СОДЕРЖАНИЕ

Том 57, номер 12, 2021

В номер вошли статьи, в которых представлены основные результаты реализации научно-технической программы Союзного государства “ДНК-идентификация”

Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства

А. В. Кильчевский, Н. К. Янковский 1345

Геногеографический атлас ДНК-маркеров, контролирующих цвет глаз и волос человека

Е. В. Балановская, И. О. Горин, С. М. Кошель, О. П. Балановский 1356

Определение хронологического возраста человека по биологическим образцам на основании анализа метилирования CpG-динуклеотидов

В. А. Лемеш, В. Н. Кипень, М. В. Богданова, А. А. Буракова, А. Г. Булгак, А. В. Байда, С. А. Брускин, О. В. Зотова, О. И. Добыш 1376

Филогенетический анализ южносибирской гаплогруппы Q-YP1102 по данным о Y-SNP- и Y-STR-маркерах у тувинцев и окружающих популяций

А. Т. Агджоян, Л. Д. Дамба, В. М. Гурьянов, В. В. Запорожченко, О. П. Балановский 1386

Генетическое разнообразие 21 аутосомного STR-маркера системы CODIS в популяциях Восточной Европы

В. Н. Харьков, С. А. Котова, Н. А. Колесников, Е. А. Спивак, К. В. Вагайцева, Т. В. Забавская, В. И. Рыбакова, А. С. Парфенова, А. Н. Верчук, Н. К. Янковский, В. А. Степанов 1396

Реконструкция происхождения гыданских ненцев на основе генетического анализа их родовой структуры с помощью нового набора YSTR-маркеров

В. Н. Харьков, Л. В. Валихова, Е. Л. Яковлева, В. Н. Сереброва, Н. А. Колесников, Т. И. Петелина, И. Ю. Хитринская, В. А. Степанов 1403

Ассоциация метилирования генов нейромедиаторных систем мозга с психоэмоциональными характеристиками человека

И. Б. Моссэ, Н. Г. Седляр, А. С. Бабенко, К. А. Моссэ, Р. С. Шулинский, А. В. Кильчевский 1415

Использование масштабного параллельного секвенирования в криминалистике: сравнительный анализ платформ для секвенирования

Т. В. Тяжелова, И. Л. Кузнецова, Т. В. Андреева, С. С. Кунижева, Е. И. Рогаев 1423

Генетико-демографические параметры населения Москвы. Миграционные процессы

О. Л. Курбатова, А. С. Грачева, Е. Ю. Победоносцева, И. Г. Удина 1438

Полногеномное исследование связи метилирования CpG-сайтов с агрессивным поведением

С. А. Боринская, А. В. Рубанович, А. К. Ларин, А. В. Казанцева, Ю. Д. Давыдова, Э. В. Генерозов, Э. К. Хуснутдинова, Н. К. Янковский 1450

Адаптивные изменения генов десатурации жирных кислот у коренного населения Северо-Востока Сибири

Б. А. Малярчук, М. В. Деренко, Г. А. Денисова 1458

Contents

Vol. 57, No. 12, 2021

Reviews and Theoretical Articles

Developing the Innovative Gene Geographical and Genomic Technologies for Identification and Revealing the Personal Features by Studying the Gene Pools of the Regional Populations

A. V. Kilchevsky and N. K. Yankovsky 1345

Gene Geographic Atlas of DNA Markers Controlling the Color of Human Eyes and Hair

E. V. Balanovska, I. O. Gorin, S. M. Koshel, and O. P. Balanovsky 1356

Determination of Human Chronological Age from Biological Samples Based on the Analysis of CpG-Dinucleotides Methylation

V. A. Lemesh, V. N. Kipen, M. V. Bahdanava, A. A. Burakova, A. G. Bulgak, A. V. Bayda, S. A. Bruskin, O. V. Zotova, and O. I. Dobysh 1376

Phylogenetic Analysis of the South Siberian Haplogroup Q-YP1102 from Y-SNP and Y-STR-Data on Tuvinian and Surrounding Populations

A. T. Agdzhoyan, L. D. Damba, V. M. Gurianov, V. V. Zaporozhchenko, and O. P. Balanovsky 1386

Genetic Diversity of 21 Autosomic STR-Markers of the Codis System in Populations of Eastern Europe

V. N. Kharkov, S. A. Kotova, N. A. Kolesnikov, E. A. Spivak, K. V. Vagaitseva, T. V. Zabavskaya, V. I. Rybakova, A. S. Parfenova, A. N. Verchuk, N. K. Yankovsky, and V. A. Stepanov 1396

Reconstruction of the Origin of the Gydan Nents Based on Genetic Analysis of Their Tribal Structure Using a New Set of YSTR-Markers

V. N. Kharkov, L. V. Valikhova, E. L. Yakovleva, V. N. Serebrova, N. A. Kolesnikov, T. I. Petelina, I. Yu. Khitrinskaya, and V. A. Stepanov 1403

Association of Methylation of Neuromediator Brain System Genes with Psychoemotional Human Characteristics

I. B. Mosse, N. G. Sedlyar, A. S. Babenko, K. A. Mosse, R. S. Shulinsky, and A. V. Kilchevsky 1415

Using NGS in Forensics: A Comparative Analysis of Sequencing Platforms

T. V. Tyazhelova, I. L. Kuznetsova, T. V. Andreeva, S. S. Kunizheva, and E. I. Rogaev 1423

Genetic Demography of the Population of Moscow. Migration Processes

O. L. Kurbatova, A. S. Gracheva, E. Yu. Pobedonostseva, and I. G. Udina 1438

Epigenome Wide Study of Association of CpG Sites Methylation and Aggressive Behavior

S. A. Borinskaya, A. V. Rubanovich, A. K. Larin, A. V. Kazantseva, Yu. V. Davydova, E. V. Generozov, E. K. Khusnutdinova, and N. K. Yankovsky 1450

Adaptive Changes of Fatty Acid Desaturation Genes in Indigenous Populations of Northeast Siberia

B. A. Malyarchuk, M. V. Derenko, and G. A. Denisova 1458

ГЕНЕТИКА, 2021, том 57, № 12, с. 1345–1355

УДК 575.17:599.9

РАЗРАБОТКА ИННОВАЦИОННЫХ ГЕНОГЕОГРАФИЧЕСКИХ И ГЕНОМНЫХ ТЕХНОЛОГИЙ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ И ИНДИВИДУАЛЬНЫХ ОСОБЕННОСТЕЙ ЧЕЛОВЕКА НА ОСНОВЕ ИЗУЧЕНИЯ ГЕНОФОНДОВ РЕГИОНОВ СОЮЗНОГО ГОСУДАРСТВА

© 2021 г. А. В. Кильчевский[1], *, Н. К. Янковский[2], **

[1]*Институт генетики и цитологии Национальной академии наук Беларуси, Минск, 220072 Республика Беларусь*

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

**e-mail: kilchev@presidium.bas-net.by*

***e-mail: yankovsky@vigg.ru*

Поступила в редакцию 29.07.2021 г.
После доработки 03.08.2021 г.
Принята к публикации 05.08.2021 г.

Мировая и отечественная практика применения ДНК-технологий в криминалистике показала их преимущества в сфере идентификации личности неизвестного индивида по сравнению с другими известными в криминалистике методами идентификации. В статье рассматриваются основные современные подходы применения генетических технологий в криминалистике, включая ДНК-идентификацию и ДНК-фенотипирование. Рассмотрены преимущества и недостатки различных подходов, основной акцент сделан на достижениях отечественных ученых, а также на перспективах дальнейших молекулярно-генетических исследований в области судебно-медицинской экспертизы. Представлены основные результаты реализации научно-технической программы Союзного государства "ДНК-идентификация" в области установления вероятной внешности, возраста, психо-эмоционального статуса, а также этногеографического и популяционного происхождения неизвестного индивида по его ДНК с применением технологий высокопроизводительного секвенирования. В ходе реализации программы также разработаны отечественные наборы реагентов для ДНК-идентификации, адаптированные под существующее и перспективное ресурсное и приборное обеспечение, используемое в криминалистических лабораториях Союзного государства.

Ключевые слова: ДНК-идентификация в криминалистике, ДНК-маркеры, базы данных, ДНК-фенотипирование, инновационные наборы реагентов, генетическая предрасположенность, персонализированная медицина.

DOI: 10.31857/S0016675821120079

В настоящее время методы ДНК-идентификации – определение принадлежности биологических следов индивиду на основе совпадения характеристик ДНК, выделенной из биологического материала, собранного на месте преступления, и полученного от подозреваемого, широко вошли в криминалистическую практику. Впервые метод выявления индивидуализирующих особенностей ДНК был описан в 1985 г. английским биологом А. Джеффрисом и вскоре применен в расследовании преступлений [1, 2]. Метод был назван "ДНК-фингерпринтом", или "геномной дактилоскопией" (в настоящее время более распространены названия "ДНК-идентификация", "генотипоскопия" и "анализ ДНК-профилей"). В России методы ДНК-идентификации получили развитие в 1990-х гг., что связано в частности с пионерскими работами Е.И. Рогаева и П.Л. Иванова [3–5]. Со времени появления методов выявления индивидуализирующих особенностей ДНК сменилось несколько подходов (гибридизация фрагментированной ДНК с радиоактивно меченными зондами, ПДРФ-анализ), но в итоге во всех странах пришли к использованию мультиаллельных аутосомных **STR**-маркеров (Short Tandem Repeats, короткие тандемные повторы). Унифицированные наборы STR-маркеров, включающие более десятка STR-локусов и маркер пола – полиморфный участок в гене амелогенина, применяются в США (система CODIS, Combined DNA Index System) и в Европейском Союзе (система ENFSI, The European Network of Forensic Science Institutes DNA working group). На основе этих систем созданы национальные базы данных ДНК-

профилей. В Великобритании национальная база данных в марте 2020 г. содержала 6.6 млн ДНК-профилей, из которых 80.3% принадлежали мужчинам, 19.1 – женщинам, а для 0.6% образцов пол не установлен. Ежегодно использовалось для анализа до 40000 образцов с места преступлений [6]. В США (по состоянию на апрель 2021 г.) в такой базе данных – более 14 млн профилей преступников, более 4 млн профилей арестованных и более 1 млн профилей с мест преступления [7]. Согласно сообщениям 110-ти (из 194) национальных центральных бюро Интерпола к 2019 г. 70 стран имели национальные базы данных ДНК-профилей и 89 стран применяли анализ ДНК в уголовном расследовании [7а].

В России для ДНК-идентификации (геномной регистрации) используются те же системы STR-маркеров, что делает получаемые ДНК-профили сопоставимыми с профилями в базах данных Интерпола и иных международных базах. Исследования ДНК-профилей проводятся в экспертно-криминалистических подразделениях СК, МВД и ФСБ России, учреждениях судебно-медицинской экспертизы Минздрава России. Согласно Федеральному Закону № 242-ФЗ “О государственной геномной регистрации в Российской Федерации” обязательной геномной регистрации подлежат лица, осужденные и отбывающие наказание в виде лишения свободы за совершение тяжких или особо тяжких преступлений, всех категорий преступлений против половой неприкосновенности и половой свободы личности, а также неопознанные трупы и биологические следы, изъятые с мест преступления. Список лиц, подлежащих обязательной геномной регистрации, предполагается расширить. Сведения по генетическим профилям хранятся в Федеральной базе данных геномной информации (**ФБДГИ**), оператором которой является МВД России. В настоящее время объем ФБДГИ составляет более 1.1 млн генетических профилей (0.6% населения России). Количество проводимых экспертиз с использованием ДНК-профилей составляет несколько сотен тысяч ежегодно [8, 8а].

В Республике Беларусь проведение ДНК-экспертиз регулируется локальными нормативно-правовыми актами Государственного комитета судебных экспертиз.

Помимо аутосомных STR-маркеров в криминалистической практике используется анализ однородительских маркеров – STR-маркеров Y-хромосомы и секвенирование гипервариабельных сегментов или полной последовательности митохондриальной ДНК (мтДНК). Все перечисленные системы могут быть использованы не только для идентификации принадлежности биологического образца конкретному индивиду, но и для установления родства. Именно анализ этих маркеров позволил идентифицировать останки всех членов семьи последнего российского императора Николая II с помощью сравнения полной последовательности мтДНК и STR-маркеров Y-хромосомы в ДНК из костных останков и из других биологических образцов представителей дома Романовых, родственников императрицы [9]. Наиболее заметным и убедительным итогом этих работ явилась идентификация останков самого императора Николая II тем же методом. При этом ДНК костных останков сравнивали с ДНК из биоматериалов, заведомо принадлежащих самому императору (кровь на рубашке после покушения на Николая II в Японии, находится в Государственном Эрмитаже) [10].

Системы однородительских маркеров, как неутрачиваемые в поколениях, могут применяться и для установления дальнего родства, тем самым указывая и на этногеографические группы, из которых происходит индивид. Для этого необходимы референсные базы данных генетических профилей представителей различных популяций, с которыми сравнивается генетический профиль исследуемого образца. Именно с помощью таких подходов были установлены регионы происхождения как исполнителя теракта в аэропорту Домодедово в 2011 г., так и новосибирского маньяка-педофила (были использованы, соответственно, популяционные базы данных Медико-генетического научного центра, Москва, и Института медицинской генетики, Томск) [11]. Полученная информация значительно ускорила раскрытие этих преступлений.

ДНК-идентификация широко применяется не только при раскрытии преступлений, но и при розыске потерявшихся людей, идентификации неопознанных тел, жертв природных и техногенных катастроф.

Для развития перечисленных направлений в России и Беларуси в 2017 г. была принята научно-техническая программа Союзного государства “Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства” (“ДНК-идентификация”), государственными заказчиками которой стали ФАНО/Минобрнауки России совместно с Национальной академией наук Беларуси. Программа была утверждена постановлением Совета Министров Союзного государства от 16 июня 2017 г. № 26 [11а] и рассчитана на реализацию в течение пяти лет

(2017–2021 гг.). В состав Программы вошли два основных блока исследований – исследования в области генетической экспертизы для криминалистики, выполняемые на единой методической платформе российскими и белорусскими участниками, и в области медицинской генетики для целей персонифицированной медицины, а также возможного применения данных подходов в криминалистике, выполняемых только белорусской стороной. Направления исследований по Программе перечислены в табл. 1.

Выявление ДНК-маркеров различных типов для определения вероятного этногеографического и популяционного происхождения и разработка баз данных генетических характеристик населения Союзного государства составили важнейшую часть исследований. Одна из задач – ликвидация "белых пятен" (областей, недостаточно исследованных по частотам встречаемости применяемых в криминалистике ДНК-маркеров) на геногеографической карте Союзного государства. Если к началу Программы по целевым маркерам было исследовано немногим более 1000 биологических образцов, представляющих несколько десятков популяций, то к ее завершению количество исследованных образцов составило более 30000, а количество этнорегиональных групп – более 200. Примеры таких исследований описаны в трех статьях данного выпуска журнала [25, 28, 29].

Столь большой объем образцов популяций Союзного государства был исследован в результате создания расширенного мультиплексного набора STR-локусов Y-хромосомы, применение которого описано в двух статьях данного номера журнала [28, 29]. В рамках Программы "ДНК-идентификация" созданы базы данных, характеризующие основные популяции Союзного государства по мужской линии. Большое количество мультиаллельных локусов ДНК индивида амплифицируются в мультиплексе в одной пробирке, и фрагменты фракционируются по размеру в единичном капилляре. Получаемый гаплотип вносится в созданную базу данных, что позволяет быстро и с высокой разрешающей способностью отнести индивида к одной из геногеографических групп. В некоторых случаях можно отнести его к конкретному народу, к определенному роду, а иногда даже к деревне данного географического региона.

Созданный набор реагентов производится почти исключительно на отечественном сырье, оборудование для ДНК-идентификации с помощью этого набора также производится в России. Созданный набор по разрешающей способности превышает зарубежные аналоги, а эффективность его применения обеспечивается уникальной базой генетических данных, разработанной в рамках Программы "ДНК-идентификация".

Развитие технологий генотипирования и геномного анализа получило значительное ускорение с появлением методов масштабного параллельного секвенирования (**МПС**), называемого также "секвенированием следующего поколения" (Next Generation Sequencing, NGS). Эти технологии позволяют перейти к использованию маркеров еще одного типа – аллелей однонуклеотидных полиморфных локусов (Single Nucleotide Polymorphism – SNP [55]). В отличие от мультиаллельных STR-маркеров, используемые SNP-маркеры являются биаллельными. Поэтому для проведения генетического анализа в целях криминалистики необходимо использовать гораздо большее количество SNP-маркеров, что дает не только определенные преимущества, но и проблемы, которые рассмотрены в статье группы Е.И. Рогаева в этом номере [49]. Полиморфный однонуклеотидный локус (в данном случае просто сайт с определенной позицией в геноме) может быть представлен в популяции не только двумя аллелями, но и большим их числом (максимум четырьмя аллелями – А, Т, Г, Ц). Хотя у индивида число аллелей такого локуса в двух гомологичных хромосомах не может превышать двух, если это локус в геноме уникален. Использование для ДНК-идентификации три- и тетра-аллельных однонуклеотидных локусов вместо биаллельных может значительно повысить разрешающую способность метода, одновременно сократив число пар праймеров в мультиплексе для их анализа. Существенно, что однонуклеотидный полиморфный локус может быть амплифицирован и определен его аллельный статус при минимальном размере амплифицируемого фрагмента ДНК (менее 50 нуклеотидов), который соответствует размеру фрагментов деградированной ДНК, встречающихся в биологических следах, с которыми работают криминалисты. Известно, что фрагменты такого размера сохраняются и в многотысячелетних останках. Практически значимыми для криминалистики являются те мультиаллельные однонуклеотидные локусы, для которых частоты аллелей примерно одинаковы во многих популяциях. Таким образом, необходимо охарактеризовать в группы населения Союзного государства по частотам аллелей SNP, уже известным в каких-либо популяциях мира.

