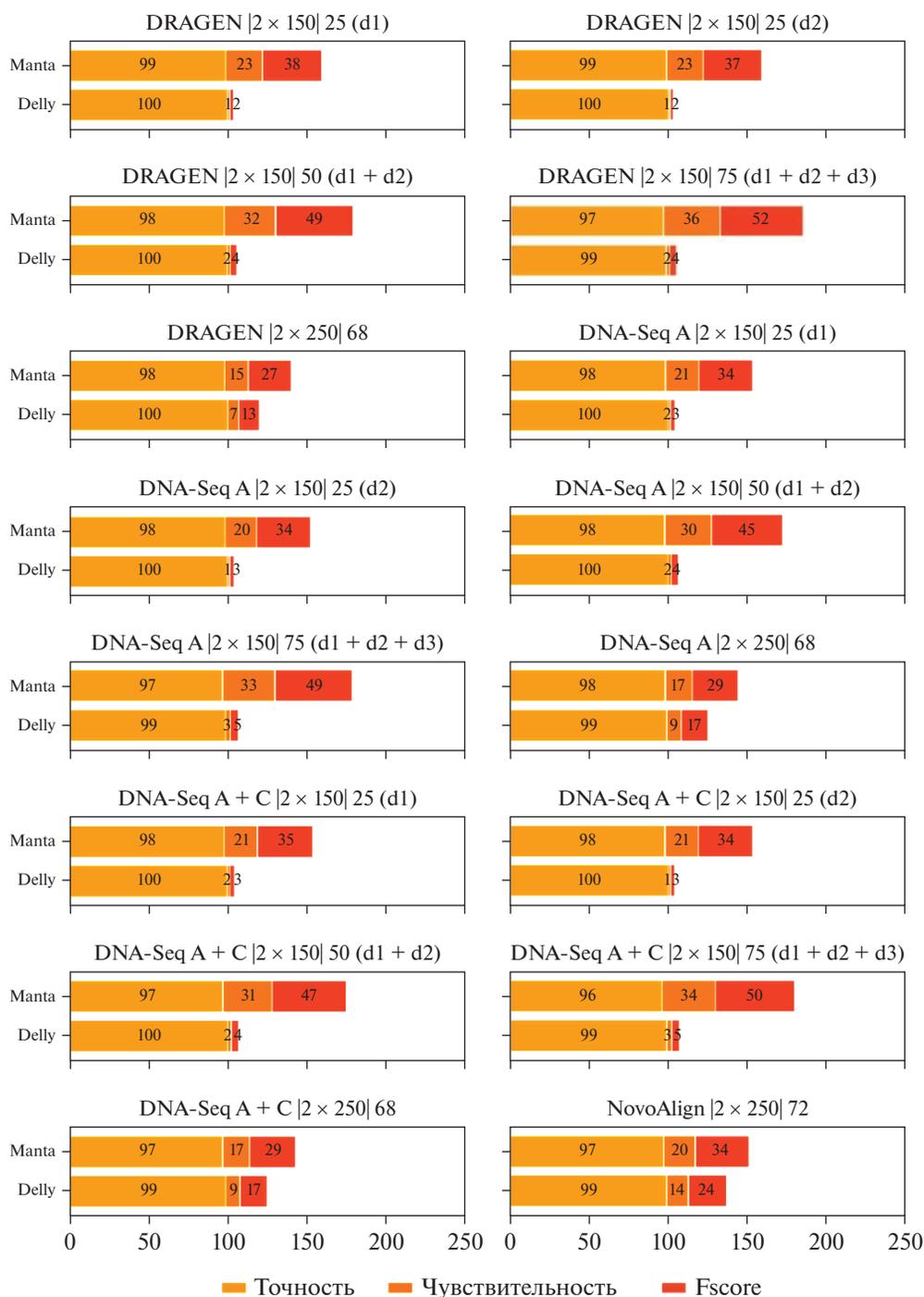


Рис. 3. Оценка качества работы разных биоинформатических инструментов по выявлению структурных вариантов типа INS.

чувствительность и Fscore, наиболее предпочтительным инструментом является Manta, который дает самые высокие показатели Fscore по сравнению с другими инструментами и их комбинациями при любом пайплайне выравнивания, включая NovoAlign, при любой глубине секвенирования и при любой из двух длин прочтений.

Сравнение результатов поиска делеций (см. рис. 2) и инсерций (см. рис. 3) демонстрирует, что поиск структурных вариантов типа делеция характеризуется более низкими показателями точности и более высокими показателями чувствительности по сравнению с поиском инсерций. При этом биоинформатический инструмент Manta детектирует

Таблица 3. Сравнение качества работы биоинформатических инструментов с помощью показателя Fscore на примере поиска структурных вариантов типа DEL

№	Пайплайн выравнивания	Длина прочтения (пн), глубина секвенирования	Показатель Fscore для биоинформатического инструмента							
			Manta	Manta с аннот.	Smooove	Delly	Manta + Smooove	Manta + Delly	Delly + Smooove	Manta + Delly + Smooove
1	DRAGEN	2 × 150, 25x (d1)	74	71	64	65	60	63	63	66
		2 × 150, 25x (d2)	74	71	65	65	60	63	63	66
		2 × 150, 50x (d1 + d2)	78	76	67	64	64	65	65	68
		2 × 150, 75x (d1 + d2 + d3)	79	77	68	64	64	64	66	68
		<i>Средняя для 2 × 150</i>	76.2	73.8	66.0	64.5	62.0	63.8	64.25	67.0
2	DNA-Seq Alignment	2 × 250, 68x	77	72	67	67	62	64	65	68
		2 × 150, 25x (d1)	73	–	66	65	61	63	64	67
		2 × 150, 25x (d2)	73	–	66	65	61	63	64	67
		2 × 150, 50x (d1 + d2)	77	–	68	64	64	65	66	68
		2 × 150, 75x (d1 + d2 + d3)	78	–	67	64	64	65	65	67
3	<i>Средняя для 2 × 150</i>		72.2	–	66.8	64.5	62.5	64.0	64.75	67.2
		2 × 250, 68x	77	–	67	72	62	70	65	74
		DNA-Seq Alignment + Co-Cleaning	73	–	66	66	61	64	64	68
		2 × 150, 25x (d2)	73	–	66	65	61	64	64	67
		2 × 150, 50x (d1 + d2)	77	–	68	66	64	66	66	69
4	NovoAlign	2 × 150, 75x (d1 + d2 + d3)	78	–	67	66	64	66	65	68
		<i>Средняя для 2 × 150</i>	75.2	–	66.8	65.8	62.5	65.0	64.75	68.0
		2 × 250, 68x	77	–	67	74	62	70	65	74
		2 × 250, 72x	73	–	14	44	7	34	14	40

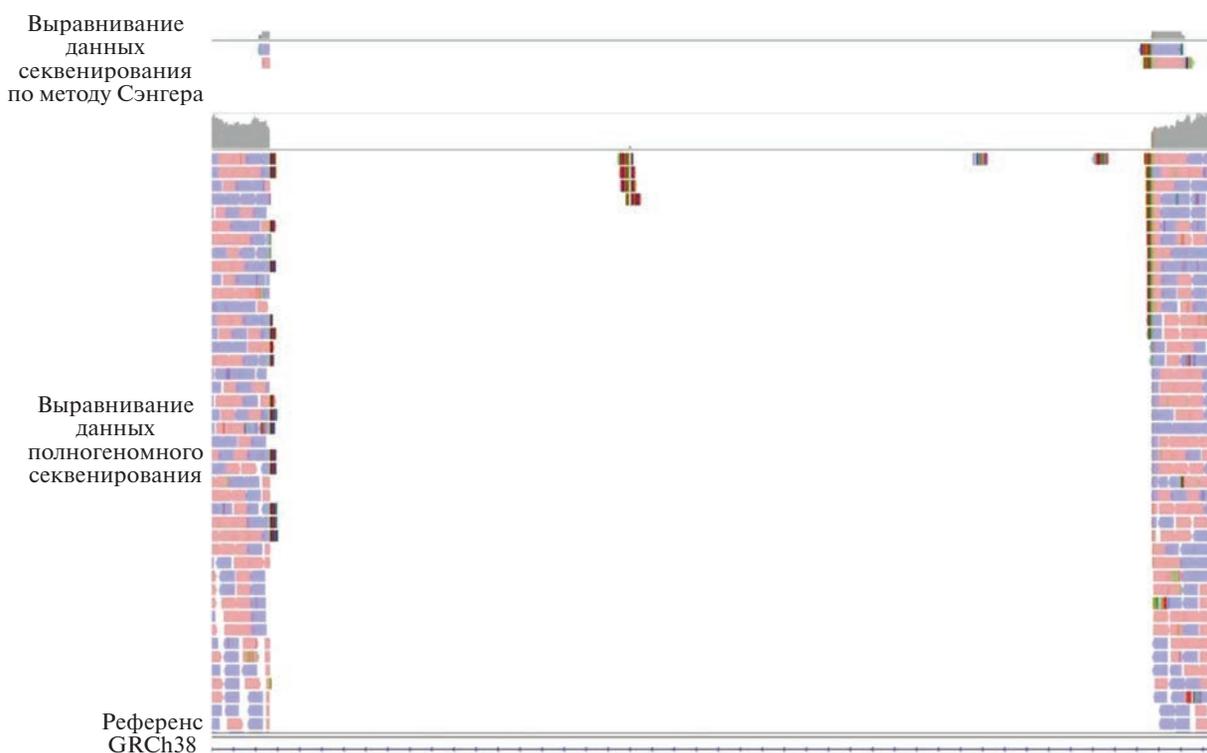


Рис. 4. Подтверждение наличия делеции chr5:12810907–12820414 (9507 пн). Здесь и далее розовая и синяя окраски прочтений обозначают выравнивание на обратную и прямую цепи ДНК соответственно.

структурные варианты лучше, чем инструменты Delly и Smoove, а также их комбинация. Как оказалось, файлы, полученные от совместной работы нескольких программ, фактически не увеличивают точности поиска и еще значительно снижают чувствительность. Фильтрация duphold, сочетаемая с инструментом Manta, несколько увеличивает точность нахождения делеций в выравниваниях, сделанных пайплайном DRAGEN, но уменьшает чувствительность и Fscore.

В ходе данного эксперимента точность поиска дупликаций и инверсий оценить не представилось возможным, поскольку формат записи таких типов структурных вариантов в эталонном файле отличается от формата записи в результирующих файлах анализируемых инструментов, однако проверка этих типов вариантов проводилась секвенированием по методу Сэнгера.

В Приложениях 2–4 показано, что точность работы биоинформатических инструментов для поиска структурных вариантов варьирует в зависимости от длины этих вариантов. Например, Manta лучше находит делеции длиной 200 и более пн, Smoove – от 200 до 10000 пн, Delly – длиной от 1000 до 10000 пн.

Валидация структурных вариантов

После полногеномного секвенирования был проведен поиск структурных вариантов в двух отобранных образцах. Найденные варианты проверялись с помощью секвенирования по методу Сэнгера. Для проверки были случайно отобраны делеции (размером от 50 до 45000 пн), инсерции, дупликации и инверсии (размером от 50 до 500 пн), которые были обнаружены инструментами Manta, Smoove и Delly, а также их комбинациями: Manta + Smoove, Manta + Delly, Smoove + Delly, Manta + Smoove + Delly.

Прочтения, полученные в результате секвенирования по методу Сэнгера, выравниваются по краям от делеции, без захвата участка собственно делеции.

После выравнивания прочтений (FASTQ) BAM-файл загружался в браузер IGV, с помощью которого визуально подтверждалось наличие или отсутствие структурного варианта в данной позиции. На рис. 4–6 показаны примеры подтвержденных делеции, инсерции и дупликации соответственно.

В табл. 4 представлена информация по количеству отобранных секвенированных и подтвержденных структурных вариантов.

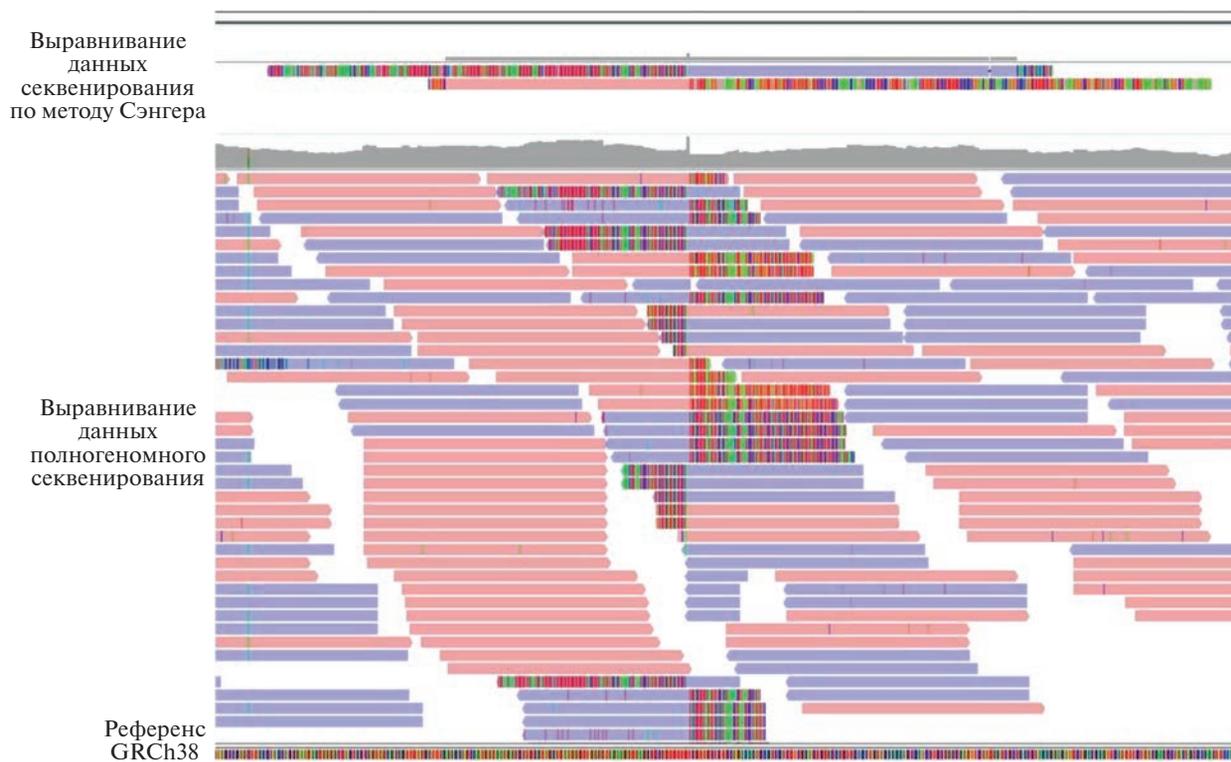


Рис. 5. Подтверждение наличия инсерции chr20:6665192 (67 пн). Здесь и далее яркие разноцветные участки по краям инсерции являются soft-clipped нуклеотидами, они не выравниваются на референсный геном, но остаются в выравнивании.

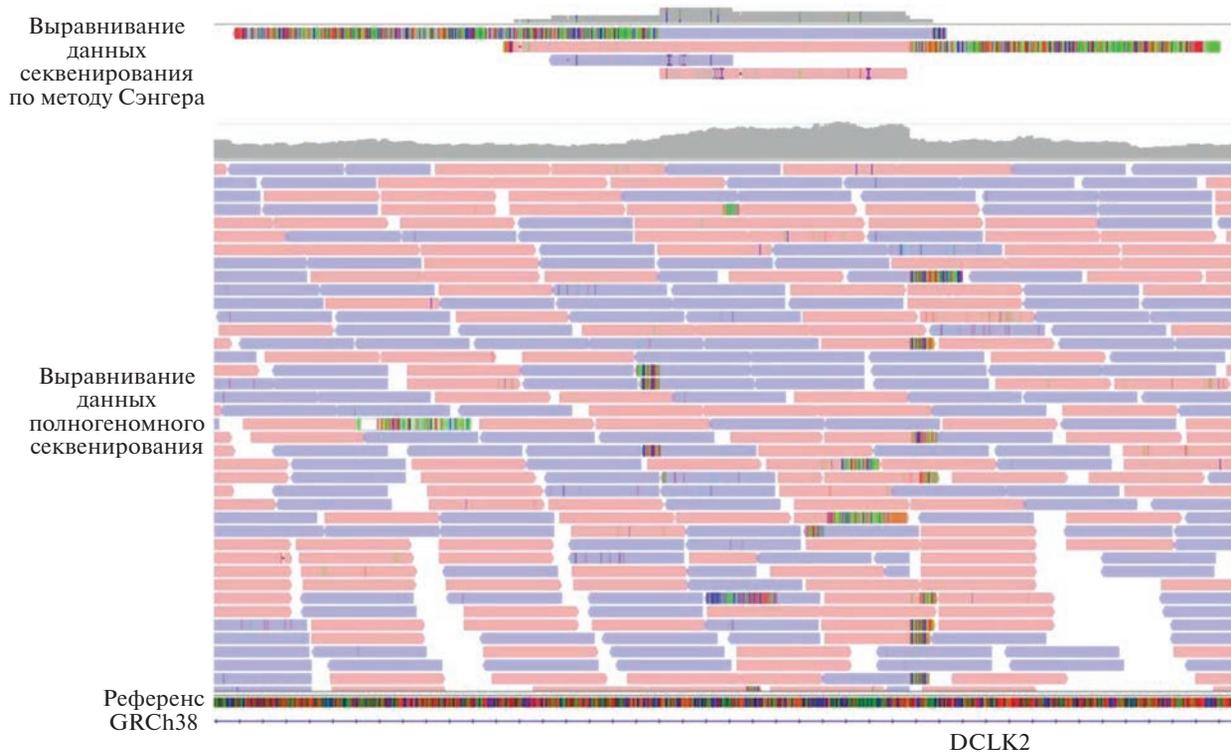


Рис. 6. Подтверждение наличия дубликации chr4:150090160–150090487 (327 пн).

Таблица 4. Сопоставление подтвержденных и секвенированных структурных вариантов

Ход эксперимента	Структурные варианты, ед.				Всего, ед.
	DEL	INS	DUP	INV	
Отобрано вариантов	60	28	23	9	120
из них секвенировано по Сэнгеру	17	12	12	2	43
из них подтвердились	11	7	6	0	24
<i>доля подтвердившихся от секвенированных</i>	<i>65%</i>	<i>58%</i>	<i>50%</i>	—	<i>56%</i>
не удалось секвенировать по методу Сэнгера	43	16	11	7	77
<i>доля несеквенированных от отобранных</i>	<i>72%</i>	<i>57%</i>	<i>48%</i>	<i>78%</i>	<i>64%</i>

В ходе данного эксперимента не все структурные варианты были секвенированы, поскольку для некоторых из них не произошло специфического связывания праймеров с матрицей, либо было невозможно подобрать праймеры из-за сложности региона (большое количество повторов во фланкирующем регионе), либо рассчитанная температура отжига праймеров не обеспечивала высокую специфичность ПЦР. Доля вариантов, которые не удалось секвенировать по методу Сэнгера, составила 78% для инверсий, 72% для делеций, 57% для инсерций и 48% для дупликаций.

Как следует из табл. 4, для 17 биоинформатически выявленных делеций, которые были проверены с помощью секвенирования по методу Сэнгера, подтвердились 11, т.е. 65%. Для инсерций аналогичная доля подтвердившихся вариантов среди всех секвенированных составила 58%. Для дупликаций эта доля составила 50%. В то же время из двух инверсий, выявленных биоинформатически и проверенных секвенированием по методу Сэнгера, не подтвердилось ни одной.

Таким образом, молекулярно-генетическая проверка показала, что делеции подтверждаются лучше, чем инсерции, дупликации и тем более инверсии. Варианты, которые были найдены при совместном использовании двух и трех биоинформатических инструментов, как оказалось, тоже подтверждаются далеко не всегда.

В табл. 5 представлены результирующие данные по вариантам, прошедшим молекулярно-генетическую проверку.

Как видно из табл. 5, из отобранных в рамках данной работы двух инверсий не подтвердилась ни одна. Следовательно, инверсии, найденные с помощью биоинформатических инструментов Manta и Smoove, требуют дополнительной проверки.

Табл. 5 и Приложение 4 также свидетельствуют о том, что структурные варианты типа DEL, INS и DUP целесообразно выявлять с помощью инструмента Manta v1.6.0 – использование двух или трех инструментов одновременно не увеличило показателя Fscore.

ОБСУЖДЕНИЕ

С развитием технологий секвенирования, со снижением стоимости и ресурсоемкости последующего анализа растет количество данных, по которым можно изучать ассоциации генетических вариантов, в том числе и структурных, с развитием целого ряда заболеваний. В настоящее время создано более 30 инструментов для поиска структурных вариантов и со временем это число будет только расти.

К сожалению, среди такого большого числа инструментов нет универсального, который бы имел высокие показатели точности и чувствительности для всех типов структурных вариантов, подходил бы для массовых вычислений, результирующие файлы которого были бы совместимы с другими инструментами для поиска структурных вариантов и инструментами объединения и создания файла с вариантами, найденными комбинацией двух и более инструментов. Поэтому в зависимости от цели исследования используемые инструменты и их режимы запуска могут различаться между собой.