Анализ SNP-маркеров позволяет на единой технологической платформе решать не только задачи ДНК-идентификации биологического материала, но и определять фенотипические призна-

Таблица 1. Направления исследований программы “ДНК-идентификация”

№	Направление	Цель исследования	Публикации
1	Разработка технологии определения вероятной внешности неизвестного индивида по характеристикам его ДНК	Разработка инновационной ДНК-технологии, позволяющей определить по биологическому материалу, содержащему ДНК, параметры внешности (цвет волос, цвет глаз) носителя этой ДНК	[12–17]
2	Разработка методики определения вероятного возраста индивида по характеристике его ДНК	Разработка методики определения вероятного возраста неизвестного индивида по образцу ДНК, полученной из его биологических следов, на основе анализа прижизненной модификации избранных участков ДНК. Данные могут быть использованы в криминалистике	[18–22]
3	Разработка технологии определения наиболее вероятного этногеографического происхождения неизвестного индивида по характеристикам его ДНК	Разработка инновационной комплексной технологии, позволяющей определить по характеристике ДНК наиболее вероятное этногеографическое происхождение неизвестного индивида	[23–27]
4	Разработка метода ДНК-идентификации для определения популяционной принадлежности неизвестного индивида	ДНК-идентификация и определение популяционной принадлежности неизвестного индивида по биоматериалам. Создание базы данных по комплексу генетических маркеров, дифференцирующих популяции населения Союзного государства, для получения информации, которая может быть использована также в криминалистике	[27–29]
5	Разработка методики определения статуса генетических локусов и прижизненной модификации участков ДНК, влияющих на психоэмоциональный статус человека	Получение методики, позволяющей оценить генетическую компоненту психоэмоционального статуса личности	[30–34]
6	Разработка методов для оценки механизмов генетического контроля широко распространенных заболеваний (сердечно-сосудистых, эндокринных, аутоиммунных, костно-мышечных, онкологических и некоронарогенных заболеваний сердца); разработка технологий выявления ДНК-маркеров болезней для применения их в криминалистике	Определение генетических рисков развития не менее восьми нозологий мультифакторных заболеваний на основе определения не менее 60 наиболее информативных генетических маркеров; создание базы данных по ДНК-маркерам широко распространенных заболеваний	
6.1	Разработка ДНК-технологии выявления генетического риска эндокринных заболеваний	Разработка методов определения риска развития разных вариантов сочетанных эндокринных заболеваний, основанных на оценке фенотипических, биохимических, молекулярно-генетических особенностей конкретного индивидуума	[35]

Таблица 1. Продолжение

№	Направление	Цель исследования	Публикации
6.2	Разработка инновационной ДНК-технологии, позволяющей определить индивидуальный генетический риск развития заболеваний, связанных с тромбозом	Определение генетического риска заболеваний тромбогенного характера (инфаркта миокарда, инсульта, тромбоэмболии легочных артерий) по ДНК индивида	[36, 37]
6.3	Разработка инновационной ДНК-технологии, позволяющей с помощью современных молекулярно-генетических методов выявить наиболее информативные варианты генов, регулирующих процессы костного метаболизма, и оценить риск развития заболеваний костно-мышечной системы	Изучение генетики заболеваний костно-мышечной системы	[38–40]
6.4	Молекулярно-генетическая оценка риска аутоиммунных заболеваний	Выявление молекулярно-генетических маркеров риска аутоиммунных заболеваний и создание на их основе геномных технологий выявления групп риска с данными заболеваниями	[41–43]
6.5	Разработка технологии выявления риска онкологических заболеваний на основе эпигенетических и молекулярно-генетических маркеров	Определение молекулярно-генетических и эпигенетических маркеров риска рецидивов и (или) прогрессирования опухолей у пациентов с онкопатологиями	[44]
6.6	Выявление ДНК-маркеров риска некоронарогенных заболеваний сердца	Определение генетических факторов патогенеза и генетических маркеров риска развития неблагоприятных клинических исходов (синдром внезапной смерти и прогрессирование сердечной недостаточности) у пациентов с некоронарогенной патологией (дилатационной кардиомиопатией, аритмогенной дисплазией правого желудочка, гипертрофической кардиомиопатией, изолированной некомпактностью миокарда левого желудочка и каналопатией)	[45–47]
7	Разработка и изготовление опытных образцов инновационных наборов реагентов для идентификации личности и установления родства на основе STR- и SNP-маркеров методом “массового параллельного секвенирования”	Изготовление наборов реагентов и рецептов к ним, а также изготовление опытных образцов таких наборов для применения в технологиях и методиках (см. мероприятия 1, 3, 4). Апробация данных наборов на практике (в криминалистических целях) с перспективой импортозамещения. Разрабатываемые наборы реагентов: 1) для определения вероятной внешности индивида на основе характеристик его ДНК, 2) для определения вероятной этногеографической и популяционной принадлежности индивида по его ДНК, 3) для ДНК-идентификации неизвестного индивида, 4) для выявления метаболитов с целью идентификации неизвестного индивида	[48, 49]

Таблица 1. Окончание

№	Направление	Цель исследования	Публикации
8	Адаптация применения разработанных отечественных инновационных наборов реагентов для ДНК-идентификации по деградированной ДНК и ДНК из малых количеств биоматериала	Расширение возможностей ДНК-идентификации применительно к биоматериалам, в которых ДНК деградирована, или материала очень мало; изготовление опытных образцов наборов для таких случаев и их апробация в криминалистических лабораториях. Разрабатываемые наборы реагентов: 1) для адаптации разработанных в мероприятии 7 наборов реагентов для генотипирования STR- и SNP-маркеров (определения внешности, этногеографического происхождения), 2) для идентификации пола, 3) для полного секвенирования митохондриальной ДНК	
9	Разработка и изготовление опытных образцов наборов реагентов для выявления ДНК-маркеров риска развития широко распространенных (сердечно-сосудистых и эндокринных) заболеваний для применения в криминалистике	Получение опытных наборов реагентов для определения генетической предрасположенности к широко распространенным заболеваниям на основе анализа не менее 60 наиболее информативных генетических маркеров, выявленных в мероприятии 6	[50]
10	Разработка методики формирования баз данных для ДНК-идентификации в смешанном населении мегаполисов и прогноза динамики генофонда мегаполиса под воздействием миграционных процессов	Опробирование методики формирования и анализа генетических баз данных для населения мегаполисов, учитывающей смешанный состав и динамичность этих популяционных структур, а также прогноз динамики генофонда населения мегаполиса под воздействием миграционных и других этнодемографических процессов	[51–54]

ки их носителей: таких как цвет глаз, цвет волос, цвет кожи, а с некоторой вероятностью и черты лица. Выявление и исследование ДНК-маркеров, ассоциированных с цветом глаз и волос, в белорусских и российских популяциях составило одно из направлений исследований (табл. 1). В данном номере журнала представлено описание геногеографического атласа частот аллелей ДНК-маркеров, контролирующих цвет глаз и волос [17].

Для реализации перечисленных подходов к ДНК-идентификации в настоящее время выпускаются различные коммерческие наборы реагентов (в основном зарубежные), поэтому актуальной является разработка отечественных наборов реагентов (табл. 1, направления 7 и 8). Важнейшая задача в криминалистике – анализ следовых количеств ДНК, причем нередко эта ДНК деградирована. При соблюдении таких подходов становится возможным анализ минимальных количеств ДНК – от 60 пикограмм, что соответствует нескольким человеческим геномам (Е.И. Рогаев, персональное сообщение). Эти особенности исследуемого в криминалистике материала требуют большой тщательности в разработке наборов реагентов для их анализа, включая синтез праймеров, а также и всю линейку реагентов свободных от контаминации человеческой ДНК. Кроме того, производство таких наборов реагентов, как и их применение нужно проводить в специально оборудованных и подготовленных особо чистых помещениях.

Опыт работы по программе Союзного государства, накопленный коллективом под руководством Е.И. Рогаева, выявил и успешно разрешил типичные проблемы, возникавшие с реагентами и инфраструктурой при разработке методов ДНК-идентификации на единичных молекулах ДНК, в том числе деградированных. Чтобы перенести разработанные методы из научных лабораторий в лаборатории практиков криминалистики, необходимо наладить на необходимом уровне произ-

водство разработанных реагентов. Данная задача представляется возможной и целесообразной на основе опыта указанного коллектива и при создании соответствующей инфраструктуры. Кроме того, возможно и целесообразно создать также инфраструктуру для проведения с таким набором реагентов особо важных работ по ДНК-идентификации в научных и практических целях.

Одна из задач криминалистики – определение возраста индивида по уровню метилирования избранных CpG-сайтов. Проведенные исследования показывают, что точность определения возраста при этом подходе составляет в среднем 3–4 года [56]. Разработка и верификация подобных методов анализа представлена в статье [22] и табл. 1 (направление 2). Кроме определения возраста, по эпигенетическим маркерам (которые формируются на основе маркеров в ДНК) возможно идентифицировать ткани и биологические жидкости, а также различать монозиготных близнецов. Уровень метилирования считается перспективным для оценки таких форм поведения, как курение и злоупотребление алкоголем, а также различных типов средовых воздействий, что также может найти применение в криминалистической практике [56, 57]. Сравнение генетических и эпигенетических предикторов показывает, что для многих признаков эпигенетические маркеры дают больший вклад, чем генетические [58]. Анализ эпигенетических маркеров проведен в исследованиях под руководством И.Б. Моссе, опубликованных в данном номере журнала [33, 34].

В исследованиях медико-генетического направления в белорусской популяции выявлены ДНК-маркеры, ассоциированные с повышенным риском развития широко распространенных заболеваний (сердечно-сосудистых, эндокринных, аутоиммунных, костно-мышечных, онкологических и некоронарогенных заболеваний сердца) (табл. 1, направление 6).

Особое направление представляет изучение генетико-демогарфической структуры и динамики мегаполисов. Мегаполисы характеризуются интенсивными миграционными процессами, смешанным составом населения в этническом и генетическом отношении, широким распространением межнациональных браков. В рамках Программы изучены четыре мегаполиса Союзного государства – Москва, Санкт-Петербург, Новосибирск и Минск. Показана неоднородность расселения этнических групп по территории мегаполисов, что необходимо учитывать при формировании баз данных. Показано, что для каждого мегаполиса необходимо создавать собственную базу данных, так как области миграционного притяжения мегаполисов и этнический состав населения различны. С учетом этого созданы первые базы данных для целей ДНК-идентификации по мтДНК и Y-хромосоме для изученных мегаполисов с привлечением генетико-демографических характеристик жителей. Разработан прогноз динамики генофонда населения мегаполиса под воздействием миграционных процессов, который показывает более выраженную динамику частот гаплогрупп Y-хромосомы в поколениях по сравнению с маркерами мтДНК. Необходимо дальнейшее развитие таких баз данных и мониторинг генетико-демографической структуры населения [52].

Результаты, полученные в ходе реализации Программы Союзного государства “ДНК-идентификация”, позволяют повысить скорость и чувствительность анализов ДНК. Они апробированы в криминалистических лабораториях Российской Федерации и Республики Беларусь и уже применяются для повышения эффективности розыскных мероприятий, установления личности неизвестного идентифицируемого лица. Результаты программы найдут применение и для решения таких актуальных гуманитарных задач, как быстрая идентификация жертв природных и техногенных катастроф, военных конфликтов, в том числе идентификация останков погибших в Великой Отечественной войне. Разработанные в рамках программы референсные базы данных частот встречаемости различных вариантов ДНК-маркеров в группах населения Союзного государства и в странах-источниках трудовой миграции отражают характеристику и динамику генофондов популяций коренного населения и смешанного населения мегаполисов. Базы данных будут дополняться по мере появления новых данных и расширения панелей ДНК-маркеров, таким образом они представляют уникальный ресурс, позволяющий оказать поддержку при выполнении следственных мероприятий. Расширение спектра генетических маркеров по различным направлениям вместе с мощным программным обеспечением перспективно для построения трехмерных моделей лиц людей по их генетическому материалу. Дальнейший прогресс возможен и в усовершенствовании методик расшифровки и интерпретации образцов деградированной ДНК или со смесью генетического материала нескольких человек.

В конечном итоге, внедрение результатов реализации программы “ДНК-идентификация” будет способствовать повышению качества жизни и безопасности населения Союзного государства.

Работа выполнена в рамках научно-технической программы Союзного государства “Разработка инновационных геногеографических и ге-

номных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства” (“ДНК-идентификация”) с белорусской стороны – при финансовой поддержке Национальной академии наук Беларуси (Государственный контракт № 1ДНК-2017/2017-28-068 от 30.08.2017), с российской стороны – при финансовой поддержке Министерства науки и образования РФ (Государственный контракт № 011-17 от 26.09.2017).

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. *Jeffreys A.J., Wilson V., Thein S.L.* Individual-specific “fingerprints” of human DNA // Nature. 1985. V. 316(6023). P. 76–79.
2. *Gill P., Jeffreys A.J., Werrett D.J.* Forensic application of DNA “fingerprints” // Nature. 1985. V. 318(6046). P. 577–579. https://doi.org/10.1038/318577a0
3. *Рогаев В.И., Сыроквашева Е.Ю., Пименов М.В., Стегнова Г.В.* Генотипоскопия человека: идентификация вида, пола и личности по генетическим отпечаткам ДНК в случае, связанном с покушением на убийство // Судебная медицина. 1992. Т. 35. № 1. С. 10–14.
4. *Иванов П.Л.* Индивидуализация человека и идентификация личности: молекулярная биология в судебной экспертизе // Вестник РАН. 2003. Т. 73. № 12. С. 1085–1097.
5. *Смирнова С.А., Омельянюк Г.Г., Стороженко И.В. и др.* Судебная молекулярно-генетическая экспертиза объектов биологического происхождения – новое направление судебно-экспертной деятельности Минюста России // Теория и практика судебной экспертизы. 2021. Т. 16. № 1. С. 6–18. https://doi.org/10.30764/1819-2785-2021-1-6-18
6. National DNA database strategy board biennial report 2018–2020. UK Home Office. Her majesty’s stationery office. September 2020 [Электронный ресурс]. Режим доступа: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/913015/NDNAD_Strategy_Board_AR_2018-2020_print.pdf. Загл. с экрана. (дата обращения 01.07.2021).
7. CODIS-NDIS Statistics. Statistics as of April 2021 [Электронный ресурс]. Режим доступа: https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics. Загл. с экрана. (дата обращения 01.07.2021).
7a. Interpol / DNA / Interpol Global DNA Profiling Survey, 2019 [Электронный ресурс] Режим доступа: https://www.interpol.int/How-we-work/Forensics/DNA (дата обращения: 01.07.2021).
8. Неизвестного “нарисует” эксперт [Электронный ресурс] / МВД-медиа. Режим доступа: https://mvd-media.ru/publications/police-of-russia/sluzhba-pol/neizvestnogo-narisuet-ekspert/ Загл. с экрана. (19.02.2021).
8а. Интервью исполняющего обязанности директора СЭЦ СК России Михаила Игнашкина информационному агентству “РИА Новости” [Электронный ресурс]. Режим доступа: http://sledcom.ru/press/interview/item/1532317/ (дата обращения 01.07.2021).
9. *Григоренко А.П., Боринская С.А., Янковский Н.К., Рогаев Е.И.* Достижения и особенности в работе с древней ДНК и с ДНК из сложных криминалистических образцов // Acta Naturae. 2009. № 3. С. 56–68.
10. *Rogaev E.I., Grigorenko A.P., Moliaka Y.K. et al.* Genomic identification in the historical case of the Nicholas II royal family // PNAS. 2009. V. 106(13). P. 5258–5263. https://doi.org/10.1073/pnas.0811190106
11. *Боринская С.А., Балановский О.П., Курбатова О.Л., Янковский Н.К.* По следам ДНК: как генетика народонаселения помогает криминалистике // Природа. 2020. № 11(1263). С. 3–14. https://doi.org/10.7868/S0032874X20110010
11а. Постановление Совета Министров Союзного государства от 16 июня 2017 г. № 26 “О научно-технической программе Союзного государства “Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства (“ДНК-ИДЕНТИФИКАЦИЯ”) [Электронный ресурс]. Режим доступа: https://nasb.gov.by/rus/activities/research/2016/dnk.pdf Загл. с экрана. (19.02.2021).
12. *Шаптуренко М.Н., Кондратюк А.В., Вакула С.И. и др.* Вариация пигментации радужки глаз белорусской популяции в связи с полиморфизмом генов HERC2 и OCA2 // Докл. Нац. акад. наук Беларуси. 2021. Т. 65. № 1. С. 59–67. https://doi.org/10.29235/1561-8323-2021-65-1-59-67
13. *Shapturenko M.N., Vakula S.I., Kandratsiuk A.V. et al.* HERC2 (rs12913832) and OCA2 (rs1800407) genes polymorphisms in relation to iris color variation in Belarusian population // Forensic Sci. Int. Genet. Suppl. Ser. 2019. V. 7. № 1. P. 331–332. https://doi.org/10.1016/j.fsigss.2019.09.127
14. *Середенко М.В., Вакула С.И., Шаптуренко М.Н. и др.* Прогностическая способность системы генетического фенотипирования hirisplex в белорусской популяции // Экол. генетика. 2021. Т. 19. № 1. С. 67–76. https://doi.org/10.17816/ecogen54547

15. *Балановский О.П., Петрушенко В.С., Горин И.О. и др.* Точность предикции пигментации волос и глаз по генетическим маркерам для популяций России // Вестник РГМУ. 2019. № 5. С. 98–114. https://doi.org/10.24075/vrgmu.2019.069

16. *Balanovska E., Kagazezheva J., Agdzhoyan A. et al.* Optimizing the genetic prediction of the eye and hair color for North Eurasian populations // BMC Genomics. 2020. V. 21. Suppl. 7. P. 527. https://doi.org/10.1186/s12864-020-06923-1

17. *Балановская Е.В., Горин И.О., Кошель С.М., Балановский О.П.* Геногеографический атлас ДНК-маркеров, контролирующих цвет глаз и волос человека // Генетика. 2021. Т. 57. № 12. С. 1356–1375.

18. *Золотаренко А.Д., Чекалин Е.В., Брускин С.А.* Современные молекулярно-генетические методы определения возраста в криминалистике // Генетика. 2019. Т. 55. № 12. С. 1397–1409. https://doi.org/10.1134/S0016675819120154

19. *Кипень В.Н., Богданова М.В., Буракова А.А. и др.* Предсказательный потенциал CpG-маркеров для определения хронологического возраста человека // Мол. и прикладная генетика. 2020. Т. 28. С. 80–90.

20. *Кипень В.Н., Богданова М.В., Буракова А.А. и др.* Обоснование минимального объема выборки для предсказания хронологического возраста человека // Мол. и прикладная генетика. 2021. Т. 30. С. 39–48. https://doi.org/10.47612/1999-9127-2021-30-39-48

21. *Лемеш В.А., Богданова М.В., Буракова А.А. и др.* Разработка MS-SNUPE праймеров для определения статуса метилирования генов человека ELOVL2, F5 и ZYG11A // Мол. генет., микробиология и вирусология. 2019. Т. 37. Спецвыпуск. С. 35–36.