Для настоящего исследования мы выбрали три биоинформатических инструмента (Manta, Smoove, Delly), которые подходят для массовой обработки данных. Вторым важным фактором, повлиявшим на выбор именно этих трех инструментов, стало то, что их результирующие файлы имеют аналогичный формат записи информации о структурных вариантах, включая аннотации. Соответ-

Таблица 5. Результаты проверки структурных вариантов с помощью секвенирования по методу Сэнгера

№	Тип SV	Длина SV, пн	Инструмент поиска структурных вариантов		
			Manta v1.6.0	Smoove v0.2.8	Delly v0.9.1
Делеции					
1	DEL	3356	chr1:115686865	chr1:115686865	chr1:115686865
2	DEL	1482	н.о.	chr1:17349677	chr1:17349677
3	DEL	319	chr1:240453690	chr1:240453707	chr1:240453690
4	DEL	68	chr2:1208249	chr2:1208263	chr2:1208249
5	DEL	414	chr2:1354296	chr2:1354296	н.о.
6	DEL	43366	chr3:130044537–130087902	chr3:130044539–130087902	chr3:130044537–130087903
7	DEL	3000	chr4:185520481	chr4:185520483	chr4:185520481
8	DEL	9507	chr5:12810905–12820411	chr5:12810907–12820411	chr5:12810905–12820412
9	DEL	474	chr11:677345	chr11:677345	chr11:677345
10	DEL	62	chr11:21767543	chr11:21767543	chr11:21767543
11	DEL	11030	chr12:12378536–12389566	chr12:12378535–12389566	chr12:12378537–12389567
12	DEL	316	chr12:1754845	chr12:1754845	chr12:1754845
13	DEL	877	chr17:17327060–17327936	chr17:17327067–17327919	chr17:17327060–17327937
14	DEL	316	chr18:79739850	chr18:79739850	н.о.
15	DEL	3133	н.о.	chr19:605301	н.о.
16	DEL	73	chr20:64212162	н.о.	chr20:64212162
17	DEL	2482	chrX:77865888–77868369	chrX:77865918–77868355	chrX:77865920–77868356
Результат*			11-2-2-0–0.85-1.00-0.92	10-0-4-1–0.71-0.91-0.80	9-1-3-2–0.75-0.82-0.78
Manta + Smoove			10-2-2-1–0.83-0.91-0.90		
Smoove + Delly			8-1-3-3–0.72-0.73-0.73		
Manta + Delly			9-2-2-2–0.82-0.82-0.82		
Manta + Delly + Smoove			8-2-2-3–0.80-0.73-0.76		
Инсерции					
18	INS	317	chr1:4228390	н.о.	н.о.
19	INS	67	chr4:136314242	н.о.	chr4:136314242
20	INS	57	chr5:134232927	н.о.	chr5:134232927
21	INS	135	chr8:1912540	н.о.	н.о.
22	INS	54	chr8:102867903	н.о.	chr8:102867903
23	INS	316	chr10:9432883	н.о.	н.о.
24	INS	56	chr13:92723353	н.о.	chr13:92723353
25	INS	80	chr15:53110125	н.о.	н.о.

Таблица 5. Окончание

№	Тип SV	Длина SV, пн	Инструмент поиска структурных вариантов		
			Manta v1.6.0	Smoove v0.2.8	Delly v0.9.1
26	INS	52	chr16:839939	н.о.	chr16:839939
27	INS	69	chr19:49867632	н.о.	н.о.
28	INS	66	chr20:6665192	н.о.	chr20:6665192
29	INS	60	chrX:145338072	н.о.	chrX:145338072
Результат*			7-0-3-0-0.70-1.00-0.82	—	4-2-1-3-0.80-0.57-0.67
Manta + Delly			4-2-1-3-0.80-0.57-0.67		
Дубликации					
30	DUP	234	н.о.	chr2:239723885	н.о.
31	DUP	372	chr3:990665	chr3:990665	chr3:990665
32	DUP	329	chr4:150090160	chr4:150090160	chr4:150090160
33	DUP	189	chr6:1051543	chr6:1051543	н.о.
34	DUP	189	chr7:40915751	chr7:40915751	н.о.
35	DUP	247	chr9:2814058	chr9:2814058	chr9:2814058
36	DUP	205	chr14:23900201	chr14:23900201	chr14:23900201
37	DUP	196	chr14:82276270	н.о.	chr14:82276270
38	DUP	162	chr15:101390179	chr15:101390179	н.о.
39	DUP	54	chr16:919548	н.о.	н.о.
40	DUP	196	н.о.	chr18:9809064	chr18:9809064
41	DUP	144	chr20:10323922	н.о.	н.о.
Результат*			5-0-4-1-0.56-0.83-0.67	4-1-3-2-0.57-0.67-0.62	3-3-1-3-0.75-0.50-0.60
Manta + Smoove			3-1-3-3-0.50-0.50-0.50		
Smoove + Delly			2-3-1-4-0.67-0.33-0.55		
Manta + Delly			3-3-1-3-0.75-0.50-0.60		
Manta + Delly + Smoove			2-3-1-4-0.67-0.33-0.44		
Инверсии					
42	INV	75	chr10:12825172	chr10:12825172	н.о.
43	INV	11	chr11:104240549	chr11:104240549	н.о.
Результат*			—	—	—

Примечание. Красный цвет – вариант не подтвердился, зеленый – вариант подтвердился, оранжевый – последовательности выравниваются таким образом, что нельзя достоверно сказать подтверждается вариант или нет (например, одна из целей не выравнивается). н.о. – не обнаружено.

* Числа даны в следующем порядке: TP (истинно положительные)-TN (истинно отрицательные)-FP (ложно положительные)-FN (ложноотрицательные)—точность-чувствительность-Fscore; варианты с сомнительными результатами проверки не учитывались.

ственно файлы от этих трех инструментов удобно сопоставлять и анализировать совместно в разных комбинациях.

Нами проведена оценка качества работы инструментов Manta, Smoove и Delly посредством проверки найденных ими вариантов с помощью секвенирования по методу Сэнгера. Рассматривались структурные варианты, найденные отдельно каждым инструментом, а также совместно двумя и тремя инструментами. Мы оценили влияние длины входных прочтений и покрытия в целом, а также влияние разных алгоритмов этапа выравнивания и показали, что с увеличением покрытия образца растет достоверность идентификации структурных вариантов биоинформатическими инструментами. В то же время длина коротких прочтений, при равном покрытии, почти не влияет на качество работы инструментов Manta, Smoove и Delly. Ограничением данного исследования стал тот факт, что влияние длины фрагментов и insert size на точность, чувствительность, Fscore нами не изучалось.

Также нам удалось показать, что выбор алгоритма выравнивания оказывает влияние на качество работы инструментов. Например, для инструмента Smoove выравнивание с помощью NovoAlign уменьшает число найденных вариантов почти в 7 раз по сравнению с другими алгоритмами выравнивания.

В результате настоящего исследования мы сделали вывод о том, что каждый из трех изученных инструментов эффективнее работает в определенном диапазоне размеров структурных вариантов. Исходя из этого, мы предполагаем, что изучение структурных вариантов, имеющих разные длины, целесообразно проводить используя совместно два и более инструмента.

Проверка данных, полученных в ходе нашего исследования с помощью секвенирования по методу Сэнгера, показала, что в действительности точность работы всех трех биоинформатических инструментов оказалась намного ниже, чем предполагалось. Фактически молекулярно-генетический метод не подтвердил наличие многих структурных вариантов, детектированных инструментами Manta, Smoove и Delly. При этом молекулярно-генетическая проверка показала, что в среднем, без учета спорных вариантов, делеции и инсерции биоинформатически определяются надежнее, чем инверсии и дубликации.

Таким образом, наше исследование позволило дать следующие рекомендации:

– для анализа данных WGS, популяционных исследований, поиска ассоциаций с фенотипом, расчета частоты встречаемости вариантов в выборке целесообразно совместно использовать хотя бы два биоинформатических инструмента для поиска структурных вариантов;

– при клинической интерпретации генетических данных наличие структурного варианта необходимо подтверждать с помощью альтернативных молекулярно-генетических методов.

Исследование не имело спонсорской поддержки.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национально-го комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников было получено информированное добровольное согласие.

Авторы заявляют, что у них нет конфликта интересов.

ПРИЛОЖЕНИЕ 1

Некоторые команды запуска инструментов, описанных в Материалах и методах:

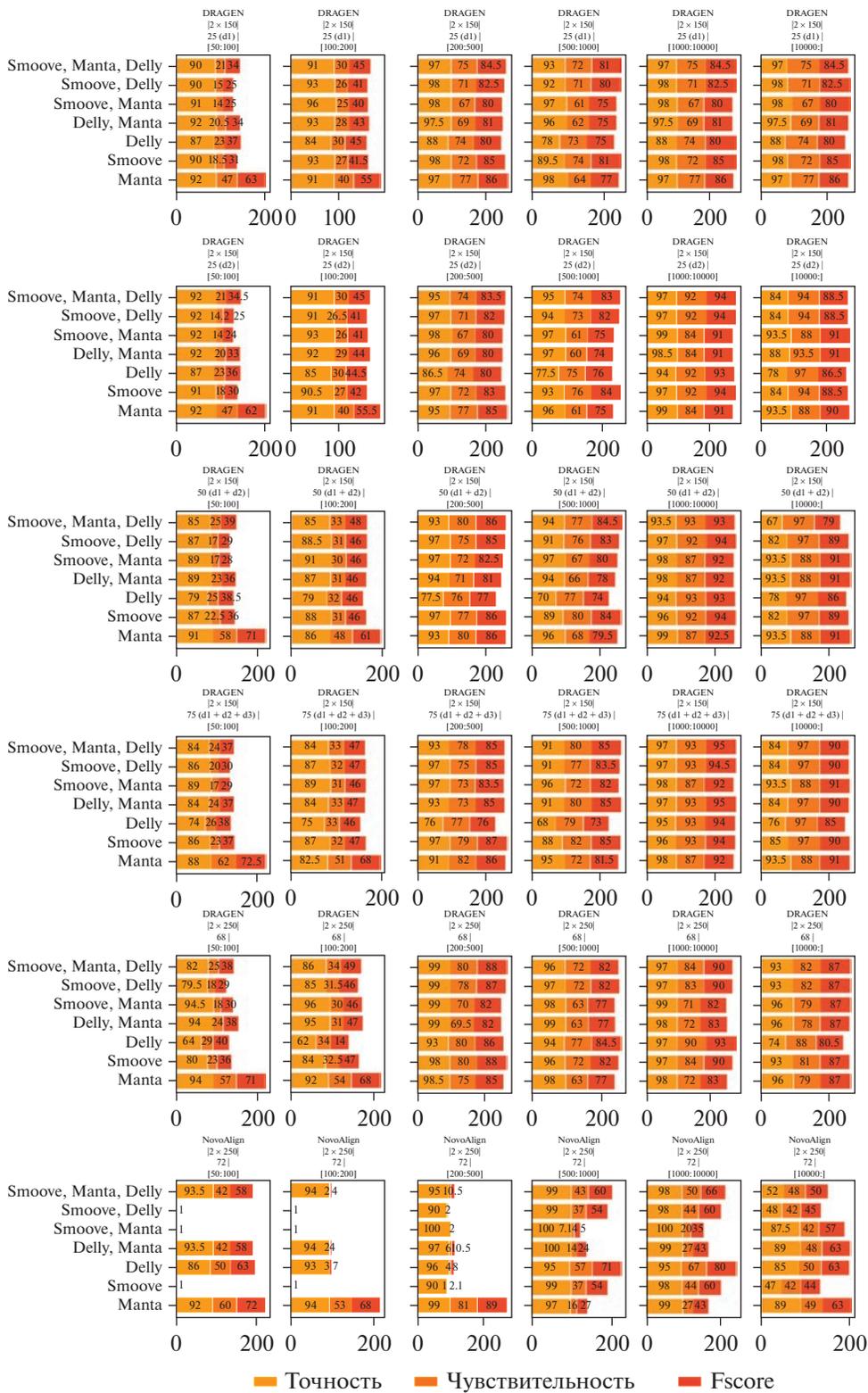
1) команда запуска инструмента bcftools для фильтрации структурных вариантов: `bcftools view -i 'FILTER="PASS" && INFO/SVTYPE != "BND";`

2) команда запуска инструмента bcftools для фильтрации структурных вариантов, аннотированных duphold: `bcftools view -i '(SVTYPE = "DEL" & FMT/DHFFC[0] < 0.7) | (SVTYPE = "DUP" & FMT/DHBFC[0] > 1.3)' $svvcf;`

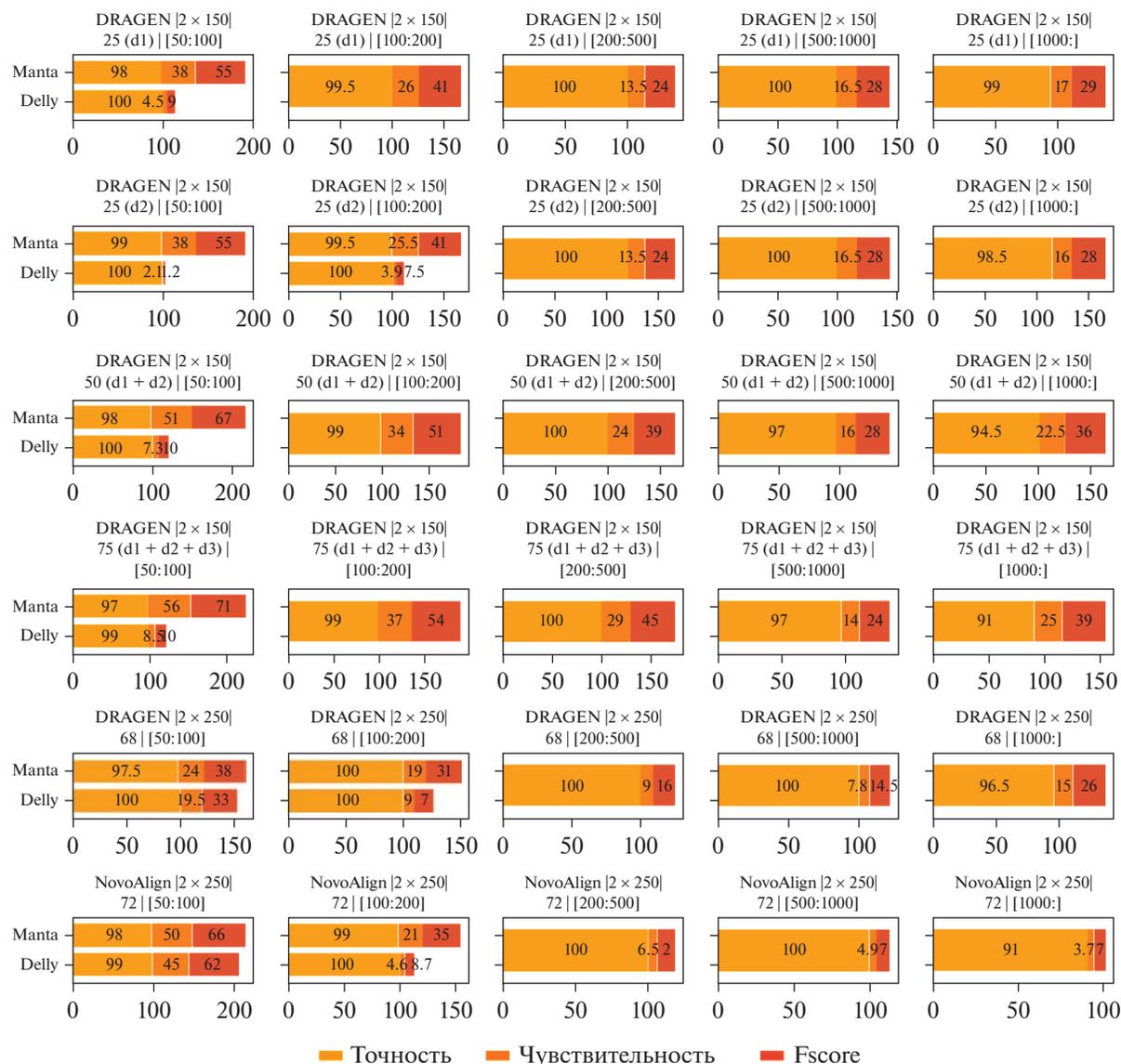
3) команда запуска инструмента SURVIVOR:

```
SURVIVOR merge <list_of_VCF_files_for_merging> 1000 2 1 1 0 50 <merged.vcf>;
```

4) команда запуска инструмента Witty.er: `wittyer -i <vcf_to_validate.vcf> -t HG002_SVs_Tier1_v0.6.vcf -em SimpleCounting -configFile config.json -includeBed HG002_SVs_Tier1_v0.6.bed -o <dir_with_comarisson_results_from_wittyer>.`



Результаты оценки чувствительности и точности работы инструментов для поиска структурных вариантов для файлов выравнивания, полученных с помощью DRAGEN на прочтениях разных длин и покрытия для разных диапазонов длин вариантов, для структурных вариантов типа делеция.



Результаты оценки чувствительности и точности работы инструментов для поиска структурных вариантов для файлов выравнивания, полученных с помощью DRAGEN на прочтениях разных длин и покрытия для разных диапазонов длин вариантов, для структурных вариантов типа инсерция.

ПРИЛОЖЕНИЕ 4

Сравнение качества работы биоинформатических инструментов с помощью показателя Fscore на примере поиска структурных вариантов типа DEL на разных диапазонах длин

№	Диапазон длин, пн	Пайплайн выравнивания	Длина прочтения (пн), глубина секвенирования	Показатель Fscore для биоинформатического инструмента						
				Manta	Smoove	Delly	Manta + Smoove	Manta + Delly	Delly + Smoove	Manta + Smoove + Delly
1	50 : 100	DRAGEN	2 × 150, 25x (d1)	63	31	37	25	34	25	34
			2 × 150, 25x (d2)	62	30	36	24	33	25	34.5
			2 × 150, 50x (d1 + d2)	71	36	38.5	28	36	29	39
			2 × 150, 75x (d1 + d2 + d3)	72.5	37	38	29	37	30	37
			<i>Средняя для 2 × 150</i>	<i>67.1</i>	<i>33.5</i>	<i>37.4</i>	<i>26.5</i>	<i>35.0</i>	<i>27.3</i>	<i>36.1</i>
			2 × 250, 68x	71	36	40	30	38	29	38
2	100 : 200	DRAGEN	2 × 250, 72x	72	0	63	1	58	0	58
			2 × 150, 25x (d1)	55	41.5	45	40	43	41	45
			2 × 150, 25x (d2)	55.5	42	44.5	41	44	41	45
			2 × 150, 50x (d1 + d2)	61	46	46	45	46	46	48
			2 × 150, 75x (d1 + d2 + d3)	63	47	46	46	47	47	47
			<i>Средняя для 2 × 150</i>	<i>58.6</i>	<i>44.1</i>	<i>45.4</i>	<i>43.0</i>	<i>45.0</i>	<i>43.8</i>	<i>46.3</i>
3	200 : 500	DRAGEN	2 × 250, 68x	68	47	44	46	47	46	49
			2 × 250, 72x	68	0	3.7	1	4	0	4
			2 × 150, 25x (d1)	86	83	80	80	81	82.5	84.5
			2 × 150, 25x (d2)	85	83	80	80	80	82	83.5
			2 × 150, 50x (d1 + d2)	86	86	77	82.5	81	85	86
			2 × 150, 75x (d1 + d2 + d3)	86	87	76	83.5	85	85	85
3	500 : 1000	DRAGEN	<i>Средняя для 2 × 150</i>	<i>85.8</i>	<i>84.8</i>	<i>78.3</i>	<i>81.5</i>	<i>81.8</i>	<i>83.6</i>	<i>84.8</i>
			2 × 250, 68x	85	88	86	82	82	87	88
			2 × 250, 72x	89	2.1	9	2	10.5	2	10.5
			2 × 150, 25x (d1)	77	81	75	75	75	80	81
			2 × 150, 25x (d2)	75	84	76	75	74	82	83
			2 × 150, 50x (d1 + d2)	79.5	84	74	80	78	83	84.5
3	1000 : 10000	DRAGEN	2 × 150, 75x (d1 + d2 + d3)	81.5	85	73	82	85	83.5	85
			<i>Средняя для 2 × 150</i>	<i>78.3</i>	<i>83.5</i>	<i>74.5</i>	<i>78.0</i>	<i>78.0</i>	<i>82.1</i>	<i>83.4</i>
			2 × 250, 68x	77	82	84.5	77	77	82	82
			2 × 250, 72x	27	54	71	14	24	54	60
			2 × 150, 25x (d1)	91	94	94	91	91	94	94
			2 × 150, 25x (d2)	91	94	93	91	91	94	94
3	10000:	DRAGEN	2 × 150, 50x (d1 + d2)	92.5	94	93	92	92	94	93
			2 × 150, 75x (d1 + d2 + d3)	92	94	94	92	95	94.5	95
			<i>Средняя для 2 × 150</i>	<i>91.6</i>	<i>94.0</i>	<i>93.5</i>	<i>91.5</i>	<i>92.3</i>	<i>94.1</i>	<i>94.0</i>
			2 × 250, 68x	83	90	93	82	83	90	90
			2 × 250, 72x	43	60	80	35	43	60	66
			2 × 150, 25x (d1)	89	90	85	90	89	90	89
4		NovoAlign	2 × 150, 25x (d2)	90	88.5	86.5	91	91	88.5	88.5
			2 × 150, 50x (d1 + d2)	91	89	86	91	91	89	79
			2 × 150, 75x (d1 + d2 + d3)	91	90	85	91	90	90	90
			<i>Средняя для 2 × 150</i>	<i>90.3</i>	<i>89.4</i>	<i>85.6</i>	<i>90.8</i>	<i>90.3</i>	<i>89.4</i>	<i>86.6</i>
			2 × 250, 68x	87	87	80.5	87	87	87	87
			2 × 250, 72x	63	44	63	57	63	45	50

СПИСОК ЛИТЕРАТУРЫ

1. Pang A.W., MacDonald J.R., Pinto D. et al. Towards a comprehensive structural variation map of an individual human genome // *Genome Biol.* 2010. V. 11. № 5. P. R52. <https://doi.org/10.1186/gb-2010-11-5-r52>
2. The International HapMap Consortium. The international hapmap project // *Nature.* 2003. P. 789–796. <https://doi.org/10.1038/nature02168>
3. Sudmant P.H., Rausch T., Gardner E.J. et al. An integrated map of structural variation in 2,504 human genomes: 7571 // *Nature.* 2015. V. 526. № 7571. P. 75–81. <https://doi.org/10.1038/nature15394>
4. Pös O., Radvanszky J., Buglyo G. et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects // *Biomed. J.* 2021. V. 44. № 5. P. 548–559. <https://doi.org/10.1016/j.bj.2021.02.003>
5. Alkan C., Coe B.P., Eichler E.E. Genome structural variation discovery and genotyping // *Nat. Rev. Genet.* 2011. V. 12. № 5. P. 363–367. <https://doi.org/10.1038/nrg2958>
6. Mahmoud M., Gobet N., Cruz-Davalos D.I. et al. Structural variant calling: the long and the short of it // *Genome Biol.* 2019. V. 20. № 1. P. 246. <https://doi.org/10.1186/s13059-019-1828-7>
7. Carvalho C.M., Lupski J.R. Mechanisms underlying structural variant formation in genomic disorders // *Nat. Rev. Genet.* 2016. V. 17. № 4. P. 224–238. <https://doi.org/10.1038/nrg.2015.25>
8. Sedlazeck F.J., Lee H., Darby C.A. et al. Piercing the dark matter: bioinformatics of long-range sequencing and mapping // *Nat. Rev. Genet.* 2018. V. 19. № 6. P. 329–346. <https://doi.org/10.1038/s41576-018-0003-4>
9. Collins R.L., Brand H., Karczewski K.J. et al. A structural variation reference for medical and population genetics // *Nature.* 2021. V. 581. P. 444–451. <https://doi.org/10.1038/s41586-020-2287-8>
10. Weischenfeldt J., Symmons O., Spitz F. et al. Phenotypic impact of genomic structural variation: Insights from and for human disease // *Nat. Rev. Genet.* 2013. V. 14. № 2. P. 125–138. <https://doi.org/10.1038/nrg3373>
11. Stankiewicz P., Lupski J. Structural variation in the human genome and its role in disease // *Annu. Rev. Med.* 2010. V. 61. P. 437–455. <https://doi.org/10.1146/annurev-med-100708-204735>
12. Schüle B., McFarland K.N., Lee K. et al. Parkinson's disease associated with pure ATXN10 repeat expansion // *NPJ Parkinson's Disease.* 2017. V. 3. P. 27. <https://doi.org/10.1038/s41531-017-0029-x>
13. Yeh Y.C., Ho H.L., Wu Y.C. et al. AKT1 internal tandem duplications and point mutations are the genetic hallmarks of sclerosing pneumocytoma // *Mod. Pathol.* 2020. V. 33. № 3. P. 391–403. <https://doi.org/10.1038/s41379-019-0357-y>
14. Schütte J., Reusch J., Khandanpour C. et al. Structural variants as a basis for targeted therapies in hematological malignancies // *Front. Oncol.* 2019. V. 9. P. 839. <https://doi.org/10.3389/fonc.2019.00839>
15. Ewing A., Meynert A., Churchman M. et al. Structural variants at the BRCA1/2 loci are a common source of homologous repair deficiency in high-grade serous ovarian carcinoma // *Clin. Cancer Res.* V. 27. № 11. P. 3201–3214. <https://doi.org/10.1158/1078-0432.ccr-20-4068>
16. Malhotra D., Sebat J. CNVs: Harbingers of a rare variant revolution in psychiatric genetics // *Cell.* 2012. V. 148. № 6. P. 1223–1241. <https://doi.org/10.1016/j.cell.2012.02.039>
17. Huse K., Taudien S., Groth M. et al. Genetic variants of the copy number polymorphic β -defensin locus are associated with sporadic prostate cancer // *Tumor Biol.* 2008. V. 29. № 2. P. 83–92. <https://doi.org/10.1159/000135688>
18. Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls // *Nature.* 2010. V. 464. P. 713–720. <https://doi.org/10.1038/nature08979>
19. Fanciulli M., Norsworthy P.J., Petretto E. et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity // *Nat. Genet.* 2007. V. 39. № 6. P. 721–723. <https://doi.org/10.1038/ng2046>
20. Fellermann K., Stange D.E., Schaeffeler E. et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to crohn disease of the colon // *Am. J. Hum. Genet.* 2006. V. 79. № 3. P. 439–448. <https://doi.org/10.1086/505915>
21. Hollox E.J., Huffmeier U., Zeeuwen P.L. et al. Psoriasis is associated with increased β -defensin genomic copy number // *Nat. Genet.* 2008. V. 40. № 1. P. 23–25. <https://doi.org/10.1038/ng.2007.48>
22. Zook J.M., Hansen N.F., Olson N.D. et al. A robust benchmark for detection of germline large deletions and insertions // *Nat. Biotechnol.* 2020. V. 38. P. 1347–1355. <https://doi.org/10.1038/s41587-020-0538-8>
23. Ye J., Coulouris G., Zaretskaya I. et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction // *BMC Bioinformatics.* 2012. V. 13. № 1. P. 134. <https://doi.org/10.1186/1471-2105-13-134>
24. Sequencing analysis viewer [Electronic resource] // Sequencing analysis viewer support. URL: https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.html (accessed: 15.09.2022).
25. bcl2fastq [Electronic resource] // bcl2fastq. URL: https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software/downloads.html (accessed: 13.05.2022).
26. FastQC [Electronic resource] // Babraham bioinformatics FastQC a quality control tool for high throughput sequence data. URL: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed: 13.05.2022).
27. Illumina DRAGEN Bio-IT Platform [Electronic resource] // Illumina DRAGEN Bio-IT platform | variant calling & secondary genomic analysis software tool. URL: <https://www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html> (accessed: 13.05.2022).
28. Schneider V.A., Graves-Lindsay T., Howe K. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference

- assembly // *Genome Res.* 2017. V. 27. № 5. P. 849–864.
<https://doi.org/10.1101/gr.213611.116>
29. *Chen X., Schulz-Trieglaff O., Shaw R. et al.* Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications // *Bioinformatics.* 2016. V. 32. № 8. P. 1220–1222.
<https://doi.org/10.1093/bioinformatics/btv710>
 30. Smoove [Electronic resource] // *Brent S.* Smoove: structural-variant calling and genotyping with existing tools. URL: <https://github.com/brentp/smoove> (accessed: 13.07.2022).
 31. *Rausch T., Zichner T., Schlattl A. et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis // *Bioinformatics.* 2012. V. 28. № 18. P. i333–i339.
<https://doi.org/10.1093/bioinformatics/bts378>
 32. NCBI [Electronic resource] // GRCh37 hg19 genome assembly NCBI. URL: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/ (accessed: 15.09.2022).
 33. *Zhao H., Sun Z., Wang J. et al.* CrossMap: A versatile tool for coordinate conversion between genome assemblies // *Bioinformatics.* 2014. V. 30. № 7. P. 1006–1007.
<https://doi.org/10.1093/bioinformatics/btt730>
 34. GDC [Electronic resource] // GDC viewer docs. URL: https://docs.gdc.cancer.gov/Data_Dictionary/viewer/#?view=table-definition-view&id=alignment_cleaning_workflow (accessed: 13.07.2022).
 35. *Li H., Durbin R.* Fast and accurate long-read alignment with Burrows–Wheeler transform // *Bioinformatics.* 2010. V. 26. № 5. P. 589–595.
<https://doi.org/10.1093/bioinformatics/btp698>
 36. *Danecek P., Bonfield J.K., Liddle J. et al.* Twelve years of samtools and bcftools // *GigaScience.* 2021. V. 10. № 2.
<https://doi.org/10.1093/gigascience/giab008>
 37. *Pedersen B.S., Quinlan A.R.* Duphold: scalable, depth-based annotation and curation of high-confidence structural variant calls // *GigaScience.* 2019. V. 8. № 4.
<https://doi.org/10.1093/gigascience/giz040>
 38. *Jeffares D.C., Jolly C., Hoti M. et al.* Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast // *Nat. Commun.* 2017. V. 8. P. 14061.
<https://doi.org/10.1038/ncomms14061>
 39. Witty.er [Electronic resource] // *Wan Y., Ho K.* Witty.er. URL: <https://github.com/Illumina/witty.er> (accessed: 13.07.2022).
 40. Unipro UGENE [Electronic resource] // Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* Oxford academic. URL: <https://academic.oup.com/bioinformatics/article/28/8/1166/195474?login=false> (accessed: 13.07.2022).
 41. *Robinson J.T., Thorvaldsdottir H., Winckler W. et al.* Integrative genomics viewer // *Nat. Biotechnol.* 2011. V. 29. № 1. P. 24–26.
<https://doi.org/10.1038/nbt.1754>

Comparative Analysis of Structural Variant Callers on the Short-Read Whole-Genome Sequencing Data

**A. A. Mkrtchian^{a,*}, K. S. Grammatikati^a, P. G. Kazakova^a, S. I. Mitrofanov^a, P. U. Zemsky^a,
A. A. Ivashechkin^a, M. N. Pilipenko^a, D. V. Svetlichny^a, A. P. Sergeev^a, E. A. Snigir^a,
L. V. Frolova^a, T. A. Shpakova^a, V. S. Yudin^a, A. A. Keskinov^a, S. M. Yudin^a, and V. I. Skvortsova^b**

^a*Federal State Budgetary Institution “Centre for Strategic Planning and Management of Biomedical Health Risks”
of the Federal Medical Biological Agency, Moscow, 119121 Russia*

^b*The Federal Medical Biological Agency of Russia, Moscow, 123182 Russia*

*e-mail: AMkrtchyan@cspfmba.ru

In this study three structural variant callers (Manta, Smoove, Delly) were analysed on the whole-genome sequencing data using four different alignment algorithms: DRAGEN, GDC DNA-Seq Alignment Workflow, GDC DNA-Seq Alignment Workflow + GDC DNA-Seq Co-Cleaning Workflow, NovoAlign, different lengths of raw reads: 2×150 bp and 2×250 bp, different mean genome coverage values. Results were compared to etalon results of GIAB team. Structural variants validation was hold also with Sanger sequencing. Structural variants deletions and insertions as it turned out were best determined with Manta tool. We’ve got 89–96% of accuracy and 59–70% of sensitivity for analysed deletions, and also 96–99% of accuracy and 15–36% of sensitivity for insertions. Smoove and Delly showed less accurate and sensitive results (Smoove: 91–95% of accuracy and 8–54% of sensitivity for deletions, Delly: 78–87% of accuracy and 31–66% of sensitivity for deletions, 99–100% of accuracy and 1–13% of sensitivity for insertions). Simultaneous using of two or even three structural variant callers didn’t give a rise of accuracy and sensitivity for deletions. Analysis showed that accuracy and sensitivity of structural variant callers rise with the rising of mean genome coverage value, increasing of reads length from 150 to 250 bp influence in to varying degrees on the accuracy and sensitivity of individual tools. Another inference of this study was that accuracy of structural variants callers vary depends on structural variants size range. For example, Manta finds better deletions in the range from 200 and more bp, Delly – from 1000 to 10000 bp, Smoove – from 200 to 10000 bp.

Keywords: structural variants, WGS, short-read sequencing, alignment, GIAB.

О ГЕНЕТИЧЕСКОЙ ДИВЕРГЕНЦИИ МИГРАЦИОННО-СВЯЗАННЫХ ПОПУЛЯЦИЙ: СОВРЕМЕННОЕ МОДЕЛИРОВАНИЕ ПО РЕЗУЛЬТАТАМ ЭКСПЕРИМЕНТОВ Ю.П. АЛТУХОВА С СОАВТОРАМИ

© 2023 г. О. Л. Жданова¹, *, Е. Я. Фрисман², **

¹Институт автоматизации и процессов управления Дальневосточного отделения
Российской академии наук, Владивосток, 690041 Россия

²Институт комплексного анализа региональных проблем Дальневосточного отделения
Российской академии наук, Биробиджан, 679016 Россия

*e-mail: axanka@iacp.dvo.ru

**e-mail: frisman@mail.ru

Поступила в редакцию 05.12.2022 г.

После доработки 26.12.2022 г.

Принята к публикации 28.12.2022 г.

С точки зрения популяционной генетики возможность существования устойчивых различий на однородном по отбору ареале выглядит несколько парадоксально, однако именно дизруптивный отбор способен привести к такой дивергенции. Примеры популяционных систем с устойчивой дивергенцией генетических структур субпопуляций, обитающих на однородном ареале, встречаются в природе и создаются экспериментально. Сопоставление этих наблюдений и результатов моделирования дает возможность объяснить сохранение устойчивости природной генетической дивергенции механизмом взаимодействия дизруптивного отбора (в виде пониженной приспособленности гибридных форм) и слабого миграционного процесса. Хорошо согласуются с этим утверждением результаты экспериментов, выполненных Ю.П. Алтуховым и соавторами с ящичными популяциями *Drosophila melanogaster*, в которых была получена первичная дивергенция генетических структур субпопуляции по локусу α -ГДГ. Можно предположить, что существенную роль в поддержании этой дивергенции сыграло наличие дизруптивного отбора, в данном случае – пониженной приспособленности гетерозигот по α -ГДГ. Альтернативной гипотезой является закрепление различий в результате генетического дрейфа. Для того чтобы, с одной стороны, выявить и обосновать возможность наличия этого фактора в данной системе, а с другой, – оценить насколько существенную роль играет дизруптивный отбор в поддержании устойчивости первичной генетической дивергенции, в настоящей работе мы проанализировали математические модели динамики частот аллелей в большой панмиктической популяции и в системе из 30 локальных миграционно-связанных популяций. Сопоставление результатов моделирования с результатами эксперимента позволяет заключить, что с большой вероятностью в рассматриваемой искусственной популяционной системе имел место дизруптивный отбор по локусу α -ГДГ, который способствовал первичной генетической дивергенции.

Ключевые слова: эволюция, естественный отбор, случайный дрейф, математическое моделирование, полиморфизм, *Drosophila melanogaster*.

DOI: 10.31857/S0016675823060139, **EDN:** SUHMUF

Сохранение сбалансированного полиморфизма в стационарной среде в рамках классической теории популяционной генетики возможно в условиях сверхдоминирования [1, 2]. Хотя обнаружить такой тип отбора удается крайне редко, полиморфизм в природе достаточно распространен, что поддерживает неослабевающий исследовательский интерес к выявлению и изучению механизмов, препятствующих потере генетического разнообразия [3–7]. Значительное внимание уделяется моделированию действия факторов буферизации (например, отбору по ограниченному по-

лом признакам [8–10], различным материнским и отцовским генетическим эффектам [11, 12], банкам семян у однолетних растений [13], перекрывающимся поколениям [14–16]), которые в отсутствие сверхдоминирования расширяют возможности поддержания полиморфизма в стационарной внешней среде, так как ограничивают влияние негативного отбора при неблагоприятных условиях и препятствуют исчезновению менее приспособленных аллелей.

Наличие полиморфизма может оказаться первой предпосылкой, первым необходимым этапом

к первичной дивергентной эволюции популяций — возникновению различий в генетических структурах субпопуляций изначально однородной популяционной системы. В то же время устойчивая первичная генетическая дивергенция сама по себе оказывается дополнительным фактором поддержания полиморфизма в популяционной системе. Наиболее интересной и даже несколько парадоксальной с точки зрения популяционной генетики является возможность установления устойчивых различий на однородном по отбору ареале, т.е. в условиях одинаковости действия отбора во всех субпопуляциях популяционной системы. Единственный тип отбора, который может привести к такой дивергенции, — дизруптивный. Моделирование действия однородного по ареалу дизруптивного отбора даже в простейшем случае при пониженной приспособленности гетерозигот системы двух менделевских однолокусных диаллельных достаточно произвольным образом лимитированных популяций приводит к возможности генетической дивергенции, т.е. к возможности закрепления устойчивых различий в генетических структурах этих популяций [17–20]. Подобное действие дизруптивного отбора, направленного на поддержание устойчивых различий в генетических структурах миграционно-связанных популяций, можно ожидать и при более сложных системах, включающих обилие субпопуляций. Есть ряд интересных примеров устойчивой первичной генетической дивергенции природных популяций, обитающих фактически на однородном ареале (например, [21–26]), однако в научной литературе встречаются они нечасто. Р. Левонтин объяснял это нежеланием исследователей публиковать результаты своих неудачных попыток найти корреляцию между географической изменчивостью и степенью действия какого-либо природного фактора [27]. Тем не менее, можно привести примеры таких исследований. Весьма впечатляющие примеры устойчивой пространственной гетерогенности генетических структур природных популяций описаны в работе Ю.Н. Новоженова с соавт. [24]. В этой работе обнаружены, в частности, сильные различия по концентрации морф в трех (двух смежных и одной пространственно удаленной) популяциях майского хруща, *Melolontha hyppocastani*, диморфных по окраске особей. Причем в одной из смежных популяций существенно преобладала одна из морф, а в другой столь же существенно преобладала альтернативная морфа. Эти различия в концентрациях морф устойчиво сохранялись в течение 5-летнего периода исследований. Авторам не удалось обнаружить различий в почвах, типах леса, растительном покрове на территориях, занятых рассматриваемыми популяциями хрущей. Бесперспективной оказалась и попытка связать окраску жуков с возрастом вырубков. Таким образом, есть веские основания пола-

гать, что здесь имеет место дивергенция на однородном ареале.

Более убедительно показать возможность генетической дивергенции на однородном ареале, максимально исключив при этом влияние гетерогенности среды обитания, можно только в лабораторных условиях. Весьма масштабные эксперименты такого рода были выполнены Ю.П. Алтуховым и соавт. [28–31] с ящичными популяциями *Drosophila melanogaster*. В ходе этих экспериментов была, в частности, получена выраженная первичная дивергенция генетических структур субпопуляции по локусу α -ГДГ при практически идентичных условиях обитания в каждом из 30 ящиков. Трудно найти экспериментальную работу, целенаправленно организованную и сопоставимую по масштабу эксперимента. Нам представляется, что в свое время этот эксперимент был явно недооценен: на его основе можно проводить масштабные модельные исследования и развивать теорию первичных основ эволюции.

В настоящем исследовании мы с помощью математического моделирования попытаемся оценить роль, которую могло играть наличие дизруптивного отбора, в данном случае наличие пониженной приспособленности гетерозигот по α -ГДГ, в поддержании первичной генетической дивергенции, наблюдаемой в системе. Альтернативной гипотезой является закрепление различий генетических структур субпопуляций, происходящее фактически только в результате генетического дрейфа.

Мы проанализируем математические модели динамики частот аллелей в большой панмиктической популяции и в системе из 30 локальных миграционно-связанных популяций и сопоставим модельные динамики частот аллелей под действием дизруптивного отбора и без него с реальными данными, полученными в экспериментальной работе Ю.П. Алтуховым и соавт. [28–30].