22. *Лемеш В.А., Кипень В.Н., Богданова М.В. и др.* Определение хронологического возраста человека по биологическим образцам на основании анализа метилирования CpG-динуклеотидов // Генетика. 2021. Т. 57. № 12. С. 1376–1385.

23. *Беленикина М.С., Галахова А.А., Балановская Е.В., Балановский О.П.* Оценка воспроизводимости флуориметрического измерения концентрации ДНК // Генетика. 2018. Т. 54. № 13. С. 113–116. https://doi.org/10.1134/S0016675818130040

24. *Балановский О.П., Горин И.О., Записецкая Ю.С. и др.* Взаимодействие генофондов русского и финноязычного населения Тверской области: анализ 4 млн SNP-маркеров // Вестник РГМУ. 2020. № 6. С. 23–31. https://doi.org/10.24075/vrgmu.2020.072

25. *Агджоян А.Т., Дамба Л.Д., Гурьянов В.М. и др.* Филогенетический анализ южносибирской гаплогруппы Q-YP1102 по данным о Y-SNP и Y-STR-маркерах у тувинцев и окружающих популяций // Генетика. 2021. Т. 57. № 12. С. 1386–1395.

26. *Парфенчик М.С., Котова С.А.* Теоретические подходы к формированию панелей ДНК маркеров для определения этногеографического происхождения индивида в судебной экспертной деятельности // Генетика. 2021. Т. 57. № 1. С. 5–14. https://doi.org/10.31857/S0016675821010100

27. *Колесников Н.А., Харьков В.Н., Зарубин А.А., Степанов В.А.* Особенности распределения регионов высокой гомозиготности в геномах представителей коренного населения Северной Евразии // Генетика. 2019. Т. 55. № 10. С. 1231–1239. https://doi.org/10.1134/S0016675819100072

28. *Харьков В.Н., Валихова Л.В., Яковлева Е.Л. и др.* Реконструкция происхождения гыданских ненцев на основе генетического анализа их родовой структуры с помощью нового набора YSTR-маркеров // Генетика. 2021. Т. 57. № 12. С. 1403–1414.

29. *Харьков В.Н., Котова С.А., Колесников Н.А. и др.* Генетическое разнообразие 21 аутосомного STR-маркера системы CODIS в популяциях Восточной Европы // Генетика. 2021. Т. 57. № 12. С. 1396–1402.

30. *Драгович А.Ю., Боринская С.А.* Генетическая и геномная основа агрессивного поведения человека // Генетика. 2019. Т. 55. № 12. С. 1381–1396. https://doi.org/10.1134/S0016675819090054

31. *Жур К.В., Моссэ И.Б., Кильчевский А.В. и др.* Ассоциация ряда полиморфизмов генов нейромедиаторных систем с психофизиологическими характеристиками спортсменов // Мол. и прикладная генетика. Т. 26. 2019. С. 136–144.

32. *Бабенко А.С., Моссэ К.А., Седляр Н.Г. и др.* Разработка панели для анализа профиля метилирования целевых эпигенетических локусов, ассоциированных с психоэмоциональным статусом человека // Молекулярная и прикладная генетика. 2020. Т. 29. С. 37–48.

33. *Моссэ И.Б., Седляр Н.Г., Бабенко А.С. и др.* Ассоциация метилирования генов нейромедиаторных систем мозга с психоэмоциональными характеристиками человека // Генетика. 2021. Т. 57. № 12. С. 1415–1422.

34. *Боринская С.А., Рубанович А.В., Ларин А.К. и др.* Полногеномное исследование связи метилирования CpG-сайтов с агрессивным поведением // Генетика. 2021. Т. 57. № 12. С. 1450–1457.

35. *Амельянович М.Д., Моссэ И.Б.* Молекулярно-генетические маркеры риска развития эндокринной патологии // Мол. и прикладная генетика. 2019. Т. 27. С. 97–107.

36. *Булгак А.Г., Моссэ И.Б., Зотова О.В., Королева Т.С., Николаева Н.В., Гончар А.Л.* Роль генетического полиморфизма в развитии инфаркта миокарда среди мужчин из республики Беларусь // Медикобиол. проблемы жизнедеятельности (Клинич. медицина). 2020. № 2(24). С. 92–102.

37. *Моссэ И.Б, Гончар А.Л., Кундас Л.А. и др.* Молекулярно-генетические факторы предрасположенности к развитию фибрилляции предсердий у представителей белорусской популяции // Кардиология. 2021. Т. 13. № 5. С. 500–511.

38. *Marozik P., Alekna V., Rudenko E. et al.* Bone metabolism genes variation and response to bisphosphonate treatment in women with postmenopausal osteoporosis // PLoS One. 2019. V. 14. № 8. e0221511. P. 1–14. https://doi.org/10.1371/journal.pone.0221511

39. *Marozik P., Rudenka A., Kobets K., Rudenka E.* Vitamin D status, bone mineral density and VDR gene polymorphism in a cohort of Belarusian postmenopausal women // Nutrients. 2021. V. 13(3). P. 837. https://doi.org/10.3390/nu13030837

40. *Rudenka A.V., Rudenka E.V., Samokhovec V.Yu. et al.* Vitamin D receptor gene polymorphism, bone mineral density and 25(OH)D level in women with osteoporosis // Proc. Nat. Acad. Sci. of Belarus. Med. Series. 2020. V. 17. № 4. P. 480–492. https://doi.org/10.29235/1814-6023-2020-17-4-480-492

41. *Яцкив А.А., Сукало А.В., Гончарова Р.И.* Ассоциация полиморфных вариантов генов иммунного и воспалительного ответа, не относящихся к главному комплексу гистосовместимости, с патологией суставов у детей в Республике Беларусь // Мед. генетика. 2020. Т. 19. № 9(218). С. 25–36. https://doi.org/10.25557/2073-7998.2020.09.25-36

42. *Яцкив А.А., Злотникова М.В., Сукало А.В., Гончарова Р.И.* Характерные особенности спектров HLA-аллелей классов I и II у пациентов с различными клиническими формами ювенильного идиопатического артрита в Республике Беларусь // Докл. Нац. акад. наук Беларуси. 2020. Т. 64. № 2. С. 209–216. https://doi.org/10.29235/1561-8323-2020-64-2-209-216

43. *Bakutenko I.Y., Hileuskaya I.D., Nikitchenko N.V. et al.* Polymorphism of proteasomal genes can be a risk factor for systemic autoimmune diseases in children // J. Pediatric Genet. 2021. V. 10(2). P. 98–104. https://doi.org/10.1055/s-0040-1714697

44. *Михаленко Е.П., Щаюк А.Н., Кильчевский А.В.* Сигнальные пути: механизм регуляции пролиферации и выживаемости опухолевых клеток (обзорная статья) // Мол. и прикладная генетика. 2019. Т. 26. С.145–157.

45. *Sivitskaya L., Vaikhanskaya T., Danilenko N. et al.* Splicing mutation in LAMP2 gene leading to exon skipping and cardiomyopathy development // Gene Reports. 2020. V. 18. P. 100564. https://doi.org/10.1016/j.genrep.2019.100564

46. *Сивицкая Л.Н., Вайханская Т.Г., Левданский О.Д. и др.* Семейный случай дилатационной кардиомиопатии, ассоциированной с мутацией в гене фактора сплайсинга RBM20 // Мед. генетика. 2018. Т. 17. № 11. С. 48–52. https://doi.org/10.25557/2073-7998.2018.11.48-52

47. *Сивицкая Л.Н., Вайханская Т.Г., Даниленко Н.Г. и др.* Новые мутации в гене интегрин-связанной киназы у пациентов с аритмогенной кардиомиопатией // Мед. генетика. 2020. Т. 19(5). С. 18–19. https://doi.org/10.25557/2073-7998.2020.05.18-19

48. *Подгорский В.В., Боринская С.А.* Анализ метаболитов в исследовании криминалистических образцов // Acta Naturae. 2021 Т. 13. № 1. С. XX–XX.

49. *Тяжелова Т.В., Кузнецова И.Л., Андреева Т.В. и др.* Использование масштабного параллельного секвенирования в криминалистике – сравнительный анализ платформ для секвенирования // Генетика. 2021. Т. 57. № 12. С. 1423–1437

50. *Гайдукевич И.В., Горькавая А.М., Грудо А.В. и др.* Разработка мультиплексной системы для определения маркеров предрасположенности к развитию сердечно-сосудистых заболеваний // Вестн. Нац. акад. навук Беларусі. Сер. хім. навук. 2021. Т. 57. № 1. С. 48–60. https://doi.org/10.29235/1561-8331-2021-57-1-48-60

51. *Курбатова О.Л., Удина И.Г., Грачева А.С. и др.* Генетико-демографические параметры населения г. Санкт-Петербурга. Миграционные процессы // Генетика. 2019. Т. 55. № 9. С. 1071–1082. https://doi.org/10.1134/S001667581909008X

52. *Курбатова О.Л., Грачева А.С., Победоносцева Е.Ю., Удина И.Г.* Генетико-демографические параметры населения г. Москвы. Миграционные процессы // Генетика. 2021. Т. 57. № 12. С. 1438–1449.

53. *Цыбовский И.С., Спивак Е.А., Котова С.А. и др.* Территориальная подразделенность популяции мегаполиса по этническому признаку в связи с проблемой создания генетических баз данных. Минск // Генетика. 2021. Т. 57. № 8. С. 955–963. https://doi.org/10.31857/S0016675821080142

54. *Цыбовский И.С., Котова С.А., Спивак Е.А. и др.* ДНК-технологии следующего поколения для новых решений в криминалистике // Наука и инновации. 2020. № 10(212). С. 17–21.

55. *Butler J.M., Coble M.D., Vallone P.M.* STRs vs. SNPs: thoughts on the future of forensic DNA testing // Forensic Sci. Med. Pathol. 2007. V. 3(3). P. 200–205. https://doi.org/10.1007/s12024-007-0018-1

56. *Parson W.* Age estimation with DNA: From forensic DNA fingerprinting to forensic (EPI)genomics: a mini-review // Gerontology. 2018. V. 64(4). P. 326–332. https://doi.org/10.1159/000486239

57. *Vidaki A., Kayser M.* Recent progress, methods and perspectives in forensic epigenetics // Forensic Sci. Int. Genet. 2018. V. 37. P. 180–195. https://doi.org/10.1016/j.fsigen.2018.08.008

58. *McCartney D.L., Hillary R.F., Stevenson A.J. et al.* Epigenetic prediction of complex traits and death // Genome Biol. 2018. V. 19. № 1. P. 136. https://doi.org/10.1186/s13059-018-1514-1

Developing the Innovative Gene Geographical and Genomic Technologies for Identification and Revealing the Personal Features by Studying the Gene Pools of the Regional Populations

A. V. Kilchevsky[a], * and N. K. Yankovsky[b], **

[a] *Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, 220072 Republic of Belarus*

[b] *Vavilov Institute of General Genetics, Moscow, 119991 Russia*

**e-mail: kilchev@presidium.bas-net.by*

***e-mail: yankovsky@vigg.ru*

Global and Russian experience in using DNA technologies in investigation of crimes showed that the technologies are superior to other forensic identification methods in personal identification of an unknown individual. The article considers the main current approaches that utilize genetic technologies in forensic studies, including DNA identification and DNA phenotyping. Advantages and drawbacks are discussed for several approaches. Emphasis is placed on the achievements of Russian researchers and the prospects of further molecular genetic studies in the field of forensic medicine. Main results obtained in the "DNA Identification" research-and-technological program of the Union State of Russia and Belarus are described with the focus on establishing the probable appearance, age, psychoemotional status, and ethnogeographic and population origin of an unknown person by examining their DNA with the use of high-throughput sequencing techniques. Work on the program additionally yielded extended version DNA identification kits, which are adapted to the available and prospective resources and equipment used in forensic labs of the Union State.

Keywords: forensic DNA identification, DNA markers, databases, DNA phenotyping, genetic predisposition, personalized medicine, innovative reagent kits.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1356–1375

УДК 575.1

ГЕНОГЕОГРАФИЧЕСКИЙ АТЛАС ДНК-МАРКЕРОВ, КОНТРОЛИРУЮЩИХ ЦВЕТ ГЛАЗ И ВОЛОС ЧЕЛОВЕКА[1]

© 2021 г. Е. В. Балановская[1, 2, *], И. О. Горин[1, 3], С. М. Кошель[1, 4], О. П. Балановский[1, 2, 3]

[1]Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, 115478 Россия

[2]АНО "Биобанк Северной Евразии", Москва, 115215 Россия

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

[4]Московский государственный университет им. М.В. Ломоносова, Москва, 119991 Россия

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

Поступила в редакцию 31.05.2021 г.
После доработки 18.06.2021 г.
Принята к публикации 21.06.2021 г.

Изучение генетического контроля пигментации выявило множество ДНК-маркеров, влияющих на формирование цвета глаз и волос человека. Зарубежная панель маркеров HIrisPlex разработана на основе генетического разнообразия населения Западной Европы и имеет несколько сниженную точность в российских популяциях, тогда как отечественная панель маркеров, включающая 31 SNP, разработана на российских выборках. В данном исследовании изучено распространение этих маркеров в популяциях России и сопредельных стран. Каждый из 31 маркеров, входящих в отечественную панель, был генотипирован у представителей 22 популяций, охватывающих основное разнообразие генофондов Северной Евразии, по суммарной выборке 1153 человек. На основе полученных аллельных частот был создан картографический атлас, который включил карты распространения 62 аллелей, а также карты чувствительности предикции фенотипа по генотипу. Картографический и статистический анализы выявили три основные закономерности генетического ландшафта маркеров пигментации. Во-первых, основным паттерном большинства карт является градиент изменений частот в направлении "восток–запад". В большинстве случаев (но есть и исключения) к западу нарастает частота аллеля, определяющего светлую пигментацию глаз и волос. Во-вторых, отсутствует связь между величиной генетического вклада ДНК-маркера в пигментацию и его межпопуляционной изменчивостью. В-третьих, обнаружено генетическое своеобразие популяций Кавказа и Урала, у которых частоты генетических маркеров пигментации несводимы ни к "западному", ни к "восточному" типу. Карты чувствительности предикции пигментации выявили также своеобразие русских популяций.

Ключевые слова: геногеография, атлас, пигментация, Северная Евразия, ДНК-маркеры, предикция, карты.

DOI: 10.31857/S0016675821120031

Степень пигментации, определяющая цвет глаз и волос человека, является традиционным и одним из наиболее интенсивно изучаемых фенотипических признаков за всю историю генетики. Генетический контроль пигментации у человека хорошо изучен, определены основные гены и их аллели, но при этом продолжают появляться все новые данные – во-первых, о новых маркерах, также вносящих свой вклад в фенотипы пигментации, и во-вторых, об особенностях генетического контроля пигментации в разных популяциях. Степень сложности изучения генетического контроля нарастает при переходе от цвета глаз к цвету волос, и далее – к цвету кожи.

Первое направление – обнаружение новых маркеров – сейчас развивается в основном за счет накопления больших выборок, которые одновременно и фенотипированы по цвету волос, и генотипированы по широкогеномным панелям. Например, исследование [1] основано на данных о более чем 350 тыс. человек из Биобанка Соединенного королевства, а работа [2] – на данных о более чем 150 тыс. человек, изученных компанией 23andMe. Если важнейшие генетические варианты, вносящие основной вклад в формирование цвета волос и глаз, единичны, то число вариантов, вносящих меньший, но заметный вклад, исчисляется десятками, а общее число SNP, ассоци-

[1] Дополнительная информация для этой статьи доступна по doi 10.31857/S0016675821120031 для авторизованных пользователей.

ированных с признаками пигментации, по некоторым данным превышает десять тысяч.

Второе направление – изучение разнообразия человечества по генетике пигментации – подразумевает не только очевидные различия в частотах ключевых генов между разными популяциями (которые и определяют разнообразие популяций по цвету глаз, волос и кожи), но и различия в том, какие именно генетические маркеры определяют характер пигментации в данном регионе. Наибольшее разнообразие по фенотипам и генотипам пигментации наблюдается у европейцев. При этом нетемные оттенки цвета волос и глаз встречаются и в некоторых других популяциях мира. Показано, что набор генетических маркеров, ассоциированных с цветом глаз, в Европе, Южной Азии и Восточной Азии неодинаков [3]: хотя в нем участвуют обычно одни и те же гены, одни и те же метаболические пути, но наборы аллелей, вовлеченных в контроль пигментации, могут различаться в разных популяциях мира. Особенным богатством и разнообразием отличаются варианты генов, контролирующие цвет кожи: их разнообразие велико даже в популяциях Африки южнее Сахары, несмотря на их кажущееся фенотипическое единообразие по цвету кожи [4]. В целом считается, что цвет кожи формировался под сильным давлением естественного отбора и определялся балансом защиты от ультрафиолета и необходимого уровня синтеза витамина D, тогда как в распространении аллелей цвета глаз и волос существенную роль играли и половой отбор, и дрейф генов.

Знания о генетическом контроле пигментации уже широко применяются по меньшей мере в двух областях – при изучении древней ДНК и в криминалистике. При анализе древних популяций часто по их геномам определяются вероятные фенотипы, и в первую очередь как раз признаки пигментации. Например, генетическое исследование населения древнего (минойского) Крита показало, что оно имело темные глаза и темные волосы, а значит вполне соответствовало своим изображениям на керамике и фресках [5], причем этот фенотип был распространен на островах Эгейского моря по меньшей мере с бронзового века [6]. Исследования древней ДНК также помогают выяснить, в какое время и в каких регионах мира распространялись аллели, связанные с пигментацией, в первую очередь “европейские” аллели светлой пигментации. Эти аллели обнаруживаются у населения Скандинавии, начиная с мезолита, но для более южных областей (как Западной, так и Восточной Европы) в мезолите характерна темная пигментация, в том числе кожи [7–9]. Повышение частоты “светлых” аллелей началось с неолита и продолжалось в последующие периоды истории [10]. При этом в предковой для неолитической Европы популяции анатолийских земледельцев эти аллели хотя и встречались [11], но были редки. Информации по древним популяциям с территории нынешней Центральной России крайне мало, но имеющиеся данные по фатьяновской культуре [12] указывают, что в бронзовом веке большинство населения было темнопигментированным, хотя около трети фатьяновцев имели голубые глаза и светлые волосы. Все эти оценки пигментации по генотипу основываются на исследованиях связи генотип–фенотип, проведенных для современных европейцев, поэтому их применение для населения других эпох и других регионов мира является экстраполяцией.

Второе направление, применяющее знания о генетическом контроле пигментации, – криминалистика. Очевидно, что возможность предсказания внешности человека по его ДНК, в том числе обнаруженной в следовых количествах, для целей криминалистики крайне важна. Поэтому в рамках именно этого направления были отобраны из всего множества генетических маркеров, ассоциированных с пигментацией, те реально значимые ДНК-маркеры, которые позволяют по генотипу определить вероятный фенотип, и разработаны конкретные алгоритмы предикции пигментации. Исследователями голландского университета им. Эразма Роттердамского и их коллегами были последовательно разработаны три такие панели маркеров и соответствующие алгоритмы предикции фенотипа по генотипу: панель IrisPlex (предикция цвета глаз [13]), панель HIrisPlex (предикция цвета глаз и волос [14]) и панель HIrisPlex-S (предикция цвета глаз, волос и кожи [15]).

Эти панели разработаны преимущественно на образцах из голландской популяции, и, хотя они и были валидированы на представителях еще трех стран зарубежной Европы, остается неизвестным, насколько они информативны для населения других регионов мира, поскольку генетический контроль пигментации неодинаков в населении разных частей света. Это особенно актуально для генофонда населения России и сопредельных стран, который характеризуется наибольшим межпопуляционным разнообразием по сравнению с генофондами коренного населения всех других регионов мира [16]: генофонды населения разных регионов России генетически значительно отличаются не только от населения Западной Европы, но и друг от друга. Одни и те же признаки внешности в населении разных регионов могут определяться разными мутациями в одних и тех же генах или разными комплексами ДНК-маркеров. Поэтому применение зарубежных криминалистических наборов для населения России может иметь различную информативность в разных регионах России. Однако точность определения пигментации по генотипу можно повысить, если корректно провести исследования связи ДНК-маркеров и при-

знаков внешности в популяциях, охватывающих генетическое разнообразие населения России и сопредельных стран. Такая задача является не только пионерской для России, но актуальна для мировой науки в целом, а также для криминалистической практики.

Проведенная нами прямая проверка точности предсказаний системы HIrisPlex на популяциях России [17] показала, что точность в целом удовлетворительная, хотя и несколько сниженная по сравнению с точностью ее работы на западноевропейских популяциях (показатели качества по разным признакам цвета глаз и волос составляют около 0.9 для населения зарубежной Европы и около 0.8 для населения России). Это особенно выражено для коренного населения Сибири: HIrisPlex предсказывает темные глаза практически для всех сибирских образцов, тогда как у некоторых из них цвет глаз в действительности светлый или по меньшей мере смешанный. Поэтому мы провели полноэкзомное секвенирование (а также секвенирование интронов генов) для трехсот образцов представителей различных популяций России и сопредельных стран, для которых было также проведено определение фенотипов пигментации. Результаты позволили выделить панель генетических маркеров, оптимальных для предикции пигментации глаз и волос коренного населения России; последний вариант этой панели включил 31 маркер [18, 19]. Из этих 31 маркера шесть маркеров входят также в панель HIrisPlex и еще семь маркеров входят в панель HIrisPlex-S (т.е. для европейских популяций они связаны с пигментацией кожи, но не глаз и волос). Но большинство маркеров (остальные 18 SNP) ранее не связывались с признаками пигментации, т.е. были выявлены как эффективные именно для населения России. Однако их пространственное распределение в популяциях коренного населения Северной Евразии остается неизвестным.

Поэтому цель данной работы – изучить распространение генетических маркеров, связанных с пигментацией (отечественная панель из 31 маркера), в популяциях Северной Евразии и для этого создать картографический атлас генетического контроля пигментации в коренном населении России и сопредельных стран.

Как известно, “атлас – это систематическое собрание карт, выполненное по единой программе как целостное произведение и изданное в виде книги или комплекта листов. Это не просто набор карт под общим переплетом, но система взаимоувязанных и взаимодополняющих друг друга карт”, а одно из предназначений атласа – проведение углубленных научных исследований [20]. Атлас представляет собой комплекс карт взаимосвязанного тематического содержания, выполненных в едином стиле, проекции, масштабе и с единым охватом территории. Поэтому данная работа нацелена на создание атласа генетического контроля пигментации коренного населения России и сопредельных стран, который, удовлетворяя всем требованиям картографической науки, позволил бы выявить основные тренды и особенности генетического ландшафта генов, ответственных за пигментацию населения Северной Евразии.

МАТЕРИАЛЫ И МЕТОДЫ

Образцы и генотипы

Атлас генетического контроля пигментации основан на ДНК-профилях 1153 образцов из множества популяций, охватывающих основное генетическое разнообразие коренного населения России и сопредельных стран. Образцы ДНК получены из Биобанка Северной Евразии и собраны на основе добровольного участия и письменного информированного согласия под контролем Этической комиссии Медико-генетического научного центра. Образцы получены от неродственных между собой представителей коренного населения, которые в трех поколениях (включая обоих бабушек и дедов) относили себя к данному этносу и происходят из одной популяции (подробно подходы к сбору образцов описаны в [21]). 1153 образцов, генотипированных в данном исследовании, не включают в себя 301 образец, секвенированный в нашей предыдущей работе [18]. Для целей данного исследования были выделены 22 метапопуляции, набор которых охватывает основное генетическое разнообразие населения Северной Евразии. Затем из коллекций биобанка отбирались образцы, относящиеся к каждой из метапопуляций, таким образом, чтобы выборка из каждой метапопуляции составила около 50 образцов. При этом мы стремились охватить и разнообразие внутри метапопуляций, например включить образцы не одной, а нескольких этнических или субэтнических групп в пределах метапопуляции. Предпочтение отдавалось образцам, по которым в биобанке имелись не только биологические образцы, но и антропологические фотографии для возможности определения фенотипа. Перечень изученных популяций представлен в табл. 1, причем номера метапопуляций соответствуют номерам, подписанным на всех картах в местах их расположения.

Генотипирование проведено с использованием чипа индивидуального дизайна Infinium iSelect HD Custom BeadChip (Illumina, США) на платформе iScan (Illumina, США). Программой GenomeStudio v2.0 сформирована для каждого образца интегральная оценка качества полученных результатов CallRate, которая для всех образцов составила не менее 0.99. С помощью программного обеспечения PLINK извлечены генотипы по

Таблица 1. Изученные метапопуляции с указанием объемов выборок (*N*) и номеров (№), соответствующих подписям на картах

Метапопуляция		Состав (этнические и региональные группы)	*N*	Географические координаты метапопуляции
№	условное название			
1	Камчатка	Коряки, чукчи	51	Средние по локальным популяциям коряков и чукчей Камчатки
2	Амур	Нанайцы	48	Средние по локальным популяциям нанайцев Приамурья
3	Тунгусы	Эвенки, эвены	54	Средние по локальным популяциям эвенов и эвенков Дальнего Востока
4	Южная Сибирь	Тоджинцы, тофалары	53	Координаты тофаларов
5	Алтае-Саяны	Алтайцы	50	Средние по локальным популяциям северных и южных алтайцев
6	Монголия	Монголы	50	Средние по локальным популяциям субэтносов монголов
7	Угры	Ханты, манси	51	Средние по локальным популяциям хантов и манси
8	Казахстан	Народы Казахстана	72	Координаты центра Казахстана
9	Таджикистан	Народы Таджикистана	52	Координаты шугнанцев (народы Памира)
10	Западный Кавказ	Адыгейцы	61	Средние по локальным популяциям адыгейцев и причерноморских шапсугов
11	Восточный Кавказ	Народы Дагестана	50	Средние по локальным популяциям народов Дагестана
12	Закавказье	Азербайджанцы	49	Координаты центра Азербайджана
13	Русский Север	Русские северные популяции	51	Координаты Пинежского района Архангельской области
14	Русские (центральные)	Русские популяции центра России	55	Координаты Ярославля (основная часть выборки)
15	Зарубежная Европа	Македонцы	50	Координаты центра Македонии
16	Западные финно-угры	Тверские карелы, ижора	50	Средние по локальным популяциям ижоры
17	Восточные славяне	Белорусы и украинцы	52	Координаты центра Белоруссии
18	Мордовия	Популяции Мордовии	50	Средние по локальным популяциям мокши, шокши и эрзян
19	Тюрки Приуралья	Чуваши	50	Средние по локальным популяциям субэтносов чувашей
20	Марий Эл	Марийцы	50	Средние по локальным популяциям горных и луговых марийцев
21	Пермь	Коми-пермяки	50	Средние по локальным популяциям коми-пермяков
22	Южный Урал	Башкиры	53	Средние по локальным популяциям башкир

31 ДНК-маркеру, используемому в отечественном предикторе пигментации: 15 маркеров наиболее важны, а 16 имеют меньший вес при расчете вероятности того или иного фенотипа [18, 19]. На основании полученных генотипов были рассчитаны частоты встречаемости каждого маркера в каждой метапопуляции. Эти частоты представлены в табл. 2 (для 15 наиболее важных маркеров) и в табл. 3 (для 16 менее значимых маркеров).

В табл. 2 и 3 для каждого аллеля приведены показатели его генетического вклада в пигментацию глаз (C_E), волос (C_H) и рыжины волос (C_R), также указан суммарный вклад (C) во все три признака (каждый из которых взят по модулю). Генетический вклад референсного и альтернативного аллелей одинаков по величине, но различен по знаку. Из двух аллелей в таблицах приведен тот, который контролирует светлую пигментацию глаз, светлую пигментацию волос и рыжину волос. Поэтому положительный знак генетического вклада указывает на контроль светлой пигментации, отрицательный знак указывает, что этот маркер контролирует темную пигментацию (или отсутствие рыжины). Поскольку со светлой пигментацией оказались связаны преимущественно альтернативные аллели, в таблице они указаны без индекса, а для референсных аллелей указан верхний индекс R (A^R, C^R, G^R, T^R).

В таблицах также указано, какие 13 из 31 ДНК-маркеров присутствуют в зарубежных криминалистических панелях: шесть ДНК-маркеров из панели HIrisPlex указаны как **H** и семь из панели HIrisPlex-S указаны как **H-S**. Остальные 18 ДНК-маркеров были выявлены в наших предыдущих работах как важные для генетического контроля пигментации коренного населения России и сопредельных стран.

Определение чувствительности предикции

При анализе чувствительности предикции пигментации анализировались генотипы только тех индивидов, для которых имелись качественные цветные фотографии. Поэтому в выборку не вошли шесть метапопуляций: три зарубежные (№ 6, 15, 17) и три российские (№ 4, 5, 13). Часть метапопуляций с малой выборкой фенотипированных образцов была объединена с более многочисленными в три региональные группы: 1) центральные русские популяции и Мордовия; 2) популяции, фенотипически пограничные между европейскими и азиатскими (угры, Казахстан, Южный Урал, Таджикистан); 3) народы Восточного Кавказа и Закавказья. Остальные восемь метапопуляций (№ 1, 2, 3, 6, 10, 19, 20, 21) обладали большими фенотипированными выборками и каждая из них сформировала собственную региональную популяцию. Суммарный объем выборки полученных 11 региональных популяций (три объединенных и восемь исходных метапопуляций) составил 425 образцов. Пигментация определялась по цветным антропологическим фотографиям по стандартным антропологическим шкалам (цвет глаз по шкале Бунака, цвет волос по шкале Фишера), а затем сводилась к трем дихотомическим признакам: цвет глаз (темный, светлый), цвет волос (темный, светлый), рыжина волос (наличие, отсутствие). Поскольку для этих образцов имелись и генотипы, и фенотипы, для них была возможность сравнить фенотип, предсказанный по генотипу, и фенотип, определенный по фотографиям. Это позволило рассчитать чувствительность предикции для 11 региональных популяций и построить карты региональной чувствительности предикции цвета глаз и волос.

Картографический анализ

По частотам аллелей, приведенным в табл. 2 и 3, построены 62 карты распространения 62 аллелей 31 ДНК-маркеров, используемых для предикции цвета глаз и волос. Для удобства чтения в таблицах и в тексте частоты выражены в процентах, но на картах традиционно приведены в долях от 1. Для каждой метапопуляции в таблицах указан номер, по которому ее легко найти на картах.

Картографирование распространенности ДНК-маркеров выполнено способом послойной окраски на основе сеточной цифровой модели, представляющей собой матрицу значений картографируемого показателя (частот встречаемости) в узлах регулярной квадратной сетки в проекции карты. Исходными данными для создания модели служат географические координаты i-той метапопуляции (точка p_i) и значение частоты ДНК-маркера (q_i) в ней. Значение частоты q в каждом узле регулярной сетки p вычисляется с использованием процедуры средневзвешенной интерполяции [22]:

$$q(p) = \frac{\sum_i w_i q_i}{\sum_i w_i}, \quad w_i = 1 \Big/ sd(p, p_i)^k, \quad k > 0,$$

где суммирование ведется по всем опорным точкам p_i, попавшим в круг радиуса R с центром в точке p, а через sd (p, p_i) обозначено расстояние между точками p и p_i на сфере (координаты точки p в проекции при этом пересчитываются в широту и долготу). Для карт атласа моделирование выполнялось с параметрами $R = 3000$ км и $k = 3$.

Тематическое содержание карт включает непрерывное поле частот признака, показанное способом послойной окраски с наложением аналитической отмывки рельефа [23]; изображение рельефа совмещено с тематическим содержанием с использованием оригинального алгоритма [24], позволяющего визуально воспринимать цвета на карте с таким синтезированным изображением в соответствии с цветами шкалы. Базовые слои картографической основы едины для всех карт атласа и включают в себя береговую линию, объекты гидрографии (крупные реки и озера), государственные границы, столицы государств в векторном формате, рельеф в виде сеточной цифровой модели (источник – GMTED2010 [25]). Для всех карт атласа использована одна и та же цвето-

Таблица 2. Частоты генетических маркеров с высоким вкладом в определение цвета глаз и волос

Гены			*HERC2*	*HERC2*	*HERC2*	*TP63*	*SLC45A2*	*MC1R*	*HERC2*	*DEF8*	*EDAR*	*SPATA33*	*SLC45A2*	*RALY*	*LOC100996492*	*RUNX2*	*GLB1*
rs			1129038	12913832	12916300	6804480	35407	1805007	2238289	8051733	3827760	71396951	28777	6059655	12915889	2820339	72856153
Аллели			*T*	*G*	*T*	*T*	*G*	*T*	A^R	*G*	A^R	*C*	*A*	A^R	T^R	*G*	T^R
Суммарный генетический вклад (*C*)			**2.03**	**0.68**	**0.45**	**0.56**	**0.63**	**0.61**	**0.40**	**0.39**	**0.38**	**0.47**	**0.45**	**0.32**	**0.31**	**0.31**	**0.40**
Вклад в светлые глаза (C_E)			1.32	0.68	0.45	0.12	0.20	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.28
Вклад в светлые волосы (C_H)			0.71	0.00	0.00	0.44	0.41	0.40	0.40	0.39	0.36	0.33	0.32	0.32	0.31	0.31	–0.12
Вклад в рыжину волос (C_R)			0.01	0.00	0.00	0.00	0.02	0.21	0.00	0.00	0.02	0.13	0.05	0.00	0.00	0.00	0.00
№	Метапопуляция	*N*	Частота, %														
1	Камчатка	51	3	3	3	21	18	0	64	9	9	0	18	1	59	75	46
2	Амур	48	1	1	1	31	13	0	68	21	1	2	13	0	48	86	51
3	Тунгусы	54	2	2	2	19	12	0	56	40	12	0	12	0	71	74	50
4	Южная Сибирь	53	5	5	5	13	19	3	65	43	22	0	19	3	46	60	49
5	Алтае-Саяны	50	17	15	15	33	27	0	70	27	49	0	31	1	54	82	67
6	Монголия	50	6	4	6	30	22	1	58	23	33	4	24	4	66	74	57
7	Угры	51	46	46	45	24	74	3	69	23	50	3	73	4	57	64	61
8	Казахстан	72	13	14	13	32	44	1	61	21	49	1	42	2	55	69	66
9	Таджикистан	52	17	17	20	55	68	1	61	28	97	1	69	8	52	65	70
10	Западный Кавказ	61	27	28	29	31	90	7	69	23	93	8	91	4	59	63	81
11	Восточный Кавказ	50	36	36	37	44	90	4	72	27	99	2	90	3	49	67	83
12	Закавказье	49	30	30	31	35	71	2	73	30	97	2	71	0	55	60	88
13	Русский Север	51	79	79	81	44	100	5	97	35	96	2	100	5	61	64	88
14	Русские (центральные)	55	83	81	83	44	91	4	91	27	97	6	91	8	72	68	88
15	Зарубежная Европа	50	49	49	49	47	94	5	79	30	99	5	93	2	62	73	87
16	Западные финно-угры	50	89	93	91	39	98	7	100	34	95	6	98	1	69	66	93

Таблица 2. Окончание

Гены		*HERC2*		*HERC2*	*HERC2*	*TP63*	*SLC45A2*	*MC1R*	*HERC2*	*DEF8*	*EDAR*	*SPATA33*	*SLC45A2*	*RALY*	*LOC100996492*	*RUNX2*	*GLB1*
rs		1129038		12913832	12916300	6804480	35407	1805007	2238289	8051733	3827760	71396951	28777	6059655	12915889	2820339	72856153
Аллели		*T*		*G*	*T*	*T*	*G*	*T*	A^R	*G*	A^R	*C*	*A*	A^R	T^R	*G*	T^R
17	Восточные славяне	52	85	84	85	56	100	4	95	14	98	6	96	8	68	64	93
18	Мордовия	50	78	78	78	42	96	0	88	32	99	1	96	4	66	59	90
19	Тюрки Приуралья	50	62	62	56	33	77	3	83	21	81	3	77	1	65	64	85
20	Марий Эл	50	74	74	72	31	83	2	86	28	90	1	83	3	63	60	76
21	Пермь	50	71	71	71	45	93	13	86	37	93	13	93	5	74	72	75
22	Южный Урал	53	38	38	38	36	64	2	73	23	73	2	64	2	60	61	73
		Показатели частоты															
q_{MIN}		**1**		**1**	**1**	**13**	**12**	**0**	**56**	**9**	**22**	**0**	**12**	**0**	**46**	**59**	**46**
q_{MAX}		**89**		**93**	**91**	**56**	**100**	**13**	**100**	**43**	**99**	**13**	**100**	**8**	**74**	**86**	**93**
q_{MEAN}		**41**		**41**	**41**	**36**	**66**	**3**	**76**	**27**	**70**	**3**	**66**	**3**	**60**	**68**	**73**
		Генетическое разнообразие															
G_{ST}		**0.389**		**0.394**	**0.387**	**0.050**	**0.450**	**0.032**	**0.089**	**0.031**	**0.543**	**0.032**	**0.433**	**0.019**	**0.025**	**0.023**	**0.118**
H_T		0.485		0.485	0.485	0.459	0.451	0.059	0.368	0.395	0.423	0.059	0.451	0.060	0.478	0.437	0.390
H_S		0.296		0.294	0.298	0.436	0.248	0.057	0.335	0.383	0.193	0.057	0.256	0.058	0.466	0.426	0.344
Наличие в панели HIrusPlex		**H**		**H-S**				**H-S**	**H**	**H**			**H-S**	**H**			

Примечание. Приведены данные по 15 ДНК-маркерам, суммарный вклад которых в формирование светлых глаз, светлых волос, рыжины волос превышает 03. Обозначения для табл. 2, 3: № – номер метапопуляции на картах; *N* – размер выборки метапопуляции (популяционный состав указан в табл. 1); q_{MIN} – минимальное значение частоты; q_{MAX} – максимальное значение частоты; q_{MEAN} – среднее значение частоты; G_{ST} – показатель межпопуляционного разнообразия; H_T – показатель общего генетического разнообразия; H_S – показатель среднего внутрипопуляционного разнообразия; *H* – ДНК-маркер входит в зарубежную криминалистическую панель HIrusPlex; H-S – ДНК-маркер входит в зарубежную криминалистическую панель HIrusPlex-S.