В экспериментальном исследовании [28–30] Ю.П. Алтухов и коллеги изучали динамику аллельных частот аутосомных диаллельных локусов α -глицерофосфатдегидрогеназы (α -ГДГ: 2 аллеля, II хромосома) и эстеразы-6 (Эст-6: 2 аллеля, III хромосома) в системе из 30 последовательно миграционно-связанных ящичных популяций *Drosophila melanogaster*. В один из отсеков популяционного ящика были помещены 150 пар гетерозиготных по обоим локусам самцов и самок *D. melanogaster* и им была предоставлена возможность свободно размножаться и расселяться по ареалу. После заселения всех отсеков и стабилизации численности мух (порядка 135 особей) была основана панмиктическая популяция той же величины и практически полностью воспроизводившая генофонд эволюционной системы. Обе популяции все время находились в термостатируемой комнате ($24 \pm 1^\circ\text{C}$) и поддерживались в совершенно

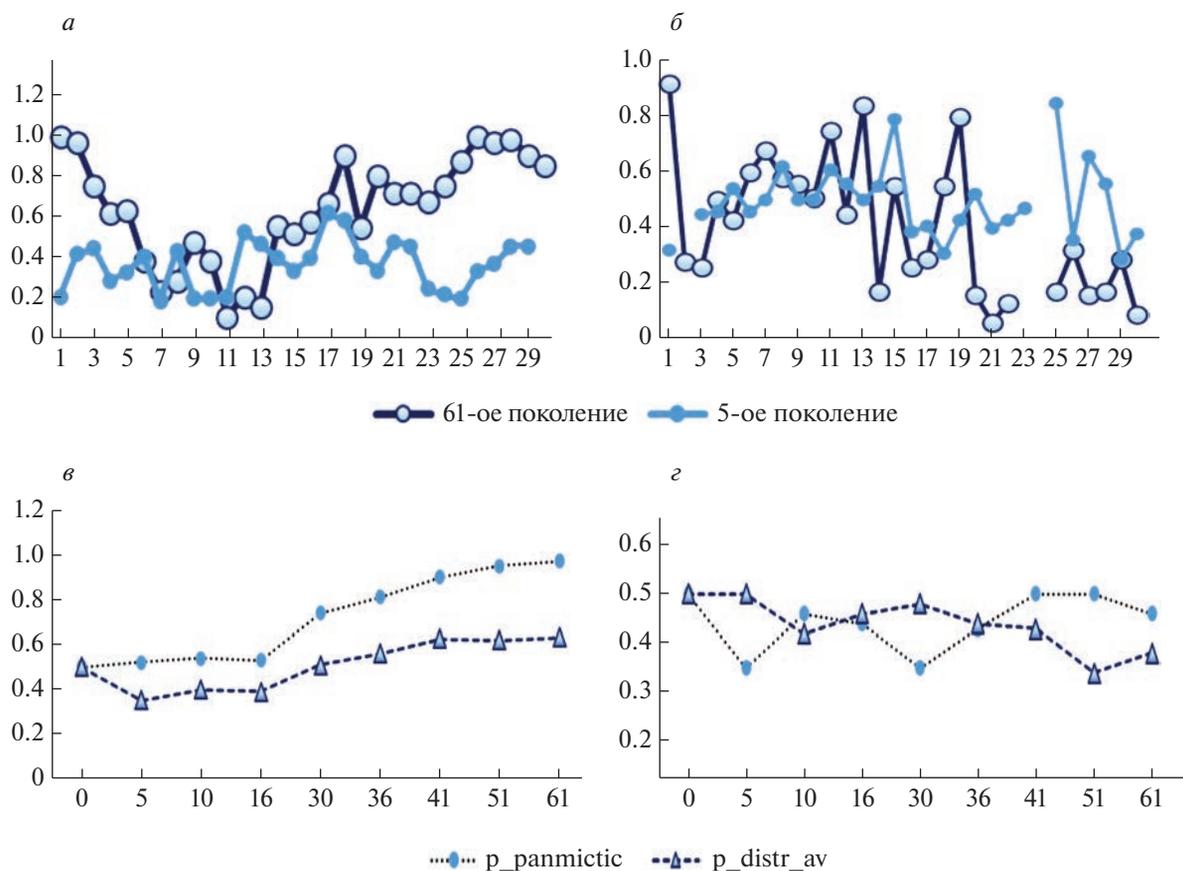


Рис. 1. Распределение частоты аллеля *F* локуса α -ГДГ (*a*) и Эст-6 (*б*) на уровне субпопуляций в экспериментально подразделенной популяции *D. melanogaster* для поколений 5 и 61; *в*, *г* – динамика частоты аллеля *F* локусов α -ГДГ (*в*) и Эст-6 (*г*) в череде поколений в панмиктической и подразделенной (средняя по всем субпопуляциям) популяциях. По абсциссе: *a*, *б* – номер ящика, *в*, *г* – номер поколения.

одинаковых условиях кормления, освещенности, влажности, смены поколений и т.п. Система эволюционировала около 70 поколений.

В результате этого эксперимента, условия которого обеспечивали низкий коэффициент миграций ($m \approx 0.03$) и сравнительно случайные начальные частоты аллелей в большинстве субпопуляций, в системе обнаружилась локальная дифференциация частот аллелей обоих локусов, несмотря на совершенно одинаковые условия в субпопуляциях (рис. 1, *a*, *б*). Кроме того, наблюдалось вытеснение одного из аллелей локуса α -ГДГ, что оказалось наиболее заметным в большой панмиктической популяции (рис. 1, *в*, *г*).

ДИНАМИКА ЧАСТОТ АЛЛЕЛЕЙ В БОЛЬШОЙ ПАНМИКТИЧНОЙ ПОПУЛЯЦИИ: ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ ОТБОРА ПО ДАННЫМ ЭКСПЕРИМЕНТА

Для моделирования экспериментальной динамики частот аллелей в большой панмиктической

популяции естественно использовать классическую модель отбора в популяции с неперекрывающимися поколениями [32]. Пусть F и f – альтернативные аллели локуса α -ГДГ, на который действует отбор, и q_n – частота одного из аллелей, например F в n -ом поколении. Считаем, что приспособленности генотипов $W_i = a_i f(N)$, где a_1 , a_2 и a_3 – относительные приспособленности генотипов FF , Ff и ff соответственно. Запишем уравнение динамики частоты аллеля F в смежных поколениях:

$$q_{n+1} = \frac{q_n(a_1 q_n + a_2(1 - q_n))}{\bar{a}_n}, \quad (1)$$

где $\bar{a}_n = a_1 q_n^2 + 2a_2 q_n(1 - q_n) + a_3(1 - q_n)^2$.

Полагая $a_1 = 1 + s_1$, $a_2 = 1$ и $a_3 = 1 + s_2$, где s_1 и s_2 – коэффициенты отбора для гомозигот, (1) можно переписать в виде:

$$q_{n+1} = \frac{q_n(1 + s_1 q_n)}{1 + s_1 q_n^2 + s_2(1 - q_n)^2}. \quad (2)$$

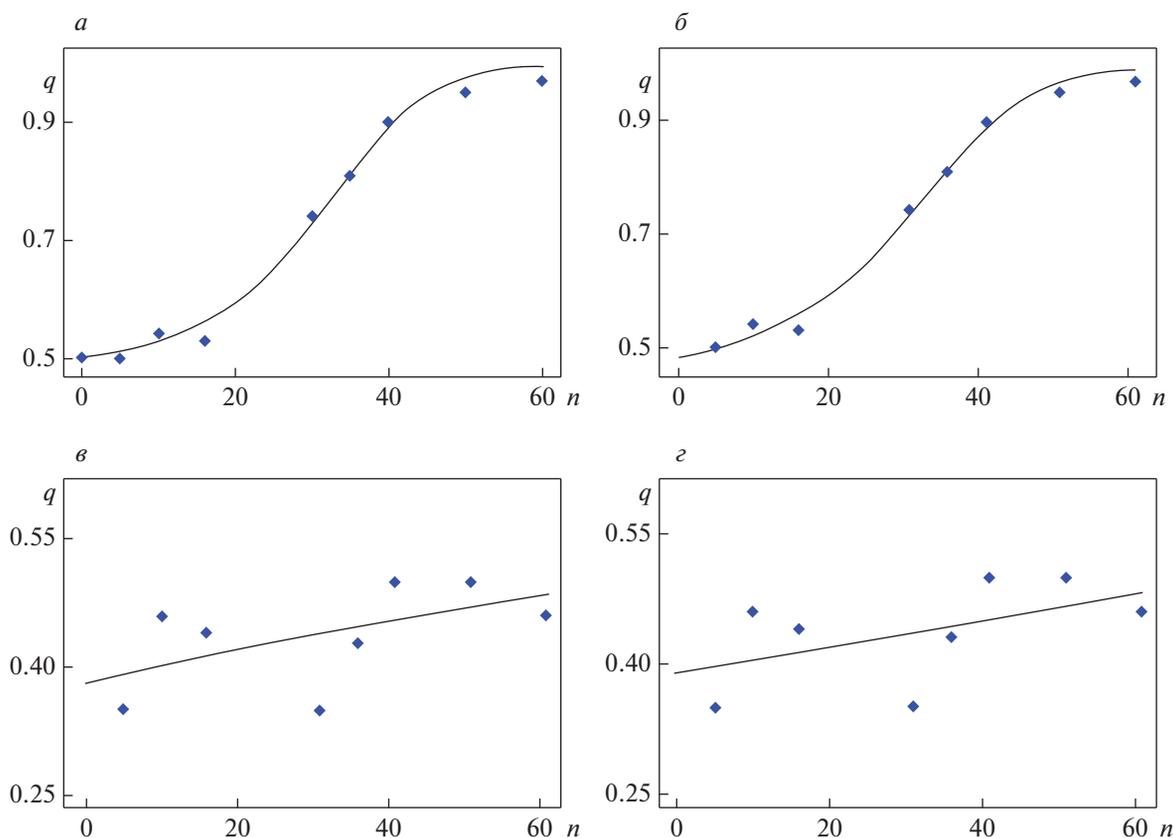


Рис. 2. Экспериментальная (синие точки) и модельная динамика частоты аллеля F локусов α -ГДГ ($a, б$) и Эст-6 ($в, г$) в череде поколений. $a: q_0 = 0.5, s_1 = 0.2, s_2 = 0.183, R^2 = 0.99$ (SSE = 0.0029); $б: q_0 = 0.483, s_1 = 0.173, s_2 = 0.14, R^2 = 0.99$ (SSE = 0.0023); $в: q_0 = 0.38, s_1 = -0.01, s_2 = -0.02, R^2 = 0.271$ (SSE = 0.0122); $г: q_0 = 0.39, s_1 = 0.009, s_2 = -0.004, R^2 = 0.274$ (SSE = 0.0116).

По экспериментальной динамике, представленной на рис. 2, видно, что частота одного из аллелей данного локуса в большой панмиктической популяции почти монотонно возрастала, и в конце концов этот аллель фиксировался. Это может происходить либо под действием обычного движущего отбора ($s_1 > 0; s_2 < 0$), либо быть вызвано деструктивным отбором ($s_1 > 0; s_2 > 0$).

Для того чтобы определить какой же тип отбора имеет место, необходимо по кривой динамики определить коэффициенты s_1 и s_2 . Для этого численно находились решения уравнения (2) при различных значениях коэффициентов s_1 и s_2 и фиксированном значении $q_0 = 0.5$. Из множества коэффициентов были выбраны значения, дающие минимальную сумму квадратов отклонения (SSE) расчетных значений от экспериментальных (рис. 2, a). Численный эксперимент был дополнен вариацией начального значения q_0 , поскольку из описания эксперимента следует, что начальная частота аллелей в панмиктической популяции определялась случайными факторами и сразу после отбора измерения ее не было, а первая точка измерения в пятом поколении расположена заметно

ниже уровня $q = 0.5$, что может свидетельствовать об отклонении начальной частоты аллеля F ниже уровня 0.5. Вариация q_0 действительно позволила улучшить аппроксимацию экспериментальной кривой, несколько уменьшив SSE и устранив гетероскедастичность (рис. 2, $б$).

Аналогично для локуса Эст-6 из множества коэффициентов (q_0, s_1 и s_2) были выбраны значения, дающие минимальную сумму квадратов отклонения расчетных значений от экспериментальных; лучшими оказались результаты, приведенные на рис. 2, $в, г$. Аппроксимация по этому локусу получилась значительно хуже. Только 27% общей вариации можно объяснить действием отбора, при этом тип отбора может быть как балансирующим (рис. 2, $в$), так и движущим (рис. 2, $г$).

МОДЕЛИРОВАНИЕ ДИНАМИКИ ЧАСТОТ АЛЛЕЛЕЙ В МНОГОПОПУЛЯЦИОННОЙ СИСТЕМЕ

Если полученные оценки верны, то становится понятен механизм возникновения локальной дифференциации в экспериментальной популя-

ционной системе. Из условий эксперимента ясно, что начальная концентрация аллеля F в каждой из субпопуляций, кроме нескольких первых (где она была равна 50%), задавалась случайным образом, так как основное ядро этой популяции основывали несколько (небольшое число) мигрантов из соседних популяций. Именно в силу малочисленности этих основателей субпопуляций основную роль при формировании начальных концентраций аллелей играли случайные процессы. Поэтому в ряде субпопуляций вполне могли изначально оказаться низкими концентрации аллеля F (причем наиболее вероятно, что такие субпопуляции располагаются смежными группами). В процессе возрастания и стабилизации численности субпопуляций в рассматриваемой популяционной системе, по-видимому, основную роль стали играть детерминистические процессы: пониженная приспособленность гетерозигот локуса α -ГДГ на фоне слабых миграций (и небольшого дрейфа генов). Именно эти процессы привели к локальной генетической дифференцировке (или генетической дивергенции) в описанной экспериментальной популяционной системе.

Таким образом, экспериментальные результаты по локусу α -ГДГ можно объяснить исходя из по крайней мере двух альтернативных гипотез: адапционной, т.е. предполагающей наличие дизруптивного отбора, и флуктуационной, т.е. предполагающей чисто случайную дрейфовую дифференциацию. Для того чтобы выяснить какая из этих гипотез более адекватна действительности, рассмотрим математическую модель динамики частот генов цепочки из 30 миграционно-связанных в кольцо популяций и проанализируем результаты ее реализации в компьютерной модели.

Блок-схема модели приведена на рис. 3, где показана связь между состояниями n -ого и $(n + 1)$ -ого поколения трех популяций, находящихся в соседних ящиках: двадцать девятом, тридцатом и первом.

Пусть n -ое поколение для этих популяций характеризуется следующими величинами: $N_1^i(n)$, $N_2^i(n)$, $N_3^i(n)$ – численности особей с генотипами FF , Ff , ff соответственно в i -ом ящике; $q_1^i(n)$, $q_2^i(n)$, $q_3^i(n)$ – частоты этих генотипов. Число мигрантов детерминированно определялось количеством особей в ящиках, между которыми происходила миграция. Генотип каждого мигранта выбирался случайно, причем вероятность выбора определенного генотипа полагалась равной его частоте в популяции, из которой данный мигрант происходил. Таким образом, фактически моделировалось влияние генетического дрейфа при пространственном распределении особей. После завершения процесса миграции и формирования новых

(“постмиграционных”) частот генотипов в каждой популяции ($\tilde{q}_1^i(n)$, $\tilde{q}_2^i(n)$, $\tilde{q}_3^i(n)$) моделировались процессы размножения и отбора, т.е. вычислялись параметры состояния $(n + 1)$ -го поколения ($N_1^i(n)$, $q_1^i(n)$, $i = 1, \dots, 30$, $j = 1, \dots, 3$). Процессы размножения и отбора считались детерминированными: переход от $\tilde{q}_j^i(n)$ к $q_j^i(n + 1)$ осуществлялся по формулам:

$$\begin{aligned} q_1^i(n + 1) &= (1 + s_1)(\tilde{q}_1^i(n) + \tilde{q}_2^i(n)/2)^2 / G, \\ q_2^i(n + 1) &= 2(\tilde{q}_1^i(n) + \tilde{q}_2^i(n)/2)(\tilde{q}_3^i(n) + \tilde{q}_2^i(n)/2) / G, \\ q_3^i(n + 1) &= (1 + s_2)(\tilde{q}_3^i(n) + \tilde{q}_2^i(n)/2)^2 / G. \end{aligned} \quad (3)$$

$$G = 1 + s_1(\tilde{q}_1^i(n) + \tilde{q}_2^i(n)/2)^2 + s_2(\tilde{q}_3^i(n) + \tilde{q}_2^i(n)/2)^2.$$

Число особей в каждом ящике после размножения и отбора считалось постоянным и равным N , поэтому количество особей каждого генотипа для $(n + 1)$ -го поколения определялось соотношениями: $N_j^i(n + 1) = q_j^i(n + 1)N$, где $N = 135$ (согласно экспериментальным данным).

В соответствии с моделируемым экспериментом мы полагали, что в начальный момент времени населен весь “ареал” популяционного ящика, однако число особей в отсеках может варьировать от 1 до 37. К сожалению, из описания эксперимента остается неясным, насколько неравномерным оказалось расселение особей по ящикам. Однако полагая, что пары разнополых особей достаточно для производства 135 особей следующего поколения, а генетический состав мужской и женской частей популяции в нулевом поколении идентичен (все мушки были гетерозиготными), можно исключить из рассмотрения половую структуру экспериментальной популяции. Неравномерность начального распределения мушек по ящикам проявится в том, что некоторые субпопуляции не оставят потомства в первый период размножения, так как одна особь или несколько однополых особей, случайно оказавшихся всем населением целой субпопуляции, не смогут произвести потомство. Такие ящики и будут, по-видимому, основными точками дифференциации генетического состава, так как в следующем поколении их заселит небольшое количество мигрантов (по 4 из каждого соседнего ящика: $135 \times 0.03 = 4$). Генетический состав этих мигрантов может значительно отличаться от среднего их родной популяции за счет ошибки выборочности. Количество ящиков (k), в которых мушки не оставят потомства в первом (а может быть и следующем) поколении, возможно, повлияет на результаты моделирования.

Учет неравномерности начального распределения мушек по ящикам и наличия “пустых” ящиков, в которых по каким-либо причинам

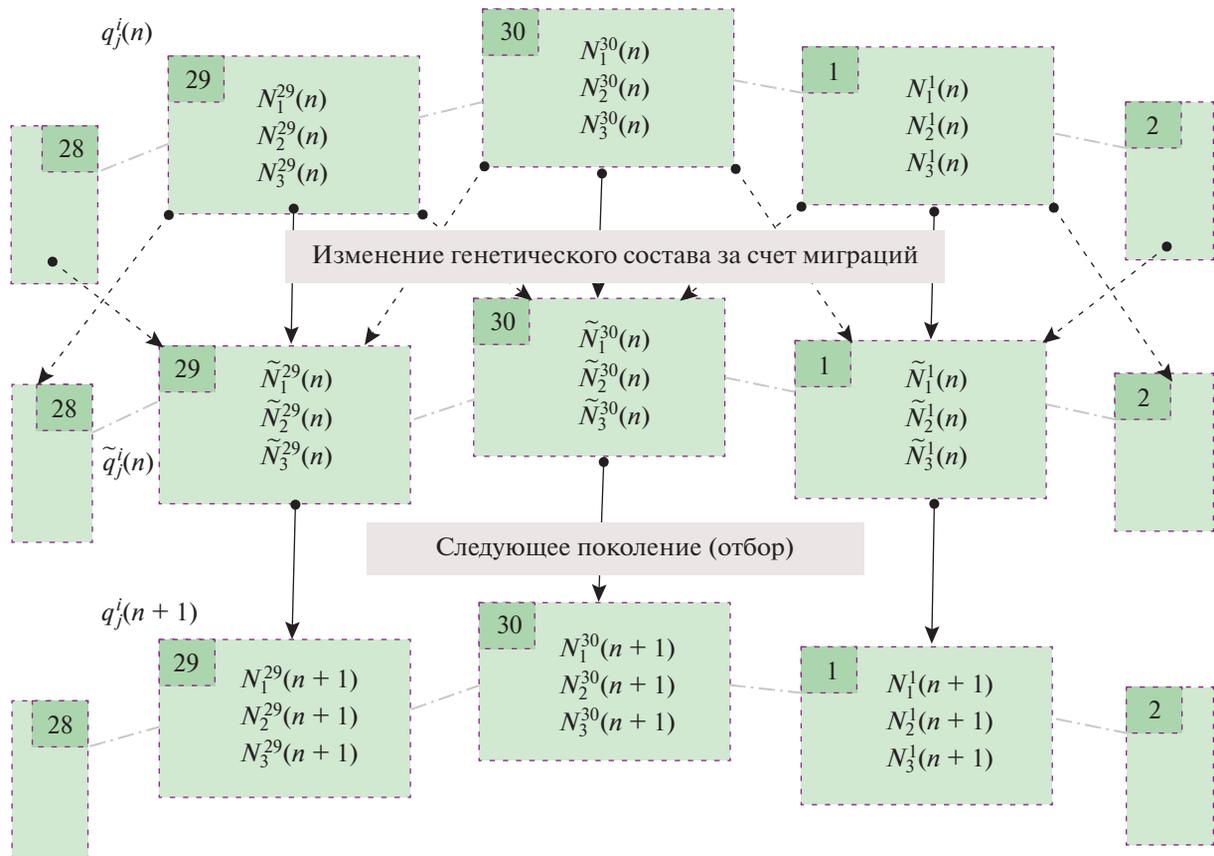


Рис. 3. Блок-схема модели ящичной популяции. Представлен переход популяций трех соседних ящиков (29-го, 30-го и 1-го) между n -м и $(n + 1)$ -м поколениями. $N_1^i(n), N_2^i(n), N_3^i(n)$ – численности особей с генотипами FF, Ff, ff соответственно в i -ом ящике; $q_1^i(n), q_2^i(n), q_3^i(n)$, – частоты этих генотипов. Миграции в соседние ящики обозначены пунктирными стрелками; $\tilde{N}_1^i(n), \tilde{N}_2^i(n), \tilde{N}_3^i(n), \tilde{q}_1^i(n), \tilde{q}_2^i(n), \tilde{q}_3^i(n)$ – постмиграционные численности и частоты соответственно.

мушки не оставили потомства в первом поколении, значительно меняет картину.

На рис. 4 представлены результаты по одной из реализаций случайного процесса динамики частот аллелей в системе ящичных популяций при неоднородном начальном распределении в отсутствие отбора и при дизруптивном отборе. При наличии отбора в модельной подразделенной популяции распределение частоты аллеля F по ящикам в 61-ом поколении в общих чертах напоминает экспериментальное, демонстрируя значительные различия по генетическому составу между группами субпопуляций: часть популяций оказывается практически мономорфной по аллелю F , другие – по альтернативному аллелю (f) и несколько полиморфных (рис. 4,а). При этом модельная динамика средней по всем популяциям частоты аллеля F оказывается подобна экспериментальной (рис. 4,б; SSE = 0.0045, $R^2 = 0.95$).