Таблица 3. Частоты генетических маркеров с небольшим вкладом в определение цвета глаз и волос

Гены			*LOC105370627*	*PRDM16*	*ANKRD11*	*AFG3L1P*	–	*OCA2*	*MC1R*	*SLC24A5*	*FANCA*	*TYR*	*HERC2*	*CACNA2D3*	*TYR*	*C19orf45*	*SLC24A4*	–
rs			12896399	1999528	3114908	4238832	34191540	2871886	2228479	1426654	75570604	1393350	7494942	11283625	1042602	685034	12588868	11228497
Аллели			*T*	*G*	T^R	C^R	*T*	*C*	G^R	A^R	*C*	*A*	*G*	*A*	*A*	*C*	T^R	A^R
Суммарный генетический вклад (*C*)			**0.27**	**0.26**	**0.25**	**0.18**	**0.18**	**0.16**	**0.12**	**0.10**	**0.10**	**0.10**	**0.06**	**0.03**	**0.03**	**0.02**	**0.02**	**0.01**
Вклад в светлые глаза (C_E)			0.00	0.00	0.00	0.00	0.18	−0.02	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.02	0.00
Вклад в светлые волосы (C_H)			0.27	0.26	0.25	0.06	0.00	0.14	0.12	0.10	0.10	0.10	0.06	0.00	0.00	0.02	0.00	0.01
Вклад в рыжину волос (C_R)			0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
№	Метапопуляция	*N*	Частота, %															
1	Камчатка	51	27	49	40	1	69	74	92	20	1	0	45	28	1	60	24	69
2	Амур	48	33	32	33	0	19	76	78	5	1	0	50	25	1	50	16	82
3	Тунгусы	54	30	34	37	3	38	83	64	9	3	1	25	24	1	49	19	76
4	Южная Сибирь	53	11	46	48	6	49	50	67	44	10	4	42	28	3	67	16	68
5	Алтае-Саяны	50	28	30	33	2	53	49	74	66	4	2	55	35	8	54	27	78
6	Монголия	50	29	36	27	3	27	57	83	24	4	4	34	38	7	55	19	77
7	Угры	51	33	47	17	8	68	57	83	76	3	8	58	26	9	77	36	87
8	Казахстан	72	28	42	31	11	49	49	90	52	4	11	44	29	19	51	39	83
9	Таджикистан	52	33	42	20	22	57	49	93	92	1	18	38	29	28	52	36	88
10	Западный Кавказ	61	37	59	33	20	67	45	98	95	7	18	49	43	48	61	40	92
11	Восточный Кавказ	50	8	45	34	19	100	39	90	99	5	26	50	53	35	70	46	94
12	Закавказье	49	22	45	31	23	67	42	95	99	2	18	47	49	47	61	45	93
13	Русский Север	51	40	56	36	29	76	30	88	99	2	24	93	52	23	59	53	92
14	Русские (центральные)	55	42	57	34	25	64	25	88	96	5	23	84	38	22	61	60	91
15	Зарубежная Европа	50	37	64	21	22	63	25	90	100	6	19	70	40	43	68	49	93
16	Западные финно-угры	50	60	48	35	25	73	19	86	99	7	31	98	50	18	57	60	97

Таблица 3. Окончание

Гены		*LOC105370627*		*PRDM16*	*ANKRD11*	*AFG3L1P*	–	*OCA2*	*MC1R*	*SLC24A5*	*FANCA*	*TYR*	*HERC2*	*CACNA2D3*	*TYR*	*C19orf45*	*SLC24A4*	–
rs		12896399		1999528	3114908	4238832	34191540	2871886	2228479	1426654	75570604	1393350	7494942	11283625	1042602	685034	12588868	11228497
Аллели		*T*		*G*	T^R	C^R	*T*	*C*	G^R	A^R	*C*	*A*	*G*	*A*	*A*	*C*	T^R	A^R
17	Восточные славяне	52	30	52	24	18	77	25	93	99	5	22	86	51	26	59	54	91
18	Мордовия	50	46	43	35	19	100	31	82	98	0	31	84	37	29	65	54	96
19	Тюрки Приуралья	50	22	45	28	12	60	26	89	91	3	12	70	39	22	58	56	86
20	Марий Эл	50	42	43	32	28	78	24	91	97	13	30	82	29	20	51	48	84
21	Пермь	50	27	52	29	20	100	39	92	96	2	12	84	35	25	49	53	85
22	Южный Урал	53	27	43	38	18	57	38	89	83	3	14	61	38	21	53	38	87
		Показатели частоты																
q_{MIN}		**8**		**30**	**17**	**0**	**19**	**19**	**64**	**5**	**0**	**0**	**25**	**24**	**1**	**49**	**16**	**68**
q_{MAX}		**60**		**64**	**48**	**29**	**100**	**83**	**98**	**100**	**13**	**31**	**98**	**53**	**48**	**77**	**60**	**97**
q_{MEAN}		**32**		**46**	**32**	**15**	**64**	**43**	**86**	**75**	**4**	**15**	**61**	**37**	**21**	**58**	**40**	**86**
		Генетическое разнообразие																
G_{ST}		**0.055**		**0.028**	**0.022**	**0.067**	**0.183**	**0.125**	**0.060**	**0.546**	**0.023**	**0.080**	**0.178**	**0.036**	**0.118**	**0.022**	**0.085**	**0.053**
H_T		0.432		0.497	0.433	0.258	0.460	0.491	0.239	0.379	0.079	0.254	0.474	0.467	0.328	0.486	0.481	0.243
H_S		0.408		0.483	0.423	0.241	0.376	0.429	0.225	0.172	0.077	0.234	0.390	0.450	0.289	0.475	0.440	0.230
Наличие в панели HIrusPlex		**H-S**			**H**				**H-S**	**H**		**H-S**			**H-S**			

Примечание. Приведены данные по 16 ДНК-маркерам, суммарный вклад которых в формирование признаков светлых глаз, светлых волос и рыжины волос менее 0.3.

вая и числовая шкала, входящая в легенду каждой карты и создающая единство атласа.

Легенда карты включает также статистическую информацию: количество метапопуляций "K", минимальное "min", максимальное "max" и среднее "avr" значения признака, рассчитанные по цифровой модели карты. Аналогичные статистические показатели рассчитаны по информации табл. 2 и 3 и приведены в этих таблицах: минимальные q_{MIN}, максимальные q_{MAX} и средние q_{MEAN} частоты, показатель межпопуляционного разнообразия G_{ST}, общего генетического разнообразия H_T, среднего внутрипопуляционного разнообразия H_S. Средние картографированные частоты (указанные в легенде карты) закономерно отличаются от табличных (указанных в таблицах), поскольку учитывают важнейший параметр популяции – ее ареал, а ареалы метапопуляций резко различны, как различны ареалы этносов, входящих в метапопуляции, – от минимальных на Кавказе до огромных в Сибири. Поэтому карты несут больше информации, чем таблицы, но при интерпретации карт полезно ориентироваться на дополнительную информацию, приведенную в таблицах.

Все построенные карты (62 карты распространения отдельных аллелей и 6 карт показателей чувствительности) приведены в Приложении. Важнейшие карты представлены также на рисунках в статье.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для изучения распространения генетических маркеров цвета глаз выбраны 22 метапопуляции, охватывающие основное разнообразие народонаселения Северной Евразии, и сформирована выборка из 1153 образцов ДНК (в среднем 53 образца из метапопуляции). Их генотипирование по 31 SNP-маркеру, входящему в отечественную систему предикции цвета глаз и волос [18, 19], позволило определить частоты каждого маркера в каждой популяции (табл. 2, 3). На основании этих частот построены карты распространения каждого маркера, причем как референсного, так и альтернативного аллеля, итого 62 карты. В таблицах ДНК-маркеры расположены по степени убывания генетического вклада хотя бы в один из трех признаков. Для наиболее значимых маркеров – с вкладом в один из признаков больше 0.4 (табл. 2) – приведены карты их распространения (рис. 1–3). Для ДНК-маркеров с вкладом в один из признаков $0.3 < C < 0.4$ или с суммарным вкладом $C > 0.4$ дано краткое описание карт (полный комплект 62 карт приведен в Приложении).

Карты распространения ДНК-маркеров с наибольшим вкладом в цвет глаз и волос

Рассмотрение атласа следует начать с карт двух ДНК-маркеров с наибольшим вкладом в генетический контроль пигментации (рис. 1,*а, б*; табл. 2) – они оба относятся к гену HERC2, оба включены в криминалистическую панель HIrisPlex-S, средние частоты обоих приближаются к 50%, частота варьирует в почти максимальных пределах ($0 < q < 93\%$) и межпопуляционное разнообразие обоих велико ($G_{ST} = 0.39$). Но когда мы переходим к оценке их вклада в контроль пигментации, обнаруживаем значительные различия между ними. Суммарный вклад *HERC2_rs1129038_T* в пигментацию максимален ($C = 2.03$) и в 3 раза больше, чем вклад *HERC2_rs12913832_G* ($C = 0.68$). Еще существеннее различия в том, какие признаки они контролируют: если второй ДНК-маркер отвечает только за светлый цвет глаз ($C_E = 0.68$), то первый не тольков 2 раза мощнее контролирует светлый цвет глаз ($C_E = 1.31$), но крайне велико и его влияние на формирование светлого цвета волос ($C_H = 0.71$). Хотя многие другие ДНК-маркеры также контролируют два или даже все три признака пигментации, но *HERC2_rs1129038* уникален крайне большой величиной вклада в контроль пигментации ($C = 2.03$): она в 5 раз сильнее, чем в среднем у ДНК-маркеров, представленных в табл. 2, и в 17 раз сильнее, чем в среднем у ДНК-маркеров, представленных в табл. 3.

Но несмотря на столь существенные различия в контроле пигментации, карты обоих маркеров практически неотличимы (рис. 1,*а, б*) и можно дать их общее описание. Минимальные значения ($1 \le q \le 6\%$) охватывают коренное население Дальнего Востока, Монголии и Южной Сибири. На Алтае-Саянах, в Казахстане и Таджикистане они несколько выше ($13 \le q \le 17\%$). Еще выше частоты "осветления" на Кавказе, в Закавказье и в степях Южного Урала ($27 \le q \le 38\%$). С частотой выше средней эти ДНК-маркеры обнаружены в столь географически разобщенных популяциях, как обские угры ($q = 46\%$) и славяноязычные македонцы Южной Европы ($q = 49\%$). Еще выше частота ДНК-маркеров, контролирующих светлую пигментацию, в Приуралье у чувашей, марийцев и коми-пермяков ($62 \le q \le 74\%$). Максимальные значения охватывают ареал от Мордовии ($q = 78\%$) до западных финноязычных популяций карел и ижоры ($q = 93\%$). Популяции всех восточных славян занимают промежуточное положение ($79 \le q \le 85\%$) между западными и приуральскими финно-уграми.

Третий ДНК-маркер этого гена – *HERC2_rs12916300_T* – обладает только вкладом в формирование светлых глаз ($C_E = 0.45$) при отсутствии вклада в цвет волос. Однако карта его распространения столь неотличима от двух

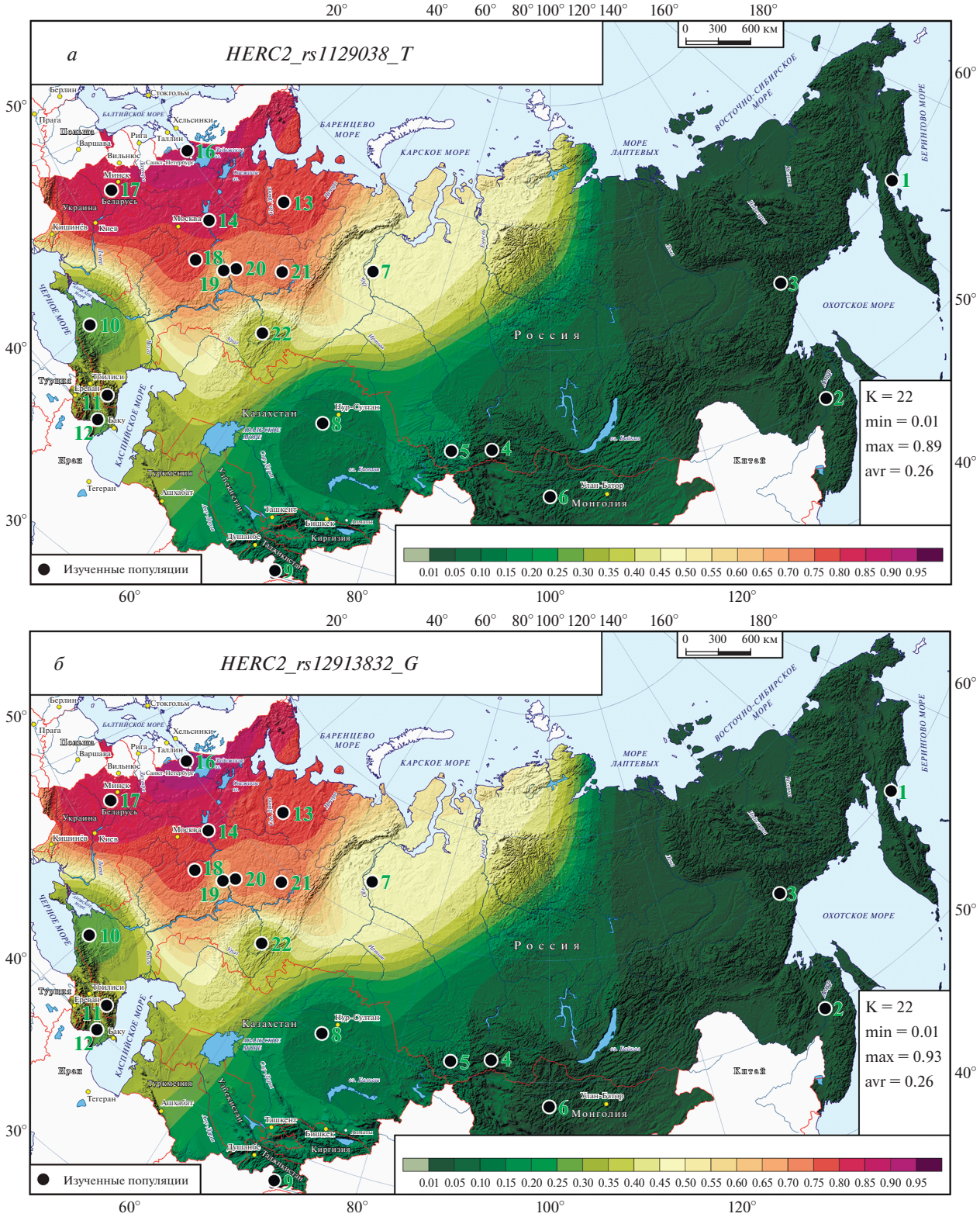

Рис. 1. Карты распространения ДНК-маркеров *HERC2_rs1129038_T* (*а*) и *HERC2_rs12913832_G* (*б*) в метапопуляциях России и сопредельных стран. Для рис. 1–3: номера метапопуляций соответствуют указанным в таблицах условным названиям метапопуляций; в статистической легенде K – число метапопуляций; min, max, avr – минимальное, максимальное и среднее значения картографированных (смоделированных) частот признаков; высокие значения частоты обозначены красно-коричневыми тонами, низкие – зелеными, шкала переходов приведена в легенде под картой; изученные популяции обозначены черными кружками.

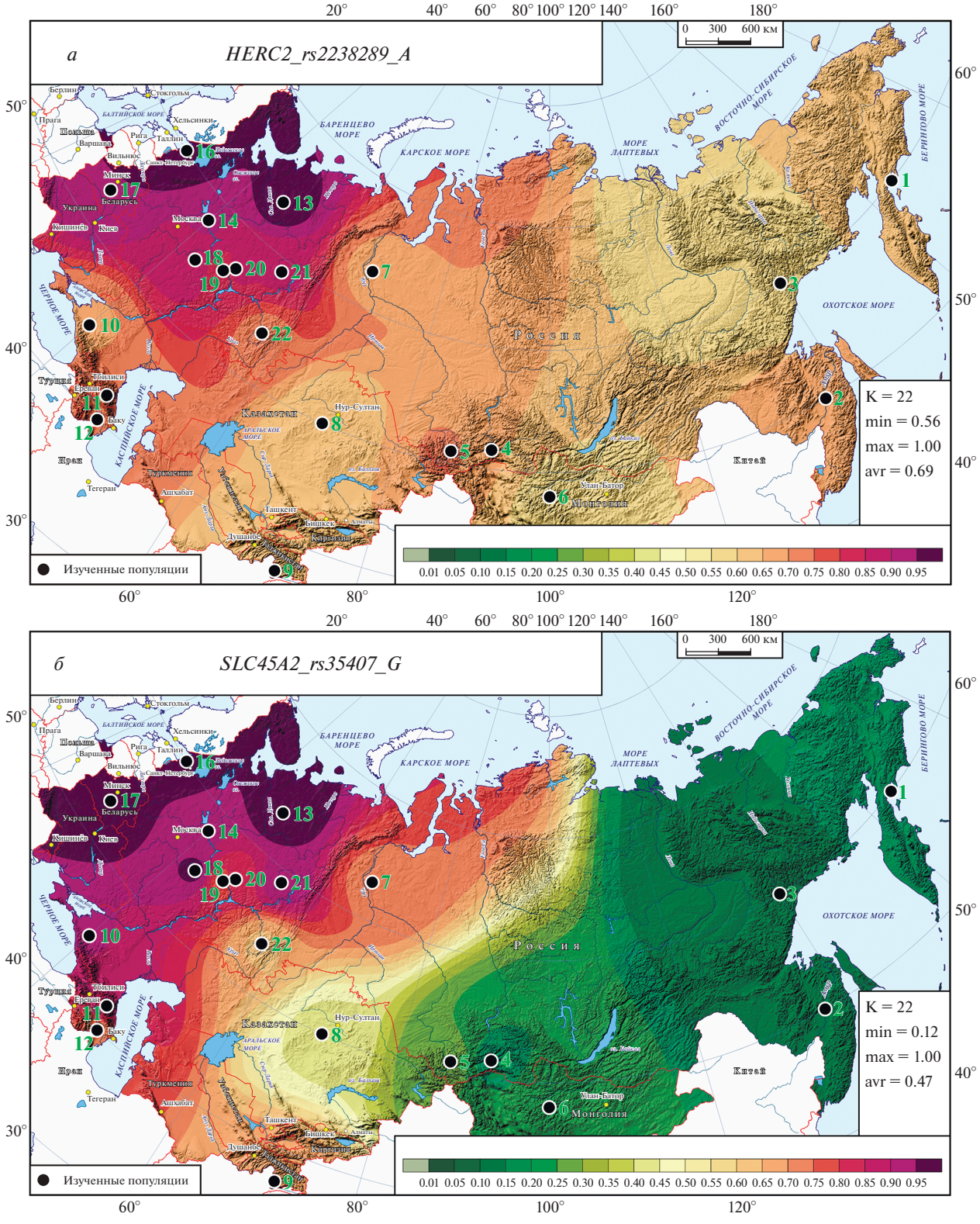

Рис. 2. Карты распространения ДНК-маркеров *HERC2_rs2238289_A* (*а*) и *SLC45A2_rs35407_G* (*б*) в метапопуляциях России и сопредельных стран.

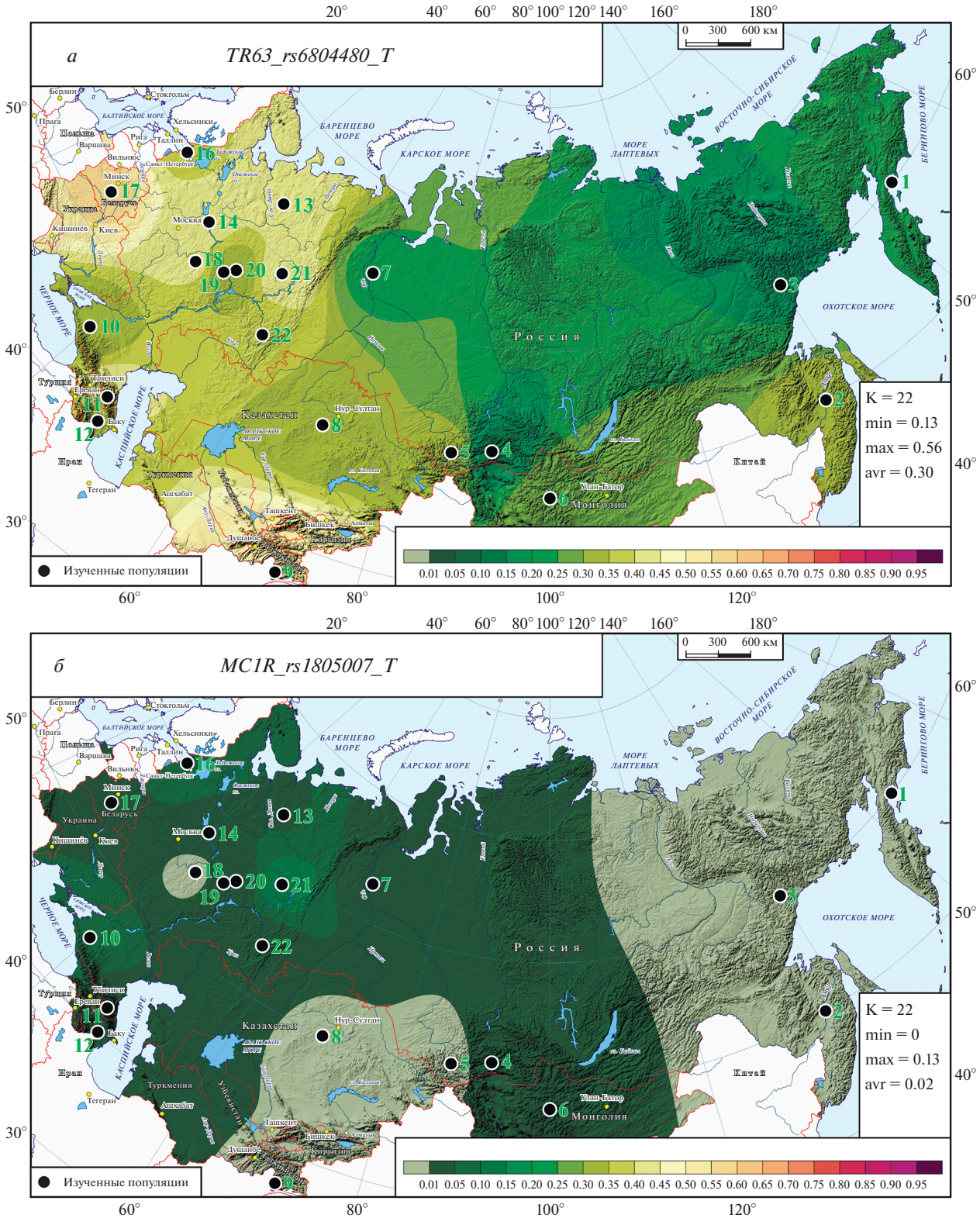

Рис. 3. Карты распространения ДНК-маркеров *TP63_rs6804480_T* (*а*) и *MC1R_rs1805007*_T (*б*) в метапопуляциях России и сопредельных стран.

предыдущих, что мы ее не приводим, тем более что почти полное сходство распространения всех трех ДНК-маркеров подтверждает табл. 2.

Четвертый ДНК-маркер гена *HERC2*, *HERC2_rs2238289_A* (рис. 2,*а*), вошедший в число семи наиболее важных генов пигментации, по ряду параметров своеобразен. Вклад его в пигментацию несколько меньше, но он входит в панель HIrusPlex (табл. 2) и в отличие от трех вышерассмотренных маркеров этого гена отвечает только за осветление волос ($C_H = 0.40$). Также в 2 раза ниже размах изменчивости его частоты ($56 \le q \le 100\%$) и в 5 раз меньше его межпопуляционное разнообразие ($G_{ST} = 0.089$). Карта его распространения (рис. 2,а) отражает эти различия: в ней нет такого буйства цвета, как на рис. 1: она окрашена в красноватые цвета высоких частот, указывающие на повсеместное распространение этого ДНК-маркера, отвечающего за осветление волос. В целом на карте сохраняется общий для всех маркеров, связанных со светлой пигментацией, градиент частоты увеличения ее с востока на запад. Но в распространении этого ДНК-маркера есть свои особенности. Например, минимальная частота *HERC2_rs2238289_A* характерна для тунгусов ($q = 56\%$, популяция № 3), в то время как у других народов Дальнего Востока частота несколько выше: на Камчатке $q = 64\%$, на Амуре $q = 68\%$. Хотя наблюдается тот же градиент увеличения частоты маркера от приуральских финнов ($q = 86\%$) к западным ($q = 100\%$), но среди популяций восточных славян наибольшая частота приходится на Русский Север ($q = 97\%$, популяция № 13).

Карта распространения *SLC45A2_rs35407_G* (рис. 2,*б*) на первый взгляд похожа на карты рис. 1: сходство возникает из-за столь же высокой изменчивости этого маркера – и частоты ($12 \le q \le 100\%$), и межпопуляционного разнообразия ($G_{ST} = 0.45$). Этот ДНК-маркер оказывает влияние на все три признака пигментации: и на осветление волос ($C_H = 0.40$), и на осветление глаз ($C_E = 0.20$), и даже небольшое влияние на рыжину волос ($C_R = 0.02$). Но при общем сходстве с картами рис. 1 распространение *SLC45A2_rs35407_G* (рис. 2,*б*) обнаруживает любопытные отличия. Во-первых, наблюдается более резкий переход от низких значений частоты к высоким. Область минимальных значений ($12 \le q \le 27\%$) простирается вплоть до Алтае-Саян. Переход к значениям выше среднего начинается лишь в Казахстане. Повышенные частоты теперь захватывают даже Таджикистан ($q = 68\%$). Область высоких значений частоты охватывает здесь не только Приуралье ($77 \le q \le 83\%$), но и обских угров ($q = 74\%$). Но наиболее ярко своеобразие этого ДНК-маркера проявляется на Кавказе: если на картах рис. 1 кавказские частоты колебались около $q = 30\%$, то на карте рис. 2,*б* они достигают $q = 90\%$. Обратим особое внимание и на высокую частоту ($q = 93\%$) у коми-пермяков (популяция № 21): если на картах рис. 1 она находилась в диапазоне между частотами у марийцев и чувашей, то теперь перешла в кластер наибольших значений ($91 \le q \le 100\%$), охватывающий популяции Мордовии (№ 18), западных финноязычных популяций и восточных славян, достигая максимума на Русском Севере (№ 13).

Особенность карт двух ДНК-маркеров, распространение которых представлено на рис. 3, связана с их низкой изменчивостью ($0.03 < G_{ST} < 0.05$). Однако их вклад в генетический контроль пигментации значителен (для гена *TP63* $C = 0.56$; для *MC1R* $C = 0.61$), поэтому детально рассмотрим карты их распространения.

ДНК-маркер *TP63_rs6804480_T* (рис. 3,*а*) оказывает влияние и на осветление волос ($C_H = 0.44$), и на осветление глаз ($C_E = 0.12$). Размах изменчивости его частот ($13 \le q \le 56\%$), всего в 2 раза меньше, чем *SLC45A2_rs35407_G* (рис. 2,*б*), но межпопуляционная изменчивость G_{ST} ниже почти в 10 раз, что отражено на карте (рис. 3,*а*) в сглаженном генетическом рельефе. Но при этом паттерны обеих карт (рис. 3,*а* и 2,*б*) в целом схожи. Обращает на себя внимание, что коми-пермяки (№ 21) вновь входят в кластер популяций с высокими частотами, превосходя по частоте и русские популяции, и мордовские. Незначительные отличия карты гена *TP63* связаны, например, с тем, что область минимальных значений достигает обских угров; Кавказ и Приуралье входят в область средних, а не высоких значений частоты; а в Таджикистане, напротив, очень высока частота этого маркера.

ДНК-маркер *MC1R_rs1805007_T* (рис. 3,*б*) демонстрирует паттерн карт с крайне низкими частотами и малой изменчивостью ($0 \le q \le 13\%$, $G_{ST} = 0.032$). При этом его генетический вклад в осветление волос велик ($C_H = 0.40$), а вклад в рыжину волос достигает максимальной величины ($C_R = 0.21$). Видимо, именно поэтому он включен в панель HIrusPlex-S (табл. 2). Генетический рельеф карты выражен очень слабо. Можно лишь отметить некоторое повышение частоты маркера на Русском Севере, у карел и на Западном Кавказе. Явным лидером вновь выступают коми-пермяки, у которых частота этого маркера, связанного с рыжиной волос, достигает максимума и в 6 раз выше средней частоты карты.

Рассмотрев семь ДНК-маркеров с наибольшим вкладом в контроль пигментации волос и глаз, мы увидели весь основной спектр паттернов пространственной изменчивости (см. Приложение, рис. П.1–П.31): от резко выраженного генетического рельефа с ярко обозначенными западным и восточным “полюсами” до крайне сглаженного рельефа. Поэтому для остальных ДНК-маркеров из табл. 2 (с вкладом в один из призна-

ков в интервале от 0.3 до 0.4) можно привести лишь краткое описание карт (напомним, что сами карты приведены в Приложении).

Карты распространения ДНК-маркеров с генетическим контролем цвета глаз и волос от 0.4 до 0.3

DEF8_rs8051733_G (рис. П.5), входящий в панель HIrisPlex-S, оказывает влияние только на осветление волос ($C_H = 0.39$). При средней величине диапазона частот ($9 \le q \le 43\%$) очень мала его межпопуляционная изменчивость ($G_{ST} = 0.031$). Поэтому на карте нет даже слабовыраженного градиента "восток–запад": минимальные частоты приходятся на народы Камчатки (коряки и чукчи) и на восточных славян (украинцев и белорусов), максимальные – на тунгусов Дальнего Востока, народы Южной Сибири, коми-пермяков и Русский Север.

EDAR_rs3827760_A (рис. П.6) оказывает такое же влияние на осветление волос ($C_H = 0.38$) и слабый вклад в рыжину $C_R = 0.02$. Но размах изменчивости ($22 \le q \le 99\%$) и особенно межпопуляционная изменчивость ($G_{ST} = 0.54$) огромны. Паттерн карты в целом схож с представленными на рис. 1,*а, б* и 2,*б*. Градиент частоты на карте выражен удивительно ярко: на Дальнем Востоке частота мала ($1 \le q \le 12\%$), а уже начиная с Приуралья и далее на запад ДНК-маркер достигает максимальный частоты ($90 \le q \le 99\%$). Но не менее яркой особенностью распространения этого ДНК-маркера, контролирующего светлый цвет волос, являются максимальные частоты на Кавказе ($93 \le q \le 99\%$) и в Таджикистане ($q = 97\%$). При этом в Казахстане его частота в 2 раза ниже ($q = 49\%$), а в Южной Сибири падает еще в 2 раза ($q = 22\%$).

SPATA33_rs71396951_C (рис. П.28) схож с ДНК-маркером *MC1R_rs1805007_T* (рис. 3,*б*) не только крайне низкими частотами и малой изменчивостью ($0 \le q \le 13\%$, $G_{ST} = 0.032$), но и генетическим вкладом как в осветление волос ($C_H = 0.33$), так и вторым по значимости вкладом в их рыжину ($C_R = 0.13$). Карты их распространения очень похожи: очень слабый градиент увеличения частоты с Дальнего Востока на запад, но самые высокие частоты вновь приходятся на коми-пермяков.

SLC45A2_rs28777_A (рис. П.26) обладает тем же паттерном, как и другой ДНК-маркер этого гена (рис. 2,*а*), при столь же высокой изменчивости ($12 \le q \le 100\%$; $G_{ST} = 0.433$). Хотя его вклад в пигментацию заметно ниже за счет меньшего влияния на цвет глаз, но и он оказывает влияние на все три анализируемых признака: и на осветление волос ($C_H = 0.32$), и на осветление глаз ($C_E = 0.08$), и на рыжину волос ($C_R = 0.05$). Карты распространения обоих аллелей очень схожи: те же закономерности при небольшом различии частот.

RALY_rs6059655_A (рис. П.20) оказывает такое же влияние на осветление волос ($C_H = 0.32$), как и предыдущий маркер, но его генетический рельеф отличается кардинально: эти два маркера представляют собой два полюса паттернов – максимальной и минимальной изменчивости. Но даже при столь низкой вариабельности *RALY_rs6059655_A* ($0 \le q \le 8\%$; $G_{ST} = 0.02$) на его карте все же есть градиент возрастания частоты от Дальнего Востока к европейской части России. При этом максимальные частоты приходятся как на восточных славян (русских, украинцев, белорусов), так и на Таджикистан.

LOC100996492_rs12915889_T (рис. П.14) вновь влияет только на осветление волос ($C_H = 0.31$). Несмотря на то что его частоты колеблются около 0.6, межпопуляционная изменчивость очень мала ($G_{ST} = 0.03$). Своеобразие генетического рельефа его карты проявляется в повышении частоты не только на западе Северной Евразии, но и на востоке (у тунгусов Дальнего Востока), и на юге (монголы).

RUNX2_rs2820339_G (рис. П.23) вторит по основным показателям предыдущему маркеру: он также связан только с осветлением волос ($C_H = 0.31$), частоты колеблются около 0.7, межпопуляционная изменчивость столь же мала ($G_{ST} = 0.02$). Но его пространственное распределение резко отличается от всех остальных ДНК-маркеров пигментации: максимальные частоты сосредоточены на востоке ($74 \le q \le 86\%$) и на юге (монголы, алтайцы; $74 \le q \le 82\%$) Северной Евразии. В то время как в Европе наибольшие значения обнаруживаются лишь у македонцев и коми-пермяков ($72 \le q \le 73\%$). Такое же нестандартное распределение – с максимумом на востоке и минимумом на западе демонстрирует карта еще одного гена – *OCA2_rs2871886_C* (рис. П.18). Он имеет небольшой положительный вклад (табл. 3) в осветление волос ($C_H = 0.14$) и слабый вклад в темные оттенки цвета глаз ($C_E = -0.02$). Однако, несмотря на высокую частоту на востоке и положительный вклад в светлые оттенки волос, ни одному из образцов на востоке Евразии, обладающих этими ДНК-маркерами, не были предсказаны светлые волосы. Это означает, что на востоке при генетическом определении цвета волос больший вес имеют другие ДНК-маркеры. Видимо, эти два ДНК-маркера улучшают предикцию светлых волос на западе, при этом практически не влияя на предикцию цвета волос на востоке ареала.

GLB1_rs72856153_T (рис. П.8) отличается от большинства других ДНК-маркеров альтернативным генетическим контролем пигментации волос и глаз: осветление цвета глаз ($C_E = 0.28$), но потемнение цвета волос ($C_H = -0.12$). При этом хорошо выражены и межпопуляционная изменчивость ($G_{ST} = 0.12$), и генетический рельеф: мини-

мальные частоты на востоке и юге ($46 \le q \le 66\%$), максимальные у всех восточных славян, в Мордовии и у западных финно-угров ($90 \le q \le 93\%$), причем область высоких частот полностью захватывает и весь Кавказ ($81 \le q \le 88\%$).