В случае без отбора модельное распределение частоты аллеля F по ящикам в 61-ом поколении оказывается более однородным, без выраженных

пиков: все субпопуляции в большей или меньшей степени полиморфны (рис. 4,б). При этом динамика средней по всем популяциям частоты аллеля F в отсутствие отбора значительно отличается от экспериментальной: средняя частота устанавливается в первые несколько поколений и далее практически не меняется; на рис. 4,г пример динамики средней частоты аллеля F , которая в 61-ом поколении демонстрирует близкое к наблюдаемому в эксперименте значение ($q \approx 0.65$), при этом SSE = 0.1879 и $R^2 < 0$.

Множественные реализации случайного процесса в многопопуляционной системе. Место экспериментальной динамики в общей статистике

Проведенное моделирование было дополнено множественными реализациями случайного процесса и оценкой его основных характеристик. Рассмотрим динамику средней (по 30 субпопуляциям) частоты аллеля F в 1000 реализациях случайного процесса при фиксированном начальном профиле

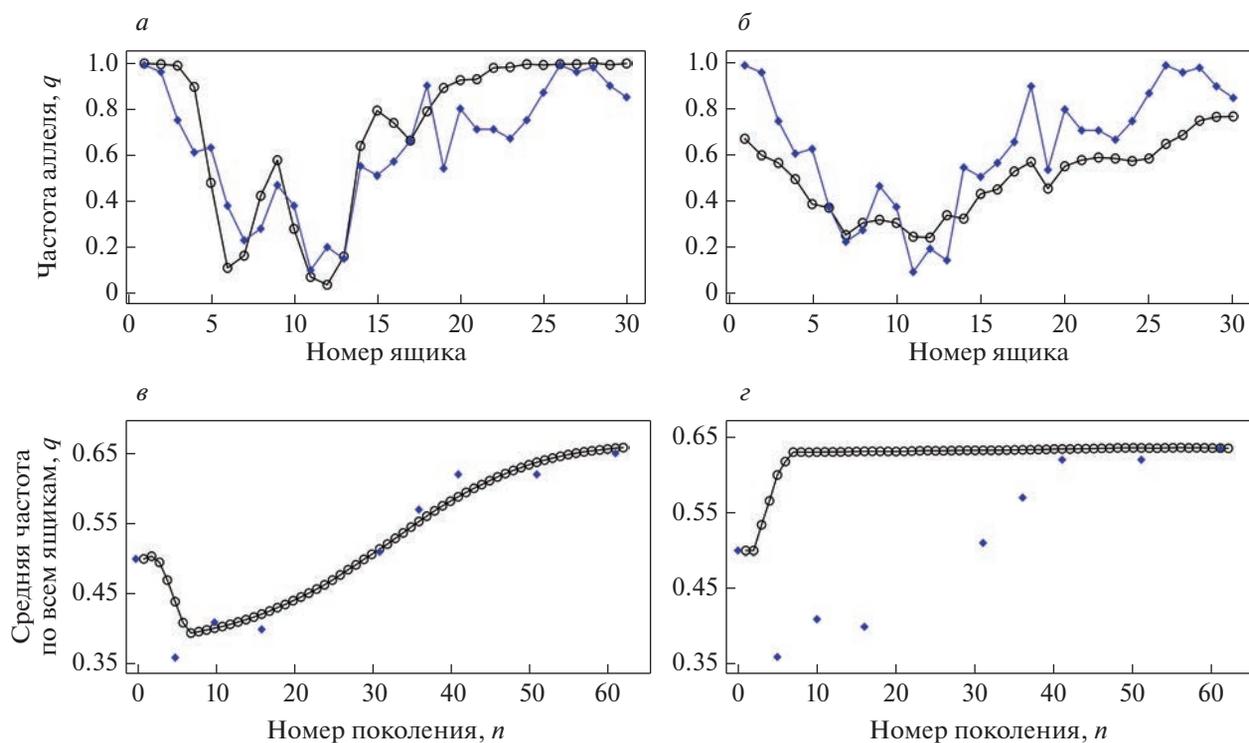


Рис. 4. Распределение частот аллеля F в 61-ом поколении в подразделенной популяции (черная линия – модель, синяя – экспериментальная) (*a*, *b*); *v*, *z* – динамика средней по всем субпопуляциям частоты аллеля F (черная линия – модель, синие точки – эксперимент); *a*, *v* – дизруптивный отбор ($s_1 = 0.175$, $s_2 = 0.139$); *b*, *z* – случайный дрейф ($s_1 = 0$, $s_2 = 0$). В начальном распределении были ящики ($k = 19$), в которых мушки не оставили потомства.

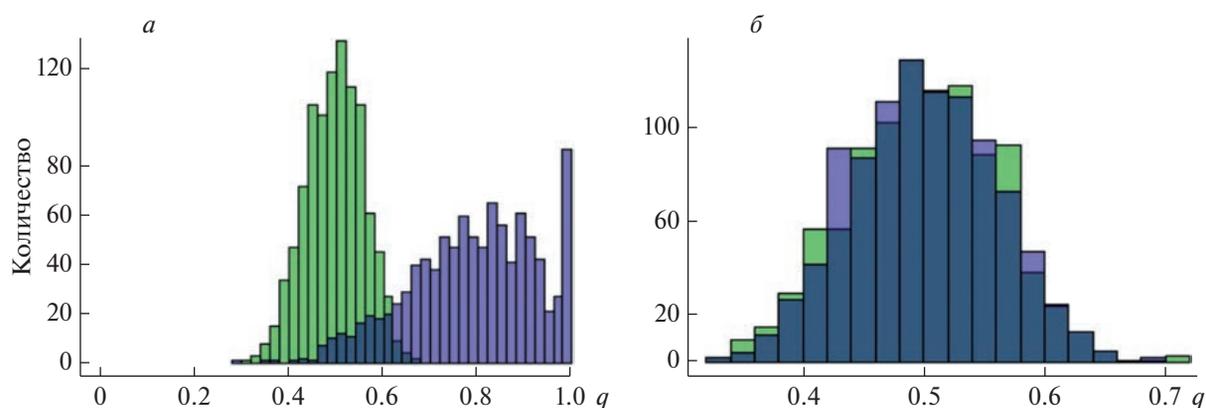


Рис. 5. Гистограмма распределения средней по всем ящикам частоты аллеля F (q) в 1000 реплик (61-ое поколение). Зеленый – без отбора, фиолетовый – с отбором (*a* – дизруптивный, *b* – балансирующий).

заселения ящиков ($k = 19$). Оказалось, что сначала (в пятом поколении) распределения в обоих случаях (без отбора и при дизруптивном отборе с количественными характеристиками, определенными по данным для локуса α -ГДГ) очень похожи друг на друга и напоминают нормальное, а далее они все больше и больше расходятся. В случае без отбора распределение становится практически идеально

нормальным, а при дизруптивном отборе смещается вправо. К 61-ому поколению различия распределений значительны (рис. 5, *a*).

Теперь естественно поставить вопрос: к какому из распределений (с большой вероятностью) принадлежат значения частот аллеля F α -ГДГ, полученные в эксперименте. Место наблюдаемой в эксперименте средней частоты аллеля F ($q = 0.65$)

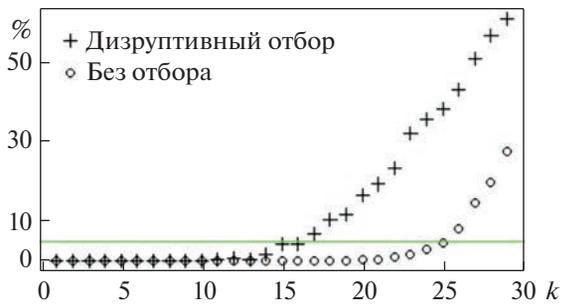


Рис. 6. Квантиль наблюдаемой в эксперименте средней частоты аллеля $F(q = 0.65)$ в ранжированном ряду реплик без отбора (круг) и с отбором (плюс) в зависимости от количества пустых ящиков в начале размножения (k). Зеленая линия — 5%-ный квантиль.

в ранжированном ряду реплик без отбора — 996, а в ряду результатов с отбором — 162. То есть наблюдаемая в эксперименте средняя частота аллеля лежит вне 99%ДИ (ограниченного значениями 5 и 995) в случае отсутствия отбора, при этом она хорошо укладывается в 95%ДИ частот с отбором (ограниченного значениями 25 и 975). Поэтому гипотезу об отсутствии отбора можно отвергнуть с доверительной вероятностью более 99%.

На рис. 5,б приведены распределения средней по всем ящикам частоты аллеля F в 1000 модельных репликах в 61-ом поколении в случаях как без отбора, так и для балансирующего отбора с количественными характеристиками, определенными по данным для локуса Эст-6. Балансирующий отбор слегка уменьшает дисперсию и нарушает симметрию распределения, поэтому здесь трудно дифференцировать распределения.

Дополнительное исследование было проведено с целью понять, насколько вариация количества пустых ячеек в начале размножения могла бы повлиять на результат эволюции ящичной популяции. Поскольку количество пустых ячеек в начале размножения из эксперимента точно не известно, вопрос о выборе фиксированного значения $k \neq 0$ вносит в моделирование дополнительную неопределенность. Множественные реализации случайного процесса (1000 реплик) для каждого значения k от 1 до 29 позволили определить место наблюдаемой в эксперименте средней частоты аллеля $F(q = 0.65)$ в ранжированном ряду реплик с отбором и без отбора (рис. 6). Оказалось, что при малом числе пустых ячеек в начале размножения ($k < 10$) вообще нельзя получить значение средней частоты q , совпадающее с наблюдаемым в эксперименте: при случайном дрейфе во всех ящиках $q < 0.65$, а при отборе $q > 0.65$. То есть при малых k ($k < 10$) при дизруптивном отборе вытеснение аллеля f идет быстрее, чем в эксперименте, а без отбора средние значения q лишь незначитель-

но отклоняются от 0.5. При больших k ($10 < k < 17$) в отдельных репликах появляются значения q близкие к наблюдаемому в эксперименте; начиная с $k = 17$ при дизруптивном отборе уже более 5% реплик демонстрируют скорость вытеснения аллеля f ниже, чем в эксперименте. При $16 < k < 25$ реализации с отбором и без него хорошо различимы друг от друга, так как более 95% реплик без отбора имеют $q < 0.65$, а при отборе экспериментальное значение $q = 0.65$ хорошо попадает в 95%ДИ. При большом количестве пустых ячеек $k > 25$ доверительные интервалы с отбором и без него начинают перекрываться и дифференцировать тот, которому принадлежит экспериментальное значение, уже невозможно.

Учитывая, что коэффициент миграции не большой, а в пятом поколении уже во всех ячейках наблюдалось успешное размножение, можно предположить, что количество пустых ячеек не достигало экстремальных уровней ($k \geq 25$) и, по видимому, находилось в пределах $16 < k < 25$.

Резюмируя, можно заключить, что в присутствии дизруптивного отбора первичная генетическая дивергенция в системе популяций, связанных стохастическими миграциями, возникает с достаточно большой вероятностью и оказывается структурно устойчивой. При этом сопоставление данных эксперимента с результатами моделирования позволяет заключить, что в построенной Ю.П. Алтуховым с сотрудниками искусственной популяционной системе наверняка имел место дизруптивный отбор, который способствовал первичной генетической дивергенции популяций.

Если взглянуть на протекающие процессы несколько шире, то можно прийти к следующему обобщению. Поскольку исходное ядро экспериментальной популяционной системы создавалось гибридными особями, можно предположить, что в ходе ее развития параллельно проходили два разнонаправленных процесса. Первый заключался в том, что формировались блоки генов с повышенной приспособленностью гетерозиготных форм. Естественный отбор в этом случае был направлен на поддержание полиморфизма аллелей этих генов. Второй процесс был связан с расщеплением гетерозиготных форм и выявлением генов, характеризующихся повышенной приспособленностью гомозигот и дизруптивным характером динамики частот аллелей. В каждый из этих процессов были вовлечены и аллели генов, адаптивно нейтральные в данной экспериментальной системе. В силу тех или иных причин, основной из которых является, по-видимому, сцепление, эти нейтральные аллели в данной ситуации оказались генетическими маркерами соответствующих адаптивных процессов. Так, локус Est-6 оказался таким маркером для процессов, направленных на поддержание полиморфизма аллелей, а локус α -ГДГ —

маркером дизруптивного отбора, протекающего в системе и способствующего первичной генетической дивергенции.

Работа выполнена в рамках Государственных заданий Института автоматики и процессов управления ДВО РАН, Института комплексного анализа региональных проблем ДВО РАН.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Wright S. The theory of gene frequencies // *Evol. Genet. Populations*. 1969. V. 2. P. 290–344.
2. Nagylaki T. *Introduction to Theoretical Population Genetics*. Berlin: Springer-Verlag, 1992. 370 p.
3. Frank S.A., Slatkin M. Evolution in a variable environment // *Am. Nat.* 1990. V. 136. № 2. P. 244–260.
4. Morozov A. Yu., Pasternak A. F., Arashkevich E. G. Revisiting the role of individual variability in population persistence and stability // *PLoS One*. 2013. V. 8. № 8. e70576. <https://doi.org/10.1371/journal.pone.0070576>
5. Mougi A. Rapid evolution of prey maintains predator diversity // *PLoS One*. 2019. V. 14. № 12. e0227111. <https://doi.org/10.1371/journal>
6. Фрисман Е.Я., Жданова О.Л., Неверова Г.П. Эколого-генетические модели в популяционной биофизике // *Биофизика*. 2020. Т. 65. № 5. С. 949–966. <https://doi.org/10.31857/S0006302920050130>
7. Неверова Г.П., Жданова О.Л., Фрисман Е.Я. Возникновение сложных режимов динамики численности в ходе эволюции структурированной лимитированной популяции // *Генетика*. 2020. Т. 56. № 6. С. 714–725. <https://doi.org/10.31857/S0016675820060065>
8. Reinhold K. Evolutionary genetics of sex-limited traits under fluctuating selection // *J. Evol. Biol.* 1999. V. 12. № 5. P. 897–902.
9. Reinhold K. Maintenance of a genetic polymorphism by fluctuating selection on sex-limited traits // *J. Evol. Biol.* 2000. V. 13. P. 1009–1014.
10. Жданова О.Л., Фрисман Е.Я. Математическое моделирование отбора по ограниченному полом признаку: к вопросу о существовании полиморфизма по размеру помета в естественных популяциях песцов // *Генетика*. 2021. Т. 57. № 2. С. 229–240. <https://doi.org/10.31857/S0016675821020156>
11. Gavrilets S. One-locus two-allele models with maternal (parental) selection // *Genetics*. 1998. V. 149. P. 1147–1152. <https://doi.org/10.1093/genetics/149.2.1147>
12. Yamamichi M., Hosono M. Roles of maternal effects in maintaining genetic variation: Maternal storage effect // *Evolution*. 2017. V. 71. № 2. P. 449–457. <https://doi.org/10.1111/evo.13118>
13. Turelli M., Schentske D.W., Bierzychudek P. Stable two-allele polymorphisms maintained by fluctuating fitnesses and seed banks: Protecting the blues in *Linanthus parryae* // *Evolution*. 2001. V. 55. P. 1283–1298. <https://doi.org/10.1111/j.0014-3820.2001.tb00651.x>
14. Svardal H., Rueffler C., Hermisson J. Comparing environmental and genetic variance as adaptive response to fluctuating selection // *Evolution*. 2011. V. 65. P. 2492–2513. <https://doi.org/10.1111/j.1558-5646.2011.01318.x>
15. Жданова О.Л., Фрисман Е.Я. Математическое моделирование механизма дифференциации репродуктивных стратегий в естественных популяциях (на примере песцов, *Alopex lagopus*) // *Компьютерные исследования и моделирование*. 2016. Т. 8. № 2. С. 213–228. <https://doi.org/10.20537/2076-7633-2016-8-2-213-228>
16. Zhdanova O.L., Frisman E. Ya. Genetic polymorphism under cyclical selection in long-lived species: The complex effect of age structure and maternal selection // *J. Theor. Biol.* 2021. V. 512. № 7. 110564. <https://doi.org/10.1016/j.jtbi.2020.110564>
17. Базыкин А.Д. Пониженная приспособленность гетерозигот в системе двух смежных популяций // *Генетика*. 1972. Т. 8. № 11. С. 155–161.
18. Базыкин А.Д. Отбор и генетическая дивергенция в системах локальных популяций и популяциях с непрерывным ареалом (математическая модель) // *Проблемы эволюции*. 1973. Т. 3. С. 231–241.
19. Фрисман Е.Я. Первичная генетическая дивергенция (теоретический анализ и моделирование). Владивосток: ДВНЦ АН СССР, 1986. 160 с.
20. Кулаков М.П., Фрисман Е.Я. Простая и сложная динамика в модели эволюции двух миграционно связанных популяций с непересекающимися поколениями // *Изв. высших учебных заведений. Прикладная нелинейная динамика*. 2022. Т. 30. № 2. С. 208–232. <https://doi.org/10.18500/0869-6632-2022-30-2-208-232>
21. Yosida T.H., Tsuchiya K., Moriwaki K. Frequency of chromosome polymorphism in *Rattus rattus* collected in Japan // *Chromosoma*. 1971. V. 33. № 1. P. 30–40.
22. Yosida T.H., Tsuchiya K., Moriwaki K. Karyotypic differences of black rats, *Rattus rattus*, collected in various localities of East and Southeast Asia and Oceania // *Chromosoma*. 1971. V. 33. № 3. P. 252–267.
23. Blake J.A. Complex chromosomal variation in natural populations of the Jamaican lizard, *Anolis grahami* // *Genetica*. 1986. V. 69. P. 3–17. <https://doi.org/10.1007/BF00122929>
24. Новоженев Ю.И., Береговой В.Е., Хохоткин М.И. Обнаружение границ элементарных популяций у полиморфных видов по частоте встречаемости форм // *Проблемы эволюции*. Новосибирск: Наука, СО, 1973. Т. 3. С. 252–260.
25. Гордеева Н.В., Салменкова Е.А., Алтухов Ю.П. Исследование генетической дивергенции горбуши, вселенной на Европейский Север России, с использованием микросателлитных и аллозимных локусов // *Генетика*. 2006. Т. 42. № 3. С. 349–360.

26. *Kandul N.P., Lukhtanov V.A., Pierce N.E.* Karyotypic diversity and speciation in *Agrodiaetus butterflies* // *Evolution*. 2007. V. 61. № 3. P. 546–559. <https://doi.org/10.1111/j.1558-5646.2007.00046.x>
27. *Левонтин Р.* Генетические основы эволюции. М.: Мир, 1977. 351 с.
28. *Алтухов Ю.П., Бернашевская А.Г.* Экспериментальное моделирование динамики генных частот в системе полуизолированных популяций // *ДАН СССР*. 1978. Т. 238. № 3. С. 712–714.
29. *Алтухов Ю.П., Бернашевская А.Г., Милюшиников А.Н., Новикова Т.А.* Экспериментальное моделирование генетических процессов в популяционной системе *Drosophila melanogaster*, соответствующей кольцевой ступенчатой модели. Сообщение I. Обоснование подхода и особенности локальной дифференциации частот аллелей α -глицерофосфатдегидрогеназы и эстеразы-6 // *Генетика*. 1979. Т. 15. № 4. С. 646–653.
30. *Алтухов Ю.П., Бернашевская А.Г.* Экспериментальное моделирование генетических процессов в популяционной системе *Drosophila melanogaster*, соответствующей кольцевой ступенчатой модели. Сообщ. 2. Стабильность аллельного состава и периодическая зависимость изменчивости частот аллелей от расстояния // *Генетика*. 1981. Т. 17. № 6. С. 1052–1059.
31. *Алтухов Ю.П.* Генетические процессы в популяциях. М.: Академкнига, 2003. 431 с.
32. *Ратнер В.А.* Математическая популяционная генетика (элементарный курс). Новосибирск: Наука, СО, 1977. 126 с.

About the Genetic Divergence of Populations Coupled by Migrations: Modern Modeling Based on Experimental Results of Yu.P. Altukhov and Coauthors

O. L. Zhdanova^{a, *} and E. Ya. Frisman^{b, **}

^a*Institute for Automation and Control Processes, Far East Branch of Russian Academy of Sciences, Vladivostok, 690041 Russia*

^b*Complex Analysis and Regional Problems Institute, Far East Branch of Russian Academy of Sciences, Birobidzhan, 679016 Russia*

*e-mail: axanka@iacp.dvo.ru

**e-mail: frisman@mail.ru

The existence possibility of stable differences in a selection-homogeneous range looks somewhat paradoxical from the population genetics point of view, but it is disruptive selection that can give such a divergence. Examples of population systems with stable divergence of the genetic structures of subpopulations living in a homogeneous area occur in nature and experimental systems. Comparison of these observations and modeling results allow us to explain the persistence of the stability of natural genetic divergence by the mechanism of interaction between disruptive selection (in the form of reduced fitness of hybrid forms) and a weak migration process. The results of experiments performed by Yu.P. Altukhov et al. with box populations of *Drosophila melanogaster*, in which the primary divergence of the genetic structures of the subpopulation at the α -GDH locus appeared, are in good agreement with this statement. One can assume the disruptive selection, in this case, the reduced fitness of heterozygotes in α -HDG, played the significant role in maintaining this divergence. An alternative hypothesis is genetic drift that fixed the differences. To identify and substantiate the possibility of the presence of this factor in a given system, and to assess the significant role that disruptive selection plays in maintaining the stability of primary genetic divergence, we analyzed mathematical models of the allele frequencies' dynamics in a large panmictic population and in a system of 30 local migration coupled populations. Comparison of the simulation results with those of the experiment allows us to conclude with a high probability, in the considered experimental population system, there was a disruptive selection at the α -GDH locus that facilitated the primary genetic divergence.