Карты распространения ДНК-маркеров с небольшим генетическим контролем цвета глаз и волос ($C < 0.3$)

Характеристики карт распространения остальных 16 ДНК-маркеров с меньшим вкладом в генетический контроль пигментации даны в табл. 3. Не входя в описание каждой карты, укажем принадлежность к уже описанным паттернам распространения в популяциях России и сопредельных стран и рассмотрим только наиболее яркие отклонения от этих паттернов.

Паттерн огромного диапазона частот, отраженный на рис. 2,*б*, еще ярче выражен на карте распространения ДНК-маркера *SLC24A5_rs1426654_A* (рис. П.25): при минимуме на Дальнем Востоке ($5 \le q \le 9\%$) максимальные частоты ($95 \le q \le 100\%$) охватывают не только всю Восточно-Европейскую равнину, но и весь Кавказ, волной высоких частот захватывая обских угров ($q = 76\%$) на севере и Таджикистан на юге ($q = 92\%$). Другие ДНК-маркеры – *SLC24A4_rs12588868_T* (рис. П.24) и *HERC2_rs7494942_G* (рис. П.11) – демонстрируют уже намного более сглаженный генетический рельеф карты, но западный и восточный "полюса" сохраняют свои ареалы. На остальных картах с очень низкой межпопуляционной изменчивостью также обнаруживаются эти два полюса, но ареалы их очень размыты – настолько, что на карте *FANCA_rs75570604_C* (рис. П.7) с малым диапазоном частоты ($0 \le q \le 13\%$) уже практически исчезают.

Многие их этих карт демонстрируют любопытные отклонения от общего паттерна. Например, на карте *TYR_rs1042602_A* (рис. П.30) при выраженной межпопуляционной изменчивости ($G_{ST} = 0.12$) и при четком градиенте роста частоты "восток–запад" максимальные частоты ярко проявляются только на Кавказе ($43 \le q \le 48\%$). На карте ДНК-маркера *C19orf45_rs685034_C* (рис. П.3), отличающегося самой низкой межпопуляционной изменчивостью ($G_{ST} = 0.02$), "западный полюс" вновь наиболее ярко проявляется на Кавказе ($61 \le q \le 70\%$). При этом максимальный всплеск частоты ($q = 77\%$) неожиданно приходится на центральную часть Северной Евразии – на обских угров. На другой карте (*ANKRD11_rs3114908_T*, рис. П.2), напротив, именно на обских угров приходятся минимальные значения ($q = 17\%$), а максимумы ($35 \le q \le 48\%$) разбросаны географически хаотично, приходясь не только на западных финно-угров, русские и мордовские популяции, но и на Южную Сибирь, и на Дальний Восток.

На карте *CACNA2D3_rs11283625_A* (рис. П.4) область максимальных частот своеобразно охватывает западной дугой популяции от Русского Севера до Восточного Кавказа, оставляя Поволжье и Приуралье в области более низких частот. А на карте *rs34191540_T* (рис. П.22), демонстрирующей значительное генетическое разнообразие ($19 \le q \le 100\%$; $G_{ST} = 0.18$), зона высоких частот, достигающая максимума ($q = 100\%$) в Приуралье и на Восточном Кавказе, заходит за Урал, включая в себя не только обских угров ($q = 68\%$), но неожиданно появляясь даже на Камчатке ($q = 69\%$). Так что для минимальных значений остается только юго-восточный ареал – от Амура ($q = 19\%$) до Монголии ($q = 27\%$).

Географическое распределение чувствительности предикции цвета глаз и волос

Чувствительность предикции определяется как отношение числа образцов, которым по генотипу был предсказан определенный фенотип (и которые действительно им обладают), к общему числу индивидов с данным фенотипом. Поскольку для этого надо иметь не только образцы ДНК, но и качественные цветные фотографии представителей коренного населения России, выборка для анализа чувствительности меньше, чем для основной части атласа: она включила 11 региональных популяций, для каждой из которых рассчитаны показатели чувствительности (для бинарных признаков они характеризуют и специфичность альтернативного фенотипа). Шесть карт (чувствительности предикции шести фенотипов) представлены в Приложении (рис. П.32–П.34). На рис. 4 приведены две из шести карт – для чувствительности предикции темных глаз и темных волос, поскольку эти фенотипы (в отличие от светлой пигментации) распространены повсеместно.

Чувствительность предикции признака "темный цвет глаз" (рис. 4,*а*) очень высока, составляя в среднем по 96% (по картографированным значениям – 98%). Минимальные значения приходятся на популяции русских и Мордовии ($p = 76\%$). Немногим выше показатель у тюрок Приуралья ($p = 89\%$) и марийцев ($p = 93\%$). В остальных региональных популяциях чувствительность максимально высока: $98 < p < 100\%$.

Чувствительность предикции признака "темный цвет волос" (рис. 4,*б*) ненамного ниже: в среднем по популяциям 88% (по картографированным значениям – 92%). Минимальные значения вновь приходятся на популяции русских и Мордовии ($p = 69\%$), но также и на коми-пермяков ($p = 70\%$). Несколько снижен ($82 < p < 89\%$) показатель вновь у тюрок Приуралья и у марий-

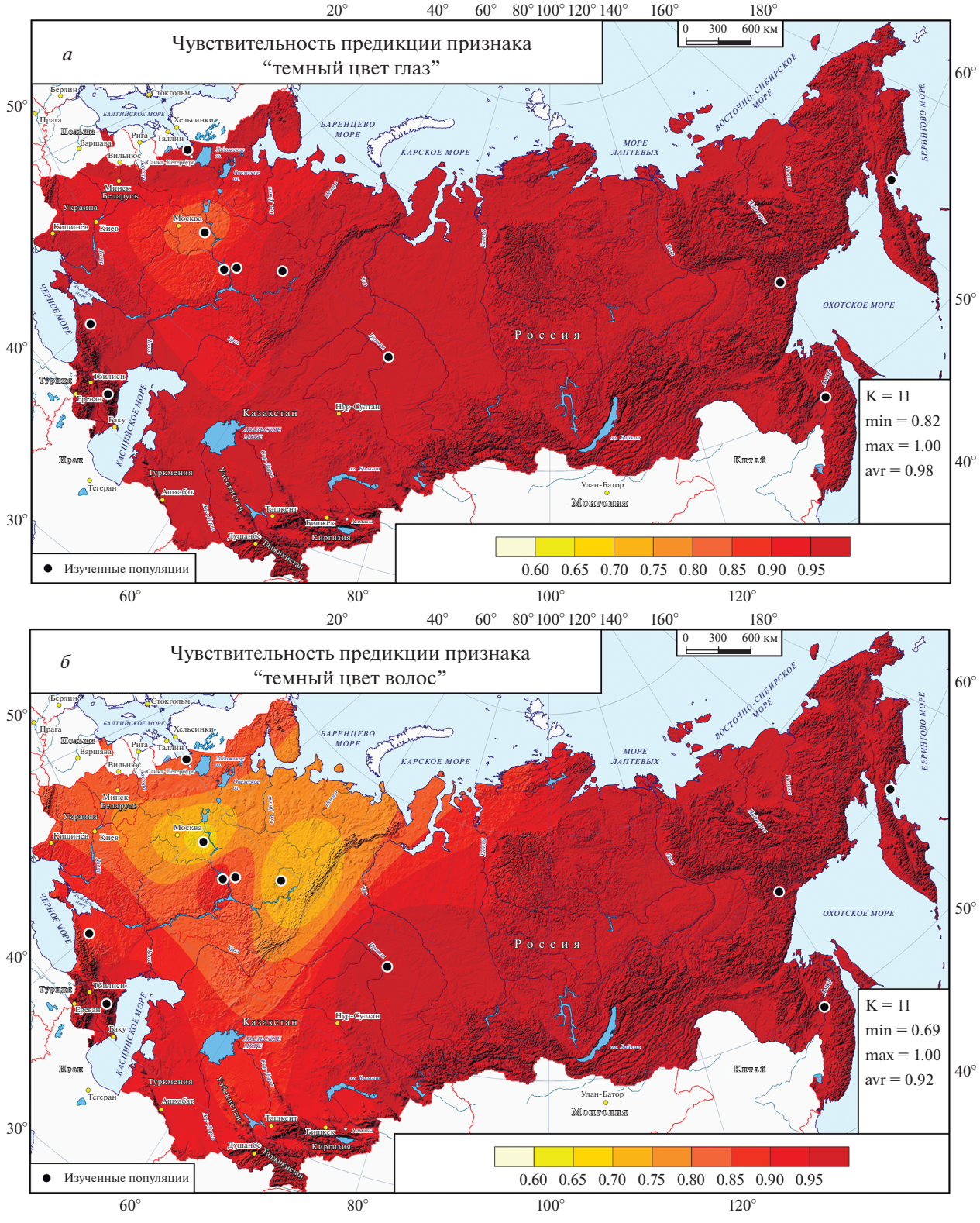

Рис. 4. Карты региональной чувствительности предикции признаков “темный цвет глаз” (*а*) и “темный цвет волос” (*б*). Предикция проведена по всей совокупности ДНК-маркеров для каждой региональной популяции с известными фенотипами. В статистической легенде K – число региональных популяций с известными фенотипами; min, max, avr – минимальное, максимальное и среднее значения картографированных (смоделированных) частот признаков; высокие значения частоты обозначены красно-коричневыми тонами, шкала переходов приведена в легенде под картой; изученные популяции обозначены черными кружками.

цев, но также у западных финно-угров и на Западном Кавказе. В остальных метапопуляциях чувствительность высока: $93 < p < 100\%$.

Чувствительность предикции "отсутствие рыжины волос" имеет очень сходный паттерн: в среднем по популяциям 84% (по картографированным значениям – 87%). Минимальные значения приходятся вновь на коми-пермяков ($p = 60\%$) и на популяции русских и Мордовии ($p = 67\%$). Несколько снижен ($71 < p < 85\%$) показатель вновь у тюрок Приуралья, Марий Эл и на всем Кавказе. В остальных метапопуляциях чувствительность высока: $90 < p < 100\%$.

Рассмотрение карт чувствительности предикции цвета глаз и волос указывает, прежде всего, на высокий уровень генетического контроля пигментации при использовании созданной нами панели 31 ДНК-маркера. Но также он выявляет те регионы, для которых характерен своеобразный генетический комплекс контроля пигментации – это коми-пермяки, народы Кавказа и неожиданно выявившийся регион русских популяций и Мордовии. В отношении угров, к сожалению, пока нет достаточной информации.

Рассмотрение всей совокупности карт атласа позволяет выделить ряд общих закономерностей. Основной из них является изменение частот в направлении "восток–запад". Этот паттерн может проявляться ярче или слабее, но присутствует почти на всех картах. Ядром "восточного полюса" практически всегда выступают популяции Дальнего Востока, к которым обычно присоединяются популяции Центральной Азии, но в более ослабленном варианте. Ядром "западного полюса", как правило, выступают популяции всех восточных славян (русских, украинцев, белорусов), западных финно-угров и мордвы. Ожидаемо, что к западу возрастают обычно частоты аллеля, определяющего светлую пигментацию глаз и/или волос, но выявлены и два случая, когда частота "светлого" аллеля возрастала к востоку.

Второй чертой, выявленной в результате картографического анализа, является особое положение населения Кавказа: по одним ДНК-маркерам оно полностью входит в "западный полюс", порой даже преобладая в нем; по другим ДНК-маркерам – резко отделяется от "западного" полюса. Ранее уже высказывалось предположение [17], что для населения Кавказа могут быть характерны особенности набора аллелей, контролирующих пигментацию волос и глаз. К тому же разные регионы Кавказа могут различаться: по одним ДНК-маркерам население Кавказа выступает единым блоком, по другим – регионы Западного, Восточного Кавказа и Закавказья обладают собственным генетическим рельефом. Это указывает на одно из возможных направлений для будущих исследований: изучение генетического контроля пигментации населения Кавказа требует не только особого, но и дифференцированного анализа.

Своеобразное положение на ряде карт занимают также две популяции финно-угорского мира – обские угры и коми-пермяки. В целом угры (ханты и манси) занимают четко выраженную промежуточную зону между двумя полюсами, часто ее и формируя. Но на ряде карт они присоединяются к "западному полюсу", на других – к "восточному", на третьих – формируют собственное ядро экстремальных частот. Поскольку обские угры – единственные в нашем популяционном массиве представители коренного населения Западной Сибири, возможно, этот регион обладает особым составом генетического контроля пигментации и требует отдельного изучения. Но если особое положение угров вполне закономерно, то своеобразие генофонда коми-пермяков неожиданно. На многих картах они занимают особое положение не только среди своих географических и лингвистических соседей, но и среди всего массива популяций Северной Евразии.

Третьей чертой, неожиданно проявившейся при совместном статистическом и картографическом анализе, является отсутствие видимой связи между уровнем изменчивости аллеля и его вкладом в пигментацию (коэффициент корреляции Спирмена $\rho = 0.2$ недостоверен): среди маркеров, вносящих основной вклад в формирование фенотипа, присутствуют как аллели, частоты которых резко различаются между популяциями, так и те, частоты которых распределены среди народонаселения Северной Евразии почти равномерно. Аналогичная закономерность наблюдается и среди генов, слабо участвующих в контроле пигментации.

Мы благодарим всех доноров образцов, которые принимали участие в данном исследовании. Коллекция образцов ДНК и антропологических фотографий была предоставлена АНО "Биобанк Северной Евразии".

Исследование проведено при финансовой поддержке Научно-технической программы Союзного государства "ДНК-идентификация" (генотипирование, картографирование, написание текста статьи) и Государственного задания для Медико-генетического научного центра (популяционно-генетический анализ межпопуляционной изменчивости).

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

1. *Morgan M., Pairo-Castineira E., Rawlik K. et al.* Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability // Nat. Communications. 2018. V. 9. P. 5271. https://doi.org/10.1038/s41467-018-07691-z
2. *Simcoe M., Valdes A., Liu F. et al.* Genome-wide association study in almost 195000 individuals identifies 50 previously unidentified genetic loci for eye color // Sci. Advances. 2021. V. 7. № 11. P. eabd1239. https://doi.org/10.1126/sciadv.abd1239
3. *Jablonski N., Chaplin G.* The colours of humanity: the evolution of pigmentation in the human lineage // Phil. Trans. R. Soc. B. 2017. V. 372. P. 20160349. https://doi.org/10.1098/rstb.2016.0349
4. *Crawford N., Kelly D., Hansen M. et al.* Loci associated with skin pigmentation identified in African populations // Science. 2017. V. 358. P. eaan8433. https://doi.org/10.1126/science.aan8433
5. *Lazaridis I., Mittnik A., Patterson N. et al.* Genetic origins of the Minoans and Mycenaeans // Nature. 2017. V. 548. P. 214–218. https://doi.org/10.1038/nature23310
6. *Clemente F., Unterla M., Dolgova O. et al.* The genomic history of the Aegean palatial civilizations // Cell. 2021. V. 184. P. 1–22. https://doi.org/10.1016/j.cell.2021.03.039
7. *Guenther T., Malmström H., Svensson E. et al.* Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation // PLoS Biology. 2018. V. 16. № 1. P. e2003703. https://doi.org/10.1371/journal.pbio.2003703
8. *Jones E., Zarina G., Moiseyev V. et al.* The neolithic transition in the Baltic was not driven by admixture with early european farmers // Curr. Biology. 2017. V. 27. P. 576–582. https://doi.org/10.1016/j.cub.2016.12.060
9. *Mathieson I., Lazaridis I., Rohland N. et al.* Eight thousand years of natural selection in Europe // bioRxiv. 2015. 016477. https://doi.org/10.1101/016477
10. *Brunel S., Bennett A., Cardin L. et al.* Ancient genomes from present-day France unveil 7000 years of its demographic history // Proc. Natl Acad. Sci. 2020. V. 117. № 23. P. 12791–12798. https://doi.org/10.1073/pnas.1918034117
11. *Feldman M., Fernández-Domínguez E., Reynolds L. et al.* Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia // Nat. Communications. 2019. V. 10. P. 1218. https://doi.org/10.1038/s41467-019-09209-7
12. *Saag L., Vasilyev S., Varul L. et al.* Genetic ancestry changes in Stone to Bronze Age transition in the East European plain // Sci. Advances. 2021. V. 7. № 4. P. eabd6535. https://doi.org/10.1126/sciadv.abd6535
13. *Walsh S., Liu F., Ballantyne K.N. et al.* IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information // Forensic Sci. Intern.: Genetics. 2011. V. 5. № 3. P. 170–180. https://doi.org/10.1016/j.fsigen.2010.02.004
14. *Walsh S., Liu F., Wollstein A. et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA // Forensic Sci. Intern.: Genetics. 2013. V. 7. № 1. P. 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005
15. *Chaitanya L., Breslin K., Zuñiga S. et al.* The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation // Forensic Sci. Intern.: Genetics. 2018. V. 35. P. 123–135. https://doi.org/10.1016/j.fsigen.2018.04.004
16. *Балановская Е.В., Рычков Ю.Г.* Этническая генетика: Этногеографическое разнообразие народов мира // Генетика. 1990. Т. 26. № 1. С. 114–121.
17. *Балановский О.П., Петрушенко В.С., Горин И.О. и др.* Точность предикции пигментации волос и глаз по генетическим маркерам для популяций России // Вестник Рос. гос. мед. ун-та. 2019. № 5. С. 25–41. https://doi.org/10.24075/vrgmu.2019.069
18. *Balanovska E., Lukianova E., Kagazezheva J. et al.* Optimizing the genetic prediction of the eye and hair color for North Eurasian populations // BMC Genomics. 2020. V. 21. № 7. P. 527. https://doi.org/10.1186/s12864-020-06923-1
19. *Балановская Е.В., Горин И.О., Балановский О.П.* Способ определения вероятности цвета глаз индивидуума, происходящего из популяций России, и панель однонуклеотидных полиморфизмов. Заявка на патент № 2020137355 от 13.11.2020.
20. *Берлянт А.М.* Картография: Учебник для вузов. М.: Аспект Пресс, 2002. 336 с.
21. *Балановская Е.В., Жабагин М.К., Агджоян А.Т. и др.* Популяционные биобанки: принципы организации и перспективы применения в геногеографии и персонализированной медицине // Генетика. 2016. Т. 52. № 12. С. 1371–1387. https://doi.org/10.7868/S001667581612002X
22. *Сербенюк С.Н., Кошель С.М., Мусин О.Р.* Методы моделирования геополей по данным в нерегулярно расположенных точках // Геодезия и картография. 1990. № 11. С. 31–35.
23. *Востокова А.В., Кошель С.М., Ушакова Л.А.* Оформление карт. Компьютерный дизайн: Учебник / Под ред. Востоковой А.В. М.: Аспект Пресс, 2002. 288 с.
24. *Koshel S.M., Nikiforov S.L., Koshel A.S.* Experiments on sea bottom relief modelling and cartographic representation: White sea case study // Proc. AutoCarto 2012 Intern. Symp. on Automated Cartography, Columbus (OH, USA), 16–18 Sept. 2012. Columbus, Ohio. 2012. https://doi.org/10.13140/RG.2.1.4353.5205
25. *Danielson J.J., Gesch D.B.* Global multi-resolution terrain elevation data 2010 (GMTED2010). Open-File Report 2011-1073. 2011. https://doi.org/10.3133/ofr20111073

Gene Geographic Atlas of DNA Markers Controlling the Color of Human Eyes and Hair

E. V. Balanovska[a, b, *], I. O. Gorin[a, c], S. M. Koshel[a, d], and O. P. Balanovsky[a, b, c]

[a] *Scientific Institution Research Centre of Medical Genetics, Moscow, 115478 Russia*

[b] *Biobank of North Eurasia, Moscow, 115215 Russia*

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

[d] *Lomonosov Moscow State University, Moscow, 119991 Russia*

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

The study of the genetic control of pigmentation has identified many DNA markers that affect the human eye and hair color formation. The foreign panel of markers HIrisPlex was developed based on the genetic diversity of the population of Western Europe and had a slightly reduced accuracy in Russian populations. In contrast, the domestic panel of markers, including 31 SNPs, was developed on Russian samples. This work studied the distribution of these markers in the populations of Russia and neighboring countries. Each of the 31 markers included in the national panel was genotyped in representatives of 22 populations, covering the main diversity of the gene pools of Northern Eurasia, for a total sample of 1153 people. Based on the obtained allelic frequencies, a cartographic atlas was created, which included maps of the distribution of 62 alleles, as well as maps of the sensitivity of the phenotype prediction by genotype. Cartographic and statistical analysis revealed three main patterns of the genetic landscape of pigmentation markers. First, the main pattern of most maps is the frequency gradient in the east-west direction. In most cases (but there are exceptions), the allele frequency that determines the light pigmentation of the eyes and hair increases to the west. Second, there is no relationship between the value of the genetic contribution of the DNA marker to pigmentation and its interpopulation variability. Thirdly, the genetic peculiarity of the populations of the Caucasus and the Urals was revealed, in which the frequencies of genetic markers of pigmentation are irreducible to either the “western” or the “eastern” type. Prediction sensitivity maps of pigmentation also revealed the uniqueness of Russian populations.