Keywords: evolution, natural selection, genetic drift, mathematical modeling, polymorphism, *Drosophila melanogaster*.

РЕПАРАЦИЯ ДНК НЕ ИНИЦИИРУЕТ ГЕНЕТИЧЕСКУЮ НЕСТАБИЛЬНОСТЬ ДРОЖЖЕВЫХ КЛЕТОК

© 2023 г. Е. С. Евстратова¹ *, В. Г. Петин², С. А. Гераскин³

¹Национальный медицинский исследовательский центр радиологии, Калужская обл., Обнинск, 249031 Россия

²Медицинский радиологический научный центр им. А.Ф. Цыба – филиал Национального медицинского исследовательского центра радиологии, Калужская обл., Обнинск, 249031 Россия

³Всероссийский научно-исследовательский институт радиологии и агроэкологии, Калужская обл., Обнинск, 249032 Россия

*e-mail: ekevs7240@mail.ru

Поступила в редакцию 30.09.2022 г.

После доработки 21.11.2022 г.

Принята к публикации 24.11.2022 г.

Приводятся экспериментально полученные кривые зависимости задержки формирования колоний от дозы ионизирующего излучения (гамма-кванты ⁶⁰Со, мощность дозы 10 Гр/мин) для шести штаммов гаплоидных и диплоидных дрожжей дикого типа (способных восстанавливаться от радиационных повреждений и характеризующихся сигмоидными кривыми выживаемости), а также для шести гаплоидных и диплоидных радиочувствительных мутантов, дефектных по репарации и характеризующихся экспоненциальными кривыми выживаемости. Задержка формирования колоний после облучения является проявлением генетической нестабильности. Показано, что для всех диплоидных штаммов генетическая нестабильность достигала 100% с увеличением дозы ионизирующего излучения, независимо от формы кривых выживаемости и способности клеток восстанавливаться от радиационных повреждений. Наоборот, для всех гаплоидных штаммов генетическая нестабильность была близка к 20%. В отличие от традиционных представлений, полученные нами данные указывают на то, что позднее формирование колоний выжившими после облучения клетками определяется главным образом плоидностью клеток и не зависит от формы кривых доза-эффект и радиочувствительности клеток. Это означает, что репарация ДНК не инициирует генетическую нестабильность дрожжевых клеток.

Ключевые слова: генетическая нестабильность, выживаемость, дрожжевые клетки, радиочувствительные мутанты, гамма-кванты.

DOI: 10.31857/S0016675823060048, **EDN:** SSBVOM

Одной из широко распространенных точек зрения на механизм возникновения генетической нестабильности является ошибочное восстановление повреждений, индуцированных ионизирующими излучениями, УФ светом или другими ДНК-повреждающими агентами. Описаны данные, показывающие, что дефекты в восстановлении ДНК вызывают генетическую нестабильность в клетках различного происхождения [1–5]. В работе [1] приводится доказательство, что двуниевые разрывы ДНК запускают сценарий задержанной репродуктивной гибели клеток китайского хомьяка, характеризующей генетическую нестабильность. Преждевременное старение человека в ряде случаев объясняют ошибочным восстановлением повреждений ДНК [2], приводящим к генетической нестабильности и, соответственно, к заболеваниям, сокращающим продолжительность жизни. Дефекты в восстановлении ДНК также индуци-

руют генетическую нестабильность и в популярном объекте исследований *Saccharomyces cerevisiae* [3–5]. Эти результаты согласуются с данными, показавшими, что эффект задержанного по сравнению с контролем колонеобразования, также характеризующий генетическую нестабильность, в большей степени характерен для диплоидных, а не гаплоидных штаммов дрожжей [6]. Известно, что способность диплоидных дрожжевых клеток восстанавливаться от радиационных повреждений более эффективно реализуется в диплоидных, а не гаплоидных дрожжевых клетках [7, 8]. Ранее нами были опубликованы данные по генетической нестабильности дрожжевых клеток различного генотипа после облучения УФ-светом [9–11]. Было показано, что генетическая нестабильность в большей степени зависит от плоидности клеток, а не от их способности восстанавливаться от УФ-повреждений.

Таблица 1. Количественные параметры, характеризующие радиочувствительность и генетическую нестабильность гаплоидных и диплоидных дрожжевых клеток

Штамм	Генотип	Максимальная генетическая нестабильность, %	LD_{90} , Гр	$\frac{LD_{90}(\text{дипл})}{LD_{90}(\text{гапл})}$	$\frac{LD_{90}(\text{дикий})}{LD_{90}(\text{мутант})}$
K7	<i>RAD</i>	15	155 ± 10	4.1	
Мегри 139-В	<i>RAD/RAD</i>	100	640 ± 44		
211-1a	<i>RAD</i>	17	140 ± 12	6.6	
211	<i>RAD/RAD</i>	100	920 ± 30		
S288C	<i>RAD</i>	20	160 ± 15	5.8	–
XS800	<i>RAD/RAD</i>	100	930 ± 28		–
XS774-4d	<i>rad51</i>	18	65 ± 5	2.3	2.5
XS806	<i>rad51/rad51</i>	100	150 ± 12		6.2
g160/2d	<i>rad52</i>	20	132 ± 14	1.7	1.2
XS1898	<i>rad52/rad52</i>	95	220 ± 19		4.2
g218/7c	<i>rad54</i>	19	45 ± 6	2.1	3.5
L9	<i>rad54/rad54</i>	86	93 ± 11		10.0

В настоящей работе мы приводим результаты о генетической нестабильности разных гаплоидных и диплоидных дрожжевых штаммов дикого типа и их радиочувствительных мутантов после облучения ионизирующим излучением (гамма-кванты ^{60}Co при мощности дозы 10 Гр/мин). Гаплоидные штаммы *Zigosaccharomyces bailii* и *Saccharomyces cerevisiae* получены от В.И. Корогодина (Россия), который получил их от R.K. Mortimer (США). Диплоидные дрожжевые клетки дикого типа (*RAD/RAD*) – штаммы Мегри 139-В, 211 и XS800, а также радиочувствительные гомозиготные мутанты диплоидного типа – штаммы XS806 и XS1898, первоначально получены от Dr. S. Nakai (Япония). Штаммы 211-1a и 211 получены от W. Laskowski (Западный Берлин), штамм L9 – от И.П. Арман (Россия).

В наших опытах перед облучением клетки выращивали до стационарной стадии на твердой питательной среде до тех пор, пока процент почкующихся клеток становился минимальным, и облучаемая суспензия содержала 1–2% почкующихся клеток. Выживаемость клеток оценивали отношением числа колоний, сформированных выжившими после облучения клетками, к числу колоний, образованных в контроле после культивации на стандартной твердой питательной среде в течение трех–четырёх суток. Для количественной оценки генетической нестабильности подсчитывали процент колоний, образованных позже появления контрольных колоний выжившими после облучения клетками. Этот тест использовали многие авторы для клеток различного происхождения [1–4, 6, 7]. Каждый опыт повторяли 2–5 раз. Для обра-

ботки экспериментальных данных использован стандартный набор статистических методов, принятых в современных биологических исследованиях. В работе принят 95%-ный уровень значимости различий, что соответствует принятому стандарту и обеспечивает необходимую степень достоверности полученных результатов.

Рис. 1, а, б, в показывают зависимость генетической нестабильности клеток от дозы ионизирующего излучения для шести штаммов дрожжей дикого типа. Все штаммы демонстрировали сигмоидные формы кривых зависимости выживаемости от дозы. Эти кривые были опубликованы нами ранее при описании влияния генотипа клеток на их радиочувствительность и ее модификацию радиопротекторами и радиосенсибилизаторами [12, 13], поэтому в настоящей работе они не приводятся, а об их радиочувствительности можно судить по значениям LD_{90} (10% выживаемости), а также по относительной радиочувствительности диплоидных и гаплоидных штаммов $LD_{90}(\text{дипл})/LD_{90}(\text{гапл})$ или по относительной радиочувствительности диких и радиочувствительных мутантов $LD_{90}(\text{дикий})/LD_{90}(\text{мутант})$ (табл. 1).

Из полученных результатов ясно, что мутантные диплоидные штаммы более радиочувствительны к облучению ионизирующим излучением по сравнению с клетками родительских штаммов: для *rad51/rad51* мутанта в 6.2 раза, для *rad52/rad52* в 4.2 раза и для *rad54/rad54* в 10 раз – на уровне 10% выживаемости клеток. Для этих штаммов поддержка формирования колоний облученными клетками практически достигала 100% (рис. 1, г, д, е) независимо от их радиочувствительности и спо-

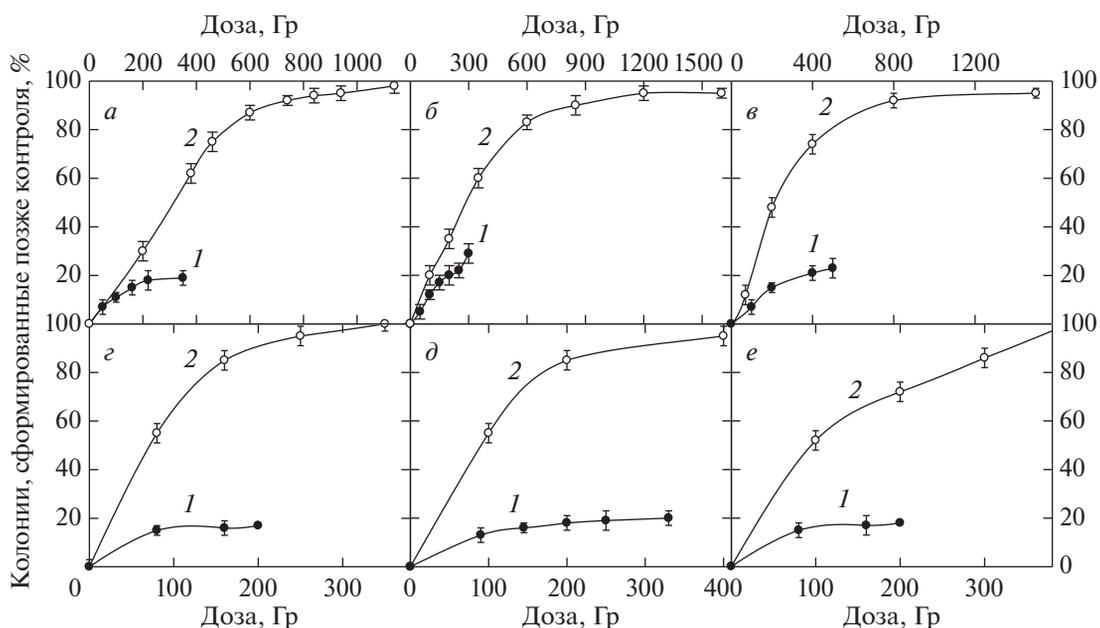


Рис. 1. Кривые зависимости задержанного по сравнению с контролем образования колоний выжившими после облучения клетками от дозы ионизирующего излучения для различных типов клеток. *a* – дрожжевые клетки дикого типа: гаплоидные клетки *Z. baillii* штамм К7 (*RAD*, кривая 1) и диплоидные клетки *S. ellipsoideus*, штамм Мегри 139-В (*RAD/RAD*, кривая 2); *б* – дрожжевые клетки *S. cerevisiae* дикого типа: гаплоидный штамм 211 (*RAD/RAD*, кривая 2) и диплоидный штамм XS800 (*RAD/RAD*, кривая 1); *в* – дрожжевые клетки *S. cerevisiae* дикого типа: гаплоидный штамм S288С (*RAD*, кривая 1), диплоидный штамм XS800 (*RAD/RAD*, кривая 2); *г* – радиочувствительные мутантные дрожжевые клетки *S. cerevisiae*: гаплоидный штамм g218/7а (*rad51*, кривая 1) и изогенный диплоидный штамм XS806 (*rad51/rad51*, кривая 2); *д* – радиочувствительные мутантные дрожжевые клетки *S. cerevisiae*: гаплоидный штамм g160/2d (*rad52*, кривая 1) и изогенный диплоидный штамм XS1898 (*rad52/rad52*, кривая 2); *е* – радиочувствительные мутантные дрожжевые клетки *S. cerevisiae*: гаплоидный штамм g218/7с (*rad54*, кривая 1) и изогенный диплоидный штамм L9 (*rad54/rad54*, кривая 2).

способности клеток восстанавливаться от радиационных повреждений. Однако этот эффект проявлялся заметнее для радиочувствительных мутантов в области меньших доз облучения ионизирующего излучения, чем для штаммов дикого типа. Но главный новый результат настоящего исследования – генетическая нестабильность имела место для радиочувствительных диплоидных мутантов, неспособных к репарации и характеризующихся экспоненциальными кривыми доза-эффект. Генетическая нестабильность гаплоидных штаммов как дикого типа, так и их радиочувствительных мутантов, оставалась близкой к 20%. В отличие от традиционных представлений эти данные указывают на то, что задержанное по сравнению с контролем формирование колоний выжившими после облучения клетками (генетическая нестабильность) определяется главным образом плоидностью клеток и не зависит от формы кривых доза-эффект, их радиочувствительности и способности клеток восстанавливаться от радиационных повреждений.

Анализ молекулярной природы повреждений, ответственных за задержку формирования колоний облученными клетками, находится вне основной задачи данной работы, хотя некоторые

результаты достойны упоминания. Полагают, что митотическая рекомбинация может быть ответственной за нестабильность генома [15, 16]. К такому же эффекту могут приводить мутации [17] и ошибки репарации ДНК [3]. Продемонстрировано, что образование двунитевых разрывов ДНК так же как и повреждения хромосом, могут инициировать задержанную репродуктивную гибель как дрожжевых клеток [18, 19], так и культивируемых клеток млекопитающих [20]. В работе [15] на дрожжах *Pichia pinus* показано, что клетки, утратившие одну или несколько негомологичных хромосом, дают начало колониям, вырастающим позже контроля. В работах [21, 22] приводятся данные, демонстрирующие, что анеуплоидия может приводить к генетической нестабильности клеток различного происхождения. Возможным механизмом проявления задержки формирования колоний могут быть дефекты в хромосомах (делеции, анеуплоидия и пр.) в диплоидных клетках как дикого типа, так и их радиочувствительных мутантов. Эти дефекты не являются летальными, если они произошли только в одной хромосоме, но могут приводить к задержке формирования колоний. Но эти события могут быть летальными для гаплоидных клеток, поэтому у них наблюдается

меньший выход колоний, сформированных позже контроля. Разнообразие возможных механизмов запуска генетической нестабильности показывает необходимость дальнейших исследований этого явления.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Chang W.P., Little J.B.* Evidence that DNA double-strand break initiate the phenotype of delayed reproductive death in Chinese hamster ovary cells // *Radiat. Res.* 1992. V. 131. № 1. P. 53–59.
2. *Brosh R.M., Bohr V.A.* Human premature aging, DNA repair and RecQ helicases // *Nucl. Ac. Res.* 2007. V. 35. № 22. P. 7527–7544. <https://doi.org/10.1093/nar/gkm1008>
3. *Degtyareva N.P., Chen L., Mieczkowski P. et al.* Chronic oxidative DNA damage due to DNA repair defects causes chromosomal instability in *Saccharomyces cerevisiae* // *Mol. Cell. Biol.* 2008. V. 28. № 17. P. 5432–5445. <https://doi.org/10.1128/MCB.00307-08>
4. *Aggarwal M., Brosh R.M.* Functional analysis of human DNA repair proteins important for aging and genomic stability using yeast genetics // *DNA Repair (Amst)*. 2012. V. 11. № 4. P. 335–348. <https://doi.org/10.1016/j.dnarep.2012.01.013>
5. *Donigan K.A., Cerritelli S.M., McDonald J.P. et al.* Unlocking the steric gate of DNA polymerase η leads to increased genomic instability in *Saccharomyces cerevisiae* // *DNA Repair (Amst)*. 2015. V. 35. № 1. P. 1–12. <https://doi.org/10.1016/j.dnarep.2015.07.002>
6. *Капульцевич Ю.Г., Корогодина В.И., Петин В.Г.* Анализ радиобиологических реакций дрожжевых клеток. I. Кривые выживания и эффект дорастания // *Радиобиология.* 1972. Т. 12. № 2. С. 267–271.
7. *Корогодина В.И.* Феномен жизни. Избранные труды. М.: Наука, 2010. Т. 1. 436 с.
8. *Петин В.Г., Белкина С.В., Жураковская Г.П.* Математические модели и реакции клеток на облучение ионизирующими излучениями разного качества. М.: ГЕОС, 2020. 263 с.
9. *Евстратова Е.С., Хрячкова А.В., Жураковская Г.П. и др.* УФ-индуцированная задержка формирования колоний дрожжевыми клетками // *Рад. биология. Радиозэкология.* 2018. Т. 58. № 3. С. 245–250. <https://doi.org/10.7868/S0869803118030037>
10. *Евстратова Е.С., Королев В.Г., Петин В.Г., Толкаева М.С.* Выживаемость и генетическая нестабильность дрожжевых клеток разного генотипа после облучения УФ-светом // *Рад. биология. Радиозэкология.* 2021. Т. 61. № 6. С. 608–614. <https://doi.org/10.31857/S0869803121060035>
11. *Евстратова Е.С., Королев В.Г., Петин В.Г.* Задержка образования колоний диплоидными клетками разного генотипа после облучения УФ-светом // *Генетика.* 2019. Т. 55. № 7. С. 832–836. <https://doi.org/10.1134/S0016675819070063>
12. *Saeki T., Machida I., Nakai S.* Genetic control of diploid recovery after gamma-irradiation in the yeast *Saccharomyces cerevisiae* // *Mut. Res.* 1980. V. 73. № 2. P. 251–265. [https://doi.org/10.1016/0027-5107\(80\)90192-x](https://doi.org/10.1016/0027-5107(80)90192-x)
13. *Petin V.G., Ryabchenko V.I.* A study of oxygen effect and radiosensitization of hypoxic cells by metronidazole in wild-type and radiosensitive mutants of yeast // *Int. J. Radiat. Biol.* 1982. V. 42. P. 491–500.
14. *Petin V.G.* Influence of misonidazole on survival of wild-type and radiosensitive mutants of yeast // *Mut. Res.* 1983. V. 108. P. 121–131.
15. *Толсторуков И.И., Близник К.М., Корогодина В.И.* Митотическая нестабильность диплоидных клеток дрожжей *Pichia pinus*. Сообщение I. Спонтанное расщепление // *Генетика.* 1979. Т. 15. № 12. С. 2140–2147.
16. *Chang W.P., Little J.B.* Persistently elevated frequency of spontaneous mutations in progeny of CHO clones surviving X-irradiation: Association with delayed reproductive death phenotype // *Mut. Res.* 1992. V. 270. № 2. P. 191–199. [https://doi.org/10.1016/0027-5107\(92\)90130-t](https://doi.org/10.1016/0027-5107(92)90130-t)
17. *Marder B.A., Morgan W.F.* Delayed chromosomal instability induced by DNA damage // *Mol. Cell. Biol.* 1993. V. 13. № 11. P. 6667–6677. <https://doi.org/10.1128/mcb.13.11.6667-6677.1993>
18. *Korogodina V.I., Bliznik K.M., Kapultcevic Yu.G. et al.* Cascade mutagenesis: Regularities and mechanisms // *Proc. 2nd Int. N.W. Timofeef-Ressovsky Conference.* Dubna: 2007. V. 1. P. 419–447.
19. *Корогодина В.И., Близник К.М., Капульцевич Ю.Г., Петин В.Г.* Оценка частоты митотической рекомбинации в клонах, вырастающих из облученных клеток // *Радиобиология.* 1974. Т. 14. № 5. С. 681–685.
20. *Sheltzer J.M., Blank H.M., Pfau S.J. et al.* Aneuploidy drives genomic instability in yeast // *Science.* 2011. V. 333. № 6045. P. 1026–1030. <https://doi.org/10.1126/science.1206412>
21. *Urushibara A., Kodama S., Yokoya A.* Induction of genetic instability by transfer of a UV-A-irradiated chromosome // *Mut. Res.* 2014. V. 766. № 1. P. 29–34. <https://doi.org/10.1016/j.mrgentox.2014.02.005>

DNA Repair Does Not Trigger Genetic Instability of Yeast Cells

E. S. Evstratova^{a, *}, V. G. Petin^b, and S. A. Geraskin^c

^aNational Medical Research Center for Radiology, Kaluga region, Obninsk, 249031 Russia

^bTsyb Medical Radiological Research Center – Branch of the National Medical Research Center for Radiology, Kaluga region, Obninsk, 249031 Russia

^cRussian Institute of Radiology and Agroecology, Kaluga region, Obninsk, 249032 Russia

*e-mail: ekevs7240@mail.ru

The dependence of the delayed formation of colonies on the dose of ionizing radiation (gamma rays of ^{60}Co , dose rate 10 Gy/min) was obtained for six strains of wild-type haploid and diploid yeast, capable of recovering from radiation damage and characterized by sigmoidal survival curves as well as for six haploid and diploid radiosensitive mutants defective in reparation and characterized by exponential survival curves. The delay in the formation of colonies by survival cells after irradiation is considered as a genetic instability display. It was shown that for all diploid strains, genetic instability reached 100% with an increase in the dose of ionizing radiation, regardless of the shape of survival curves and the cell ability to recover from radiation damage. Conversely, for all haploid strains, genetic instability was only close to 20%. In contrast to traditional concepts, these data indicate that the late formation of colonies by surviving yeast cells after irradiation is determined mainly by cell ploidy and does not depend on the shape of the dose-response curves and cell radiosensitivity. This means that DNA repair does not trigger genetic instability in yeast cells.

Keywords: genetic instability, gamma rays, yeast cells, haploid and diploid, radiosensitive mutants.

ИДЕНТИФИКАЦИЯ ГИБРИДНЫХ ОСОБЕЙ СКАЛЬНЫХ ЯЩЕРИЦ *Darevskia armeniaca* × *Darevskia valentini* НА ОСНОВЕ МИКРОСАТЕЛЛИТНОГО ГЕНОТИПИРОВАНИЯ

© 2023 г. А. Е. Гирнык¹, *, А. А. Вергун^{1, 2}, А. П. Рысков¹

¹Институт биологии гена Российской академии наук, Москва, 119334 Россия

²Московский педагогический государственный университет, Москва, 119991 Россия

*e-mail: aegirnyk@mail.ru

Поступила в редакцию 06.10.2022 г.

После доработки 16.11.2022 г.

Принята к публикации 22.11.2022 г.

В работе представлены результаты микросателлитного генотипирования гибридных ящериц *D. armeniaca* × *D. valentini* из симпатрических популяций Армении. По совокупности локусов *Du281* и *Du47G* идентифицированы триплоидные особи и определено их происхождение. Показано, что гибриды содержат как аллели материнского партеновида — *D. armeniaca*, так и аллели отцовского вида — *D. valentini*.

Ключевые слова: *Darevskia*, триплоиды, симпатрия, микросателлиты, генотипирование.

DOI: 10.31857/S0016675823060073, **EDN:** SSLMKR

Гибридогенное видообразование лежит в основе теории ретикулярной (сетчатой) эволюции [1]. В ряде случаев этот феномен приводит к образованию однополых, клонально размножающихся с помощью партеногенеза, видов. Среди позвоночных облигатный партеногенез был впервые обнаружен у кавказских скальных ящериц рода *Darevskia*, который сегодня включает 29 двуполых и семь однополых (партеногенетических) диплоидных видов гибридного происхождения [2]. Отличительными особенностями рода *Darevskia* являются как происхождение и большое разнообразие партеновидов, так и продолжающиеся процессы межвидовой гибридизации однополых самок с самцами двуполых видов в зоне частичного перекрытия их ареалов. При этом образующиеся в результате полиплоидные гибридные особи рассматриваются как возможные участники промежуточных этапов сетчатой эволюции в этой группе рептилий [3, 4]. Полиплоидные гибриды также известны и среди многих других родов ящериц: *Aspidoscelis* [5, 6], *Liolaemus* [7], *Hemidactylus* [8]. Гибриды рода *Darevskia* изучаются более 50 лет [9, 10], однако до сих пор накоплено не так много генетических данных, которые бы позволяли оценить вклад таких ящериц в эволюционную историю рода. Молекулярно-генетические исследования близкородственных однополых и двуполых видов, а также их гибридов, могут помочь в изучении различных путей эволюции у рептилий, механизмов формиро-

вания генетического разнообразия и процессов адаптации.

Среди кавказских скальных ящериц гибридные особи были зарегистрированы во многих районах Армении, а также в Нагорном Карабахе и Грузии в зонах симпатричного обитания разных однополых и двуполых видов [11, 12]. Частичное перекрытие ареалов партеногенетического вида *D. armeniaca* и двуполого вида *D. valentini* и появление в них гибридов отмечено в зонах Лчашен, Кучак, Теж, Сотк [3, 11–15], однако известны и другие симпатрические ареалы этих видов в районах: Ахалкалаки, Артаваза, Мец Сепасара и оз. Ханчали [11, 15]. Одной из уникальных смешанных популяций в Армении является популяция Кучак, поскольку в ней образуются полиплоидные гибриды сразу двух партеногенетических видов ящериц *D. armeniaca* и *D. unisexualis*, а в качестве отцовского вида выступает двуполый вид *D. valentini*. По данным исследований 2001–2008 гг. в популяции Кучак процент гибридных особей был особенно высок (свыше 35%) по сравнению с другими симпатрическими популяциями (7–12%) [3, 16]. При этом гибриды *D. armeniaca* × *D. valentini* составляли лишь небольшую часть [3, 13]. Образующиеся в симпатрических зонах триплоидные гибриды разного происхождения объединяют в себе морфологические характеристики обоих родительских видов, поэтому существуют трудности в дифференциации триплоидных ги-

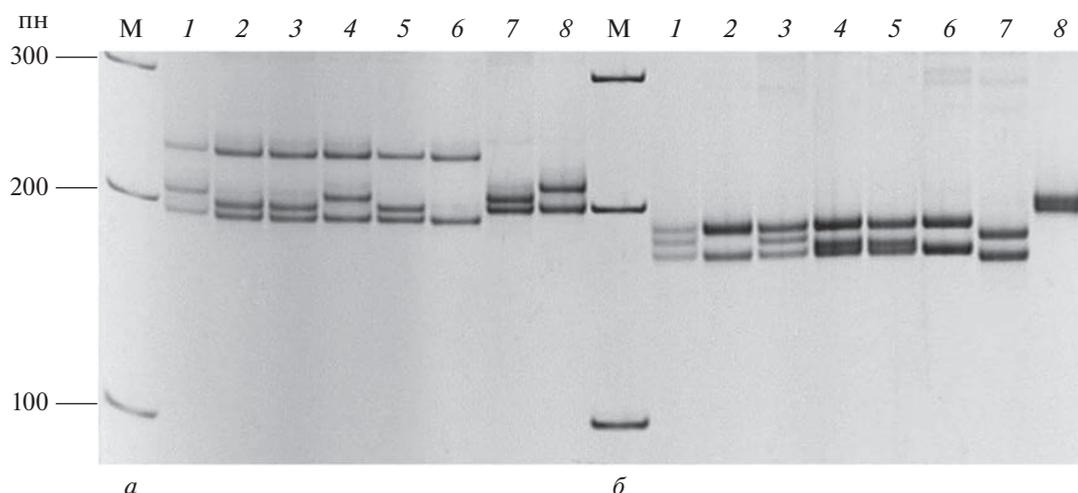


Рис. 1. Картина электрофоретического фракционирования продуктов ПЦР-амплификации микросателлитных локусов: *a* – *Du281*, *б* – *Du47G* в 8%-ном ПААГ. М – маркер молекулярного веса 100 бр (пн); 1–5 – триплоидные особи *D. armeniaca* × *D. valentini*; 6 – контрольная особь партеновида *D. armeniaca*; 7, 8 – контрольные особи родительского (отцовского) вида *D. valentini*.

бридов и их родителей, что требует использования более эффективных методов анализа [3].

В этой связи, в настоящей работе решалась задача по идентификации и дифференциации предполагаемых триплоидных гибридов из двух смешанных популяций, а также по определению происхождения этих полиплоидов. Ранее мы разработали методику молекулярно-генетической идентификации триплоидных гибридов *D. unisexualis* × *D. valentini*, используя для генотипирования микросателлитный локус *Du215* [17]. Однако для идентификации триплоидов *D. armeniaca* × *D. valentini* этот локус оказался мало пригоден. В данном исследовании для идентификации этих гибридов был использован вариант микросателлитного генотипирования на основе сочетания двух локусов ядерного генома (*Du281* и *Du47G*), которые были эффективны в популяционных исследованиях *D. armeniaca* [18].

Биологические образцы для анализа ДНК предположительных триплоидов *D. armeniaca* × *D. valentini* (четыре особи), *D. unisexualis* × *D. valentini* (одна особь) были собраны и определены герпетологом И.С. Даревским в период с 1999 по 2000 гг. в популяциях Теж (Памбакский хребет; 40°42′8.08″ N 44°36′30.80″ E) и Кучак (восток Арагацотнской области; 40°31′49.81″ N 44°17′3.43″ E) Армении. ДНК выделяли из фрагмента хвоста ящериц стандартным фенол-хлороформным методом [19]. Генотипирование проводили по двум микросателлитным локусам ядерного генома (*Du281* и *Du47G*), используя ПЦР-амплификацию с последующим разделением аллелей в 8%-ном полиакриламидном геле (ПААГ) и их полноразмерным секвенированием, как это описано ранее [18, 20].

В настоящей работе были также апробированы микросателлитные локусы *Du215* и *Du323*, однако локусы *Du281* и *Du47G* оказались наиболее эффективными в связи с лучшим разделением аллелей.

По результатам монолокусной ПЦР все особи были гетерозиготными по исследованным локусам и имели по три аллеля, за исключением одной особи, которая имела только два аллеля по локусу *Du47G*. На рис. 1 показано, как разделяются видоспецифичные аллели локусов *Du281* и *Du47G* у этих особей (дорожки 1–5), у контрольной особи партеновида *D. armeniaca* (дорожка 6) и двуполого вида *D. valentini* (дорожки 7–8).

Для проведения видовой идентификации выявленных аллелей локусов *Du281* и *Du47G* было проведено их полноразмерное секвенирование, на основе которого определены аллель-специфические маркеры (микросателлиты и SNPs в прилежащих к микросателлиту областях аллеля). Эти аллель-специфические полиморфные маркеры локусов сравнивались с таковыми, ранее полученными для *D. valentini* и партеновидов *D. armeniaca* и *D. unisexualis* из разных популяций Армении [21, 22]. Сочетание этих маркеров отражает гибридное происхождение триплоидов и позволяет определить видовую принадлежность аллелей. Структурные различия выявленных аллелей исследуемых особей представлены в табл. 1.

Все исследованные гибридные особи по локусу *Du281* (рис. 1, *a*) несли по два аллеля материнского вида *D. armeniaca*: *Du281(arm)2* и *Du281(arm)4* и отличались только по отцовскому аллелю *D. valentini* – у особей 1 и 4 выявлен аллель *Du281(val)2*, а у особей 2, 3, 5 выявлен аллель *Du281(val)3* (табл. 1). Аллель *Du281(arm)4*, общий у триплоидных осо-

Таблица 1. Структурные различия аллельных вариантов локусов *Du281* и *Du47G* исследованных особей

Аллельные варианты локусов	Размер (пн)	Структурные различия микросателлитного кластера	SNP*
<i>Du281</i>	<i>Du281(arm)2</i>	225 5'(GGTA)(GATA) ₉ GAT(GATA) (GGTA) ₂ (GAT)(GATA) ₄ 3'	T (+15)
	<i>Du281(arm)4</i>	183 5'(GATA) ₈ 3'	C (+15)
	<i>Du281(val)2</i>	195 5'(GATA) ₁₁ 3'	C (+15)
	<i>Du281(val)3</i>	191 5'(GATA) ₁₀ 3'	C (+15)
	<i>Du281(val)5</i>	183 5'(GATA) ₈ 3'	C (+15)
<i>Du47G</i>	<i>Du47G(arm)2</i>	184 5'(GATA) ₉ (GACA) ₄ (GATA) ₂ GAT(GATA) ₂ 3'	T (+7), A (+21), G (+52), T (+56)
	<i>Du47G(arm)3</i>	176 5'(GATA) ₁₁ (GACA)(GATA) GAT(GATA) ₂ 3'	T (+7), T (+21), C (+52), A (+56)
	<i>Du47G(arm)4</i>	172 5'(GATA) ₁₀ (GACA)(GATA) GAT(GATA) ₂ 3'	T (+7), T (+21), C (+52), A (+56)
	<i>Du47G(arm)5</i>	168 5'(GATA) ₉ (GACA)(GATA) GAT(GATA) ₂ 3'	T (+7), T (+21), C (+52), A (+56)
	<i>Du47G(val)8</i>	176 5'(GATA) ₁₁ (GACA)(GATA) GAT(GATA) ₂ 3'	T (+7), T (+21), C (+52), A (+56)
	<i>Du47G(val)9</i>	168 5'(GATA) ₉ (GACA)(GATA) GAT(GATA) ₂ 3'	T (+7), T (+21), C (+52), A (+56)

Примечание. * – в скобках даны расстояния в парах оснований: (+) – после микросателлитного кластера. Нумерация аллелей локусов *Du281* и *Du47G* для *D. armeniaca* и *D. valentini* дана как в статье [21] и как они депонированы в GenBank.

бей, совпадал с аллелем *Du281(val)5*, ранее выявленным в популяциях двуполого вида *D. valentini*, что говорит о гибридном происхождении как триплоидных особей, так и самого партеновида *D. armeniaca*, возникшего путем межвидовой гибридизации двуполого вида *D. mixta*, выступающего в качестве материнского и двуполого вида *D. valentini*, выступающего в качестве отцовского [21].

По локусу *Du47G* (рис. 1,б) две особи (1 и 3) несли аллели *D. armeniaca Du47G(arm)2*, *Du47G(arm)3*, *Du47G(arm)5*. Аллели *Du47G(arm)3* и *Du47G(arm)5* совпадали с аллелями *D. valentini Du47G(val)8* и *Du47G(val)9*. Две другие особи (4 и 5) также несли аллели *Du47G(arm)2* и *Du47G(arm)5*, но отличались наличием аллеля *Du47G(arm)4*. Особь 2, по результатам ПЦР-анализа по локусу *Du47G* имела только два аллеля, также общие для гибридов *Du47G(arm)2* и *Du47G(arm)5*. Эти данные могут говорить о том, что у этой особи два аллеля, наследуемые от партеновида *D. armeniaca* и двуполого вида *D. valentini* идентичны, а третий аллель, наследуемый также от *D. armeniaca*, отличался как по подвижности в ПААГ, так и по нуклеотидной последовательности. Такой факт был выявлен ранее для локуса *Du215*, который был использован для идентификации триплоидных гибридов *D. uni-*

sexualis × *D. valentini* [17], но оказался недостаточно эффективным в отношении гибридов *D. armeniaca* × *D. valentini* из-за идентичности “отцовских” аллелей *D. valentini*. Один из образцов (3) фенотипически был определен как гибрид *D. unisexualis* × *D. valentini*. Однако по картине электрофоретического фракционирования аллелей используемых локусов и по результатам их секвенирования этот образец мы отнесли к гибриду *D. armeniaca* × *D. valentini*. О возможных ошибочных идентификациях предполагаемых гибридов ранее уже сообщалось в литературе [13].

Таким образом, приведенный нами вариант микросателлитного генотипирования по совокупности локусов *Du281* и *Du47G* позволил идентифицировать триплоидных особей, а секвенирование выявленных аллельных вариантов – определить происхождение гибридов *D. armeniaca* × *D. valentini*. Дополнительное использование для генотипирования локуса *Du215* позволяет достоверно различить триплоидов разного происхождения.

Авторы благодарны И.С. Даревскому за предоставление биологических образцов. А также И.А. Мартиросян за ценные советы при подготовке рукописи.

Исследование выполнено при поддержке гранта Российского научного фонда № 19-14-00083.

Работу с животными проводили с соблюдением стандартов, разработанных Министерством защиты природы Армении (код разрешения 5/22.1/51043). Исследование одобрено этическим комитетом МГУ (номер разрешения 24-01) и проведено в строгом соответствии с этическими принципами и научными стандартами.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Dobzhansky Th. Genetic nature of species differences // Am. Natur. 1937. V. 71. № 735. P. 404–420. <https://doi.org/10.1086/280726>
2. Uetz P., Freed P., Aguilar R., Hošek J. The Reptile Database (Online resource) // URL: <http://www.reptile-database.org> (accessed 10.10.2022)
3. Danielyan F., Arakelyan M., Stepanyan I. The progress of microevolution in hybrids of rock lizards of genus *Darevskia* // Biol. Z. Armen. 2008. V. 60. № 1–2. P. 147–156.
4. Spangenberg V., Arakelyan M., Galoyan E. et al. Reticulate evolution of the rock lizards: Meiotic chromosome dynamics and spermatogenesis in diploid and triploid males of the genus *Darevskia* // Genes. 2017. V. 8. № 6. P. 149. <https://doi.org/10.3390/genes8060149>
5. Moritz C.C., Wright J.W., Brown W.M. Mitochondrial DNA analyses and the origin and relative age of parthenogenetic lizards (genus *Cnemidophorus*). III. *C. velox* and *C. exsanguis* // Evolution. 1989. V. 43. № 5. P. 958–968. <https://doi.org/10.1111/j.1558-5646.1989.tb02542.x>
6. Taylor H.L., Walker J.M., Cole C.J., Dessauer H.C. Morphological divergence and genetic variation in the triploid parthenogenetic teiid lizard, *Aspidoscelis neotesselata* // J. Herp. 2015. V. 49. № 3. P. 491–501. <https://doi.org/10.1670/14-057>
7. Araya-Donoso R., Véliz D., Vidal M., Lamborot M. Relationships of the morphological variation in diploids, triploids and mosaics of *Liolaemus chiliensis* (Sauria: Liolaemidae) // Amphib. Reptil. 2017. V. 38. № 4. P. 503–515. <https://doi.org/10.1163/15685381-00003132>
8. Kluge A.G., Eckardt M.J. *Hemidactylus garnotii* Dumeril and Bibron, a triploid all-female species of gekkonid lizard // Copeia. 1969. № 4. P. 651–664. <https://doi.org/10.2307/1441789>
9. Даревский И.С., Куликова В.Н. Систематические признаки и некоторые особенности оогенеза гибридов между обоеполой и партеногенетическими формами скальной ящерицы // Цитология. 1962. Т. 4. № 2. С. 160–170.
10. Darevsky I.S., Danielyan F.D. Diploid and triploid progeny arising from natural mating of parthenogenetic *Lacerta armeniaca* and *L. unisexualis* with bisexual *L. saxicola valentini* // J. Herp. 1968. V. 2. № 3–4. P. 65–69. <https://doi.org/10.2307/1563104>
11. Petrosyan V.G., Osipov F.A., Bobrov V.V. et al. New records and geographic distribution of the sympatric zones of unisexual and bisexual rock lizards of the genus *Darevskia* in Armenia and adjacent territories // Biodivers. Data J. 2020. V. 8. P. 1–30. <https://doi.org/10.3897/BDJ.8.e56030>
12. Arakelyan M.S., Danielyan F.D., Corti C. et al. Herpetofauna of Armenia and Nagoro-Karabakh. Salt Lake City: Society for Study of Amphibians and Reptiles, 2011. 149 p.
13. Freitas S.N., Harris D.J., Sillero N. et al. The role of hybridisation in the origin and evolutionary persistence of vertebrate parthenogens: a case study of *Darevskia* lizards // Heredity. 2019. V. 123. № 6. P. 795–808. <https://doi.org/10.1038/s41437-019-0256-5>
14. Carretero M., García-Muñoz E., Argaña E. et al. Parthenogenetic *Darevskia* lizards mate frequently if they have the chance: A quantitative analysis of copulation marks in a sympatric zone // J. Nat. Hist. 2018. V. 52. № 7–8. P. 405–413. <https://doi.org/10.1080/00222933.2018.1435832>
15. Galoyan E., Bolshakova A., Abrahamyan M. et al. Natural history of Valentin's rock lizard (*Darevskia valentini*) in Armenia // Zool. Res. 2019. V. 40. № 4. P. 277–292. <https://doi.org/10.24272/j.issn.2095-8137.2019.036>
16. Даревский И.С., Даниелян Ф.Д. Естественная аллотриплоидия у некоторых видов скальных ящериц Кавказа и вероятное эволюционное значение этого феномена // Эволюция, экология, видообразование. Материалы конференции памяти Н.Н. Воронцова (1934–2000). М.: УНЦ ДО, 2001. С. 131–133.
17. Гирнык А.Е., Вергун А.А., Аракелян М.С., Рысков А.П. Молекулярно-генетическая идентификация триплоидных гибридов *Darevskia unisexualis* × *D. valentini* из симпатрических популяций Армении // Мол. генетика, микробиология и вирусология. 2021. Т. 39. № 3. С. 46–48. <https://doi.org/10.17116/molgen20213903146>
18. Гирнык А.Е., Вергун А.А., Омельченко А.В. и др. Молекулярно-генетическая характеристика аллельных вариантов микросателлитных локусов *Du215*, *Du281*, *Du323* и *Du47G* у партеногенетических ящериц *Darevskia armeniaca* (сем. Lacertidae) // Генетика. 2017. Т. 53. № 4. С. 468–479.
19. Mathew C.G.P. The isolation of high molecular weight eukaryotic DNA // Nucleic Acids. Methods in Molecular Biology. United States: Humana Press, 1984. V. 2. P. 31–34. <https://doi.org/10.1385/0-89603-064-4:31>
20. Korchagin V.I., Badaeva T.N., Tokarskaya O.N. et al. Molecular characterization of allelic variants of (GATA)n microsatellite loci in parthenogenetic lizards *Darevskia unisexualis* (Lacertidae) // Gene. 2007. V. 392. № 1–2. P. 126–133. <https://doi.org/10.1016/j.gene.2006.11.020>
21. Girnyk A.E., Vergun A.A., Semyenova S.K. et al. Multiple interspecific hybridization and microsatellite mutations provide clonal diversity in the parthenogenetic rock lizard *Darevskia armeniaca* // BMC Genomics. 2018. V. 19. № 1. P. 1–12. <https://doi.org/10.1186/s12864-018-5359-5>
22. Vergun A.A., Girnyk A.E., Korchagin V.I. et al. Origin, clonal diversity, and evolution of the parthenogenetic lizard *Darevskia unisexualis* // BMC Genomics. 2020. V. 21. № 351. P. 1–10. <https://doi.org/10.1186/s12864-020-6759-x>

**Identification of Hybrid Specimens of Rock Lizards
Darevskia armeniaca × *Darevskia valentini* Based on Microsatellite Genotyping**

A. E. Girnyk^{a, *}, A. A. Vergun^{a, b}, and A. P. Ryskov^a

^a*Institute of Gene Biology Russian Academy of Sciences, Moscow, 119334 Russia*

^b*Moscow Pedagogical State University, Moscow, 119991 Russia*

**e-mail: aegirnyk@mail.ru*

The paper presents the results of microsatellite genotyping of hybrid lizards *D. armeniaca* × *D. valentini* from sympatric populations of Armenia. Triploid individuals were identified based on the totality of the *Du281* and *Du47G* loci and their origin was determined. It was shown that the hybrids contain both alleles of the maternal parthenospecies *D. armeniaca* and alleles of the paternal species *D. valentini*.

Keywords: *Darevskia*, triploids, sympatry, microsatellites, genotyping.

ГЕНЕТИЧЕСКАЯ ВАРИАБЕЛЬНОСТЬ ЛОКУСА ГЕНА *MAOA* У АГРЕССИВНЫХ ЖИВОТНЫХ НЕКАНОНИЧЕСКОЙ ПОВЕДЕНЧЕСКОЙ МОДЕЛИ *Neogale vison*¹

© 2023 г. А. Д. Манахов^{1, 2, 3, *}, Н. А. Дудко^{1, 3}, Ф. Е. Гусев^{1, 3},
Т. В. Андреева^{1, 2, 3}, О. В. Трапезов^{4, 5}, Е. И. Рогаев^{3, 6, **}

¹Центр генетики и наук о жизни, “Научно-технологический университет “Сириус”,
Краснодарский край, пгт. Сириус, 354340 Россия

²Центр генетики и генетических технологий, Московский государственный университет
им. М.В. Ломоносова, Москва, 119234 Россия

³Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

⁴Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук, Новосибирск, 630090 Россия

⁵Новосибирский государственный университет, Новосибирск, 630039 Россия

⁶Медицинская школа Чан Массачусетского университета, департамент психиатрии, Шрусбери, 01545 США

*e-mail: manakhov@rogaevlab.ru

**e-mail: rogaev@vigg.ru

Поступила в редакцию 09.12.2022 г.

После доработки 26.12.2022 г.

Принята к публикации 19.01.2023 г.

Роль гена *MAOA* в процессах регуляции агрессивного поведения у человека и животных широко известна. В рамках данного исследования мы впервые провели анализ генетической варибельности гена *MAOA* и его промоторной области у представителей неканонической поведенческой модели – американской норки (*Neogale vison*). Проведенный анализ не выявил значимых генетических изменений животных из линии, характеризующейся агрессивным поведением, что позволяет предположить наличие генетических и/или эпигенетических вариаций в других системах, связанных с регуляцией агрессии, у данной модели.

Ключевые слова: американская норка, *Neogale vison*, агрессивное поведение, *MAOA*.

DOI: 10.31857/S0016675823060097, **EDN:** SSRBNW

Ген *MAOA* – один из важнейших и наиболее изученных генов, вовлеченных в процессы, лежащие в основе агрессивного поведения у человека и животных, за что в литературе данный ген часто называют “геном война” или “геном убийцы”. Данный ген кодирует фермент моноаминоксидазу А – белок, который связывается с мембранами митохондрий и катализирует окислительное дезаминирование моноаминовых нейротрансмиттеров, в том числе серотонина, норадреналина и дофамина, выполняющих важнейшую роль в регуляции эмоционального и социального поведения [1]. Преимущественно моноаминоксидаза А обнаруживается в катехоламинергических нейронах головного мозга [2].

Впервые участие гена *MAOA* в регуляции агрессивного и антисоциального поведения у человека было показано при исследовании родословной крупной голландской семьи, мужчины из которой имели склонность к насилию и вспышками агрессии, которые проявлялись в попытках убийств, изнасилования и поджогах (синдром Бруннера). У представителей этой родословной была выявлена нонсенс-замена в гене *MAOA* (с.886 C>T, p.Gln296*), а пациенты с мутацией характеризовались значительным снижением уровня метаболитов серотонина и катехоламинов [3, 4]. Проведенное нокаутирование гена *MAOA* у мышей приводит к схожим поведенческим и биохимическим изменениям [5]. Более поздние исследования выявили и другие мутации в гене *MAOA* у человека (например, p.Arg45Trp и p.Cys266Phe), которые также приводят к снижению активности фермента, нарушению когнитивного развития и поведенческим изменениям, характерным для рас-

¹ Дополнительная информация для этой статьи доступна по doi 10.31857/S0016675823060097 для авторизованных пользователей.

Таблица 1. Описание образцов и результатов секвенирования геномов американской норки

№	Образец	Коэффициент поведения	Окраска	% картирования	% дубликатов	Глубина покрытия	Ссылка
Агрессивные							
1	mink_0-541	−2	с.т.к.	98.62	13.73	19.50	Данная работа
2	mink_0-543	−2	То же	97.17	30.58	13.28	То же
3	mink_3-453	−2	«	99.25	2.56	8.40	«
4	mink_3-473	−2	«	98.50	2.37	8.12	«
5	mink_0-529	−2	«	97.77	15.09	7.66	«
6	mink_3-193	−2	«	99.45	2.42	6.43	«
7	mink_1-371	−3	«	98.21	40.77	6.34	«
8	mink_3-457	−2	«	98.72	2.96	5.96	«
9	mink_3-191	−2	«	99.45	2.72	5.91	«
10	mink_8-67	−2	«	98.98	2.76	5.60	«
11	mink_3-227	−2	«	99.16	13.15	5.36	«
Неселектированные							
12	mink_7-317	Нет данных	<i>a/a m/m p/p</i>	99.60	7.40	37.21	[14]
13	mink_4-523	«	<i>Sh/+ p/p</i>	99.70	8.87	21.52	[15]
14	mink_3-261	«	с.т.к.	99.06	3.14	8.14	[13]
15	mink_3-265	«	с.т.к.	99.33	1.53	7.65	«
16	mink_7-331	«	<i>m/m</i>	99.49	0.30	7.30	[14]
17	mink_3-247	«	с.т.к.	99.23	1.85	6.24	[13]
18	mink_0-329	«	<i>p/p</i>	99.22	2.32	4.82	«
19	mink_9-431	«	<i>p/p</i>	99.04	3.48	4.64	«

Примечание. Статистика рассчитана для данных секвенирования, картированных относительно референсного генома американской норки (GCF_020171115.1), с.т.к. – стандартная, темно-коричневая окраска; все приведенные образцы – от особей мужского пола.

стройств аутистического спектра [6, 7], нокаутные мыши по аналогии характеризуются основными нарушениями, наблюдаемыми при расстройствах аутистического спектра [8].

Связь между генетическими вариантами в *MAOA* и агрессивным поведением была выявлена и у других животных. Так, у быков породы Лидийский боевой бык (*toro de lidia*), специально выведенных для участия в испанской корриде и характеризующихся высоким уровнем агрессии, в промоторной области гена *MAOA*, в предлагаемом сайте связывания транскрипционного фактора Sp1, было выявлено снижение экспансии нуклеотида C (g.105,462,494delsinsC) по сравнению с домашними породами крупного рогатого скота [9]. У свиней также была показана ассоциация четырех сцепленных однонуклеотидных полиморфизмов в гене *MAOA*, один из которых локализован в промоторной области, с агрессивным поведением [10].

Изменения в уровнях моноаминовых нейротрансмиттеров в головном мозге были обнаружены и у сформированных в ходе искусственного отбора агрессивных и ручных линий лис (*Vulpes*

vulpes), американской норки (*Neogale vison*) и крыс (*Rattus norvegicus*), полученных в Институте цитологии и генетики РАН (ИЦиГ СО РАН) [11].

В рамках данного исследования мы впервые провели изучение генетической вариабельности локуса гена *MAOA* у американской норки с агрессивным поведением. В работе были использованы биологические образцы, полученные от самцов американской норки стандартной темно-коричневой окраски, которые содержались на экспериментальной звероферме ИЦиГ СО РАН и подверглись отбору на агрессивное поведение, все исследуемые особи были охарактеризованы по оборонительной реакции на человека [12].

Для 11 агрессивных норок были получены данные полногеномного секвенирования на платформе Illumina HiSeq 2000/2500, в качестве группы сравнения использовались геномы восьми норок, не подвергавшихся отбору по поведению, полученные нами ранее (табл. 1) [13–15]. Полученные данные геномного секвенирования с помощью программы BWA mem [16] были картированы относительно референсного генома самки американской норки (GCF_020171115.1). Генетические вари-

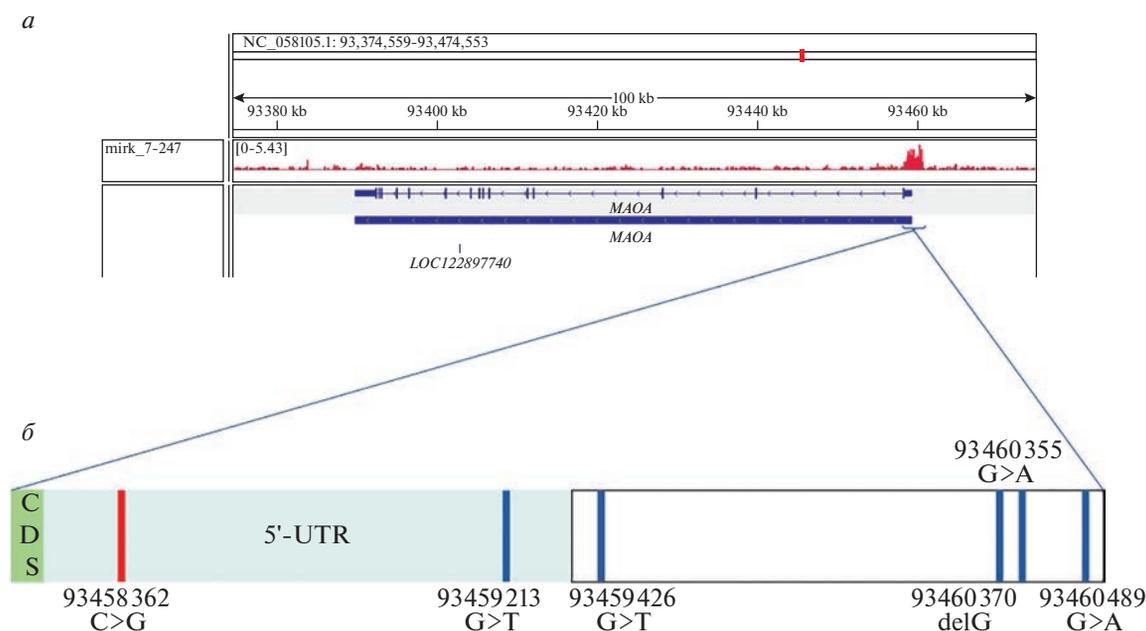


Рис. 1. Генетическая вариабельность промоторной области гена *MAOA*. *a* – профиль секвенирования хроматина, иммунопреципитированного с антителами к H3K4me3, из фронтальной коры головного мозга агрессивной норки; *б* – генетические варианты, обнаруженные в промоторной области гена *MAOA*; красным цветом отмечен вариант, встречающийся только в группе агрессивных норок (частота аллеля, характерного для агрессивных животных – 0.222); синим цветом отмечены варианты, встречающиеся в обеих группах норок.

анты в области гена *MAOA* были выявлены с помощью функции HaplotypeCaller (*-sample-ploidy 1*) из пакета программ GATK v4 [17] и анализировались в программе PLINK [18].

Для одного образца фронтальной коры головного мозга агрессивной норки были получены данные секвенирования хроматина, иммунопреципитированного с антителами к H3K4me3. Полученные данные с помощью программы BWA mem были картированы относительно референсного генома американской норки и нормализованы на общее число прочтений с помощью программы DeepTools [19]. Определение локусов обогащения проводили с помощью программ MACS2 [20] и BEDTools [21].

Ген *MAOA* американской норки, как и у других млекопитающих, локализован на X-хромосоме и, согласно аннотации генома, имеет координаты (NC_058105.1:93,389,771-93,459,335). Основываясь на сравнительном анализе канонической, самой протяженной по белковому продукту, изоформы транскрипта *MAOA* человека (ENST00000338702, Ensembl v105), мы несколько скорректировали координаты кодирующей области первого экзона гена *MAOA* норки (исходные координаты – NC_058105.1:93,458,122-93,459,335; обновленные координаты – NC_058105.1:93,458,122-93,458,194). По результатам секвенирования хроматина, иммунопреципитированного с антителами к H3K4me3 у норки с агрессивным поведением, были опреде-

лены предполагаемые координаты промоторного региона гена *MAOA* – NC_058105.1:93,458,194-93,460,530 (рис. 1*a, б*).

В общей сложности, по сравнению с референсным геномом, в исследованной выборке норок в области гена *MAOA* и его промоторном регионе было выявлено 277 генетических вариантов, из них ни один не попал в кодирующую область гена, 271 локализованы в интронах и 6 попали в промоторную область (Приложение 1). Анализ ассоциации выявленных вариантов с поведением, проведенный с помощью точного теста Фишера в программе PLINK, не выявил ни одного статистически значимого, с учетом поправки на множественные тестирование, генетического варианта (Приложение 2).

Таким образом, полученные нами результаты демонстрируют отсутствие значимых генетических изменений как в кодирующей, так и в промоторной области гена *MAOA* у американских норок из линии, характеризующейся агрессивным поведением. Это позволяет предположить, что агрессивное поведение у этих животных обусловлено генетическими и/или эпигенетическими вариациями в других системах, связанных с регуляцией агрессии, что делает перспективным дальнейшее изучение данной модели для поиска молекулярно-генетических механизмов формирования агрессивного поведения.

Работа выполнена при финансовой поддержке Российского научного фонда в рамках научного проекта № 19-75-30039.

Все применимые международные, национальные и институциональные принципы ухода и использования животных были соблюдены. Исследования людей в качестве объекта не проводились.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Kolla N.J., Bortolato M.* The role of monoamine oxidase A in the neurobiology of aggressive, antisocial, and violent behavior: A tale of mice and men // *Prog. Neurobiol.* 2020. V. 194. P. 101875. <https://doi.org/10.1016/j.pneurobio.2020.101875>
2. *Nelson R.J., Trainor B.C.* Neural mechanisms of aggression // *Nat. Rev. Neurosci.* 2007. V. 8. № 7. P. 536–546. <https://doi.org/10.1038/nrn2174>
3. *Brunner H.G., Nelen M.R., van Zandvoort P. et al.* X-linked borderline mental retardation with prominent behavioral disturbance: Phenotype, genetic localization, and evidence for disturbed monoamine metabolism // *Am. J. Hum. Genet.* 1993. V. 52. № 6. P. 1032–1039.
4. *Brunner H.G., Nelen M., Breakefield X.O. et al.* Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A // *Science.* 1993. V. 262. № 5133. P. 578–580. <https://doi.org/10.1126/science.8211186>
5. *Cases O., Seif I., Grimsby J. et al.* Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking MAOA // *Science.* 1995. V. 268. № 5218. P. 1763–1766. <https://doi.org/10.1126/science.7792602>
6. *Palmer E.E., Leffler M., Rogers C. et al.* New insights into Brunner syndrome and potential for targeted therapy // *Clin. Genet.* 2016. V. 89. № 1. P. 120–127. <https://doi.org/10.1111/cge.12589>
7. *Piton A., Poquet H., Redin C. et al.* 20 ans après: A second mutation in MAOA identified by targeted high-throughput sequencing in a family with altered behavior and cognition // *Eur. J. Hum. Genet.* 2014. V. 22. № 6. P. 776–783. <https://doi.org/10.1038/ejhg.2013.243>
8. *Bortolato M., Godar S.C., Alzghoul L. et al.* Monoamine oxidase A and A/B knockout mice display autistic-like features // *Int. J. Neuropsychopharmacol.* 2013. V. 16. № 4. P. 869–888. <https://doi.org/10.1017/S1461145712000715>
9. *Eusebi P.G., Sevane N., Cortés O. et al.* Aggressive behavior in cattle is associated with a polymorphism in the MAOA gene promoter // *Anim. Genet.* 2020. V. 51. № 1. P. 14–21. <https://doi.org/10.1111/age.12867>
10. *Chen R., Chu Q., Shen C. et al.* Identification of single nucleotide polymorphisms in porcine MAOA gene associated with aggressive behavior of weaned pigs after group mixing // *Animals (Basel).* 2019. V. 9. № 11. P. 952. <https://doi.org/10.3390/ani9110952>
11. *Kulikov A.V., Bazhenova E.Y., Kulikova E.A. et al.* Interplay between aggression, brain monoamines and fur color mutation in the American mink // *Genes, Brain and Behavior.* 2016. V. 15. № 8. P. 733–740. <https://doi.org/10.1111/gbb.12313>
12. *Трапезов О.В.* Гомологические ряды изменчивости окраски меха у американской норки (*Mustela vison* Schreber, 1777) в условиях domestikации // *Вестник ВОГиС.* 2007. Т. 11. № 3/4. С. 547–560.
13. *Manakhov A.D., Andreeva T.V., Trapezov O.V. et al.* Genome analysis identifies the mutant genes for common industrial Silverblue and Hedlund white coat colours in American mink // *Sci. Reports.* 2019. V. 9. № 1. P. 4581. <https://doi.org/10.1038/s41598-019-40918-7>
14. *Manakhov A.D., Mintseva M.Y., Andreev I.A. et al.* Genome analysis of American minks reveals link of mutations in Ras-related protein-38 gene to Moyle brown coat phenotype // *Sci. Reports.* 2020. V. 10. № 1. P. 15876. <https://doi.org/10.1038/s41598-020-72239-5>
15. *Manakhov A.D., Mintseva M.Y., Andreeva T.V. et al.* Shadow coat colour in American mink associated with a missense mutation in the KIT gene // *Animal Genetics.* 2022. V. 53. № 4. P. 522–525. <https://doi.org/10.1111/age.13202>
16. *Li H., Durbin R.* Fast and accurate short read alignment with Burrows-Wheeler transform // *Bioinformatics.* 2009. V. 25. № 14. P. 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
17. *McKenna A., Hanna M., Banks E. et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data // *Genome Research.* 2010. V. 20. № 9. P. 1297–1303. <https://doi.org/10.1101/gr.107524.110>
18. *Purcell S., Neale B., Todd-Brown K. et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses // *Am. J. Hum. Genetics.* 2007. V. 81. № 3. P. 559–575. <https://doi.org/10.1086/519795>
19. *Ramírez F., Ryan D.P., Grüning B. et al.* DeepTools2: A next generation web server for deep-sequencing data analysis // *Nucl. Ac. Res.* 2016. V. 44. № W1. P. 160–165. <https://doi.org/10.1093/nar/gkw257>
20. *Zhang Y., Liu T., Meyer C.A. et al.* Model-based analysis of ChIP-Seq (MACS) // *Genome Biol.* 2008. V. 9. № 9. P. R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
21. *Quinlan A.R., Hall I.M.* BEDTools: A flexible suite of utilities for comparing genomic features // *Bioinformatics (Oxford, England).* 2010. V. 26. № 6. P. 841–842. <https://doi.org/10.1093/bioinformatics/btq033>

Genetic Variability of *MAOA* Gene among Aggressive Animals from the Non-Canonical Behavioral Model *Neogale vison*

A. D. Manakhov^{a, b, c, *}, N. A. Dudko^{a, c}, F. E. Gusev^{a, c}, T. V. Andreeva^{a, b, c},
O. V. Trapezov^{d, e}, and E. I. Rogaev^{c, f, **}

^aCenter for Genetics and Life Science, “Sirius University of Science and Technology”,
Krasnodar region, pgt. Sirius, 354340 Russia

^bCenter for Genetics and Genetic Technologies, Lomonosov Moscow State University, Moscow, 119234 Russia

^cVavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia

^dFederal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian
Academy of Sciences, Novosibirsk, 630090 Russia

^eNovosibirsk State University, Novosibirsk, 630039 Russia

^fDepartment of Psychiatry, UMass Chan Medical School, Worcester, MA, 01545 USA

*e-mail: manakhov@rogaevlab.ru

**e-mail: rogaev@vigg.ru

The *MAOA* gene is widely known regulator of aggressive behavior among human and animals. Here, we analyzed the genetic variability of the *MAOA* gene and its promoter region in non-canonical behavioral model – American mink (*Neogale vison*). We didn't observe any significant genetic variations among animals with aggressive behavior, that suggests the presence of genetic and/or epigenetic variations in other systems involved in regulation of aggression in this model.

Keywords: American mink, *Neogale vison*, aggressive behavior, *MAOA*.