Keywords: gene geography, atlas, pigmentation, Northern Eurasia, DNA markers, prediction, maps.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1376–1385

УДК 575.113.1:575.21:577.24

ОПРЕДЕЛЕНИЕ ХРОНОЛОГИЧЕСКОГО ВОЗРАСТА ЧЕЛОВЕКА ПО БИОЛОГИЧЕСКИМ ОБРАЗЦАМ НА ОСНОВАНИИ АНАЛИЗА МЕТИЛИРОВАНИЯ CpG-ДИНУКЛЕОТИДОВ

© 2021 г. В. А. Лемеш[1, *], В. Н. Кипень[1, **], М. В. Богданова[1], А. А. Буракова[1], А. Г. Булгак[2], А. В. Байда[3], С. А. Брускин[4], О. В. Зотова[2], О. И. Добыш[1]

[1]*Институт генетики и цитологии Национальной академии наук Беларуси, Минск, 220072 Республика Беларусь*

[2]*Республиканский научно-практический центр "Кардиология", Минск, 220089 Республика Беларусь*

[3]*Белорусская медицинская академия последипломного образования, Минск, 220013 Республика Беларусь*

[4]*Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия*

**e-mail: v.lemesh@igc.by*

***e-mail: v.kipen@igc.by*

Поступила в редакцию 27.05.2021 г.
После доработки 17.06.2021 г.
Принята к публикации 22.06.2021 г.

На основании биоинформатического и статистического анализов GEO-проектов по определению полногеномного профиля метилирования ДНК человека сформирован перечень из 41 CpG-динуклеотида с высоким прогностическим потенциалом для создания моделей предсказания возраста человека по биологическим образцам. Определен уровень метилирования для 1208 образцов индивидов из Республики Беларусь (275 – кровь, 466 – буккальный эпителий, 467 – сперма), рассчитаны коэффициенты корреляции *R* и построены математические модели определения возраста человека. Среднее значение точности предсказания возраста по образцам крови с использованием 12 CpG-динуклеотидов составило 3.4 года (для мужчин – 3.3, для женщин – 3.5); по образцам буккального эпителия с использованием 6 CpG-динуклеотидов – 4.6 года (для мужчин – 4.5, для женщин – 4.7); по образцам спермы с использованием 5 CpG-динуклеотидов – 3.0 года. Полученные результаты будут положены в основу при разработке калькуляторов предсказания возраста индивида по биологическим следам для экспертов-криминалистов.

Ключевые слова: криминалистика, определение возраста, кровь, буккальный эпителий, сперма, метилирование ДНК, human methylation 450K BeadChip.

DOI: 10.31857/S0016675821120092

Определение возраста человека по образцам биологических жидкостей и фрагментам тканей играет важную роль в криминалистической практике: способствует ограничению поисков при идентификации останков, сужению круга подозреваемых, тем самым позволяя сэкономить время, которое зачастую является лимитирующим фактором в процессе расследования. В связи со значимостью проблемы обеспечения безопасности граждан Союзного государства (Республика Беларусь и Российская Федерация) разработка методик определения возраста несомненно актуальна и востребована.

В настоящее время в судебно-медицинской экспертизе для оценки возраста человека применяются морфологический (характеристика внешних особенностей зубов или частей скелета) и биохимический (возрастание рацемизации аспарагиновой кислоты в дентине при старении) методы. Однако ошибка при использовании этих методов довольна высока (3–10 лет). Кроме того, при минимальном количестве биологического материала или высокой степени его деградации их применение является затруднительным или невозможным. В таких случаях незаменимыми становятся методы, в которых матрицей для исследования выступает наследственный материал клетки.

Сравнение характеристик основных молекулярно-генетических подходов для определения возраста индивида представлено в обзоре А.Д. Золотаренко и др. [1]. Для решения криминалистической задачи по определению хронологического возраста человека наиболее воспроизводимым, линейно-зависимым, чувствительным и специфичным является подход, основанный на выявлении уровня метилирования ДНК в конкретных CpG-динуклеотидах [2–4].

Самыми распространенными биологическими образцами криминалистического характера являются следы крови, буккального эпителия и спермы. В мире предложено значительное количество методик определения возраста индивида по уровню прижизненной химической модификации ДНК [5–19]. Однако эффективность применения указанных методик для определения возраста представителей населения регионов Союзного государства не установлена и требует проработки, поскольку было продемонстрировано, что уровень метилирования CpG-динуклеотидов может различаться в зависимости от этногеографического происхождения индивидов и факторов окружающей среды [20, 21].

Цель данной работы – оценить точность предсказания возраста человека с использованием математических моделей на основании анализа метилирования CpG-динуклеотидов в образцах крови, буккального эпителия и спермы индивидов из Республики Беларусь.

МАТЕРИАЛЫ И МЕТОДЫ

Биологические образцы. Материалом для исследования служили образцы ДНК индивидов из Республики Беларусь. Забор венозной крови в пробирки BD Vacutainer K2E осуществлен у 275 индивидов (мужчины – 119, женщины – 156). Средний возраст добровольцев составил 57.3 ± ± 16.9 лет (для мужчин – 58.4 ± 16.8 лет, для женщин – 56.4 ± 17.0 лет), диапазон – от 18 до 93 полных лет. Забор буккального эпителия на ватный тампон-зонд осуществлен у 466 индивидов (мужчины – 262, женщины – 204). Средний возраст добровольцев составил 45.9 ± 18.1 лет (для мужчин – 46.7 ± 18.1 лет, для женщин – 44.8 ± 18.2 лет), диапазон – от 9 до 93 полных лет. Образцы спермы от 382 мужчин были получены на основании запроса в Республиканский банк ДНК (Институт генетики и цитологии НАН Беларуси). Дополнительно 85 образцов спермы получены от добровольцев, средний возраст которых составил 34.2 ± ± 5.2 лет, диапазон – от 22 до 53 полных лет.

Забор биологического материала был осуществлен с использованием процедур, одобренных Комитетом по биоэтике при Институте генетики и цитологии НАН Беларуси. Биологический материал хранили при температуре –20°C в течение 1–2 мес., либо при –80°C при долговременном хранении. ДНК экстрагировали с использованием набора MagMAX™ DNA Multi-Sample Kit (ThermoFisher, США) в соответствии с инструкциями производителя. Выделенную ДНК количественно оценивали с использованием NanoPhotometer N50 (IMPLEN, США).

Для того чтобы целенаправленно подобрать CpG-динуклеотиды для последующего молекулярно-генетического исследования образцов ДНК индивидов из Республики Беларусь, проведены биоинформатический и статистический анализы данных 18 GEO-проектов. Информация об уровне метилирования ДНК с использованием IHM 450K BeadChip (Illumina) представлена в открытом доступе на платформе Gene Expression Omnibus NCBI: для образцов крови – GSE40279, GSE42861, GSE51032, GSE50660, GSE55763, GSE77696, GSE106648, GSE125105 (5935 человек, диапазон возраста – от 17 до 93 лет); для образцов буккального эпителия – GSE39560, GSE59505, GSE72120, GSE72556, GSE78874, GSE92767, GSE99029, GSE110128, GSE111223 (690 человек, диапазон возраста – от 1 мес. до 91 года); для образцов спермы – GSE114753 (141 человек, диапазон возраста – от 19 до 57 лет).

На основании проведенного анализа отобран 41 CpG-динуклеотид, уровень метилирования которого коррелировал с хронологическим возрастом человека: для образцов крови 27 CpG-динуклеотидов – cg00481951 (*SST*), cg02872426 (*DDO*), cg05213896 (*IL4I1*), cg06784991 (*ZYG11A*), cg06874016 (*NKIRAS2*), cg07547549 (*SLC12A5*), cg07553761 (*TRIM59*), cg08128734 (*RASSF5*), cg08468401, cg08928145 (*TSSK6*), cg10501210 (*MIR29B2CHG*), cg11807280 (*MEIS1*), cg12757011 (*SLC4A10*), cg14361627 (*LOC105375508*), cg16054275 (*F5*), cg16867657 (*ELOVL2*), cg18384097 (*PTPN7*), cg18473521 (*HOXC4*), cg19283806 (*CCDC102B*), cg19671120 (*CNGA3*), cg22454769 (*FHL2*), cg24079702 (*FHL2*), cg25410668 (*RPA2*), Chr.3:g.51707136 (*GRM2*), Chr.4:g.16573797 (*LDB2*), Chr.7:g.130734357 (*KLF14*) и Chr.20:g.63980491 (*SAMD10*) [6, 22]; для образцов буккального эпителия 7 CpG-динуклеотидов – cg00481951 (*SST*), cg07547549 (*SLC12A5*), cg08928145 (*TSSK6*), cg12757011 (*SLC4A10*), cg14361627 (*LOC105375508*), cg18384097 (*PTPN7*), cg19671120 (*CNGA3*) [23, 24]; для образцов спермы 14 CpG-динуклеотидов – cg00881487 (*MXRA5*), cg03762081, cg04123357 (*COL18A1*), cg06304190 (*TTC7B*), cg06979108 (*NOX4*), cg10836509, cg12837463, cg13014709 (*CDK4*), cg13030797, cg13837679 (*ALX4*), cg14750551 (*PARP14*), cg21843517 (*SNRNP35*), cg23488376 (*LINC01595*), cg24958325 (*ANKMY1*) [15].

Бисульфитное секвенирование с использованием SNaPshot. Праймеры для анализа CpG-динуклеотидов cg02872426, cg06784991, cg06874016, cg07553761, cg11807280, cg16054275, cg16867657, cg18473521, cg25410668, Chr.3:g.51707136, Chr.4:g.16573797, Chr.7:g.130734357, Chr.20:g.63980491 приведены в работе J. Naue с соавт. [6]. Праймеры для анализа CpG-динуклеотидов cg00481951, cg07547549, cg08928145, cg12757011, cg14361627, cg18384097, cg19671120 представлены в работе

S.R. Hong et al. [24]. Праймеры для анализа CpG-динуклеотидов cg06304190, cg06979108, cg12837463, cg00881487, cg13030797, cg14750551, cg23488376 представлены в работе H.Y. Lee et al. [15]. Праймеры и SBE-олигонуклеотиды для анализа CpG-динуклеотидов cg03762081, cg04123357, cg05213896, cg08128734, cg08468401, cg10501210, cg10836509, cg13014709, cg13837679, cg19283806, cg21843517, cg22454769, cg24079702 и cg24958325 смоделированы с использованием программы BiSearch (http://bisearch.enzim.hu/) – табл. 1.

Бисульфитная конверсия ДНК проводилась для 200–500 нг с использованием MethylEdge® Bisulfite Conversion System Kit (Promega, США). ПЦР проводили в объеме 20 мкл, содержащем 10–15 нг конвертированной ДНК, 1 ед. ArtStart ДНК-полимеразы (ArtBioTech, Беларусь), 2 мкл 10× буфера для ПЦР (в состав которого входил хлорид магния в концентрации 3.0 мМ), 200 нМ каждого дНТФ и 0.4–1.0 мкМ каждого праймера. Протокол амплификации на термоциклере ProFlex PCR System (ThermoFisher, США) включал следующие этапы: 95°C – 4 мин, далее 34 цикла (94°C – 20 с, 56°C – 30 с, 72°C – 45 с), финальная элонгация при 72°C в течение 7 мин. В дальнейшем 5 мкл амплификата очищали с использованием Exo-CIP™ Rapid PCR Cleanup Kit (NEB, США). Реакцию микросеквенирования (SBE, Single-base extension) проводили с использованием SNaPshot Kit (Applied Biosystems, США) согласно протоколу производителя, каждая реакция включала также 3 мкл очищенного на предыдущей стадии амплификата и 0.2–0.4 мкМ соответствующего SBE-олигонуклеотида. В дальнейшем 10 мкл амплификата очищали с использованием 1 ед. FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher).

Электрофоретическая детекция продуктов реакции микросеквенирования проводилась с использованием генетического анализатора ABI PRISM 3500 и программного обеспечения GeneMapper® 5.0 (Applied Biosystems).

Процент метилирования (0–100%) для каждого CpG-динуклеотида рассчитывали согласно формуле: интенсивность флуоресцентного сигнала для G (метилированные, т.е. неконвертированные цитозины; согласно правилу комплементарности C-G) разделить на сумму флуоресцентных сигналов G и A (неметилированные, т.е. конвертированные в T-цитозины; согласно правилу комплементарности T-A).

Статистический анализ данных. В ходе биоинформатического и статистического анализов были рассчитаны: коэффициенты корреляции R по Спирману; коэффициенты детерминации (R^2), а также скорректированные R^2, равные $1 - (1 - R^2)(n - 1)/(n - k)$, где k – число независимых переменных и n – объем выборки; значения среднего абсолютного отклонения (Mean Absolute Deviation, MAD) от истинного возраста и среднеквадратическая ошибка модели (Root Mean Square Error, RMSE). Данные уровня метилирования CpG-динуклеотидов были нормализованы с использованием алгоритма Batch normalization ($n = 30$). Статистический анализ проводили с использованием программного обеспечения SPSS v20.0.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Коэффициенты корреляции R

Коэффициенты R для CpG-динуклеотидов, рассчитанные по результатам молекулярно-генетического анализа для образцов крови и буккального эпителия с указанием половой принадлежности, представлены в табл. 2, для спермы – в табл. 3. Наибольшие расчетные значения R показаны для следующих CpG-динуклеотидов (в порядке уменьшения абсолютного значения): для образцов крови – cg19283806, cg14361627, cg08128734, cg16867657, cg22454769, cg07553761, cg08468401 и cg11807280; для образцов буккального эпителия – cg07547549, cg14361627, cg00481951, cg12757011 и cg08928145; для образцов спермы – cg13014709, cg12837463, cg06304190, cg00881487 и cg23488376. Из-за слабой или статистически незначимой ассоциации с хронологическим возрастом из дальнейшего моделирования исключены CpG-динуклеотиды cg08928145, cg18384097 (табл. 2) и cg13837679 (табл. 3).

Регрессионная модель предсказания возраста

В модели предсказания возраста по биологическим образцам отобраны CpG-динуклеотиды с использованием процедуры Stepwise, реализованной в SPSS, оказывающие достоверное и значительное влияние на изменение процента объясненной дисперсии R^2: 12 CpG-динуклеотидов для крови – cg19283806, cg14361627, cg22454769, cg11807280, cg08468401, cg05213896, cg16867657, cg07547549, cg25410668, cg08128734, cg10501210 и Chr.4:g.16573797; 6 CpG-динуклеотидов для буккального эпителия – cg00481951, cg07547549, cg08928145, cg12757011, cg14361627 и cg19671120; 5 CpG-динуклеотидов для спермы – cg13014709, cg06304190, cg04123357, cg12837463, cg06979108.

При использовании множественной линейной регрессии в качестве модели предсказания возраста по образцам *крови* в общей группе индивидов (без разделения по полу) значение MAD составило 3.4 года, RSME – 2.4 года. При разделении по полу для мужчин MAD – 3.3 года, RSME – 2.3 года; для женщин MAD – 3.5 года, RSME – 2.4 года. Скорректированный процент объяснен-

Таблица 1. Праймеры и SBE-олигонуклеотиды, смоделированные с использованием программы BiSearch

№	CpG-динуклеотид	Ген	Хромосомная позиция*	Нуклеотидная последовательность
1	cg03762081	–	Chr.19:g.51020309	ACTCTGTGGC[CG]TGAGTGAGCT
	F-праймер (5' > 3'): GAGGTTAAGAGAAATAGAAATAAG R-праймер (5' > 3'): AATATAAAACAAAACCACAACCC SBE-олигонуклеотид (5' > 3'): $(T)_n$-AACACACACAAAACTCACTCAC			
2	cg04123357	*COL18A1*	Chr.21:g.45509586	ACAGCCACCG[CG]ACTTCCAGCC
	F-праймер (5' > 3'): AGATGATATTTTGGTTAGTT R-праймер (5' > 3'): AAACTAAACCCTCTTCTACTA SBE-олигонуклеотид (5' > 3'): $(T)_n$-CACTCACCACCRACTAAAAATC			
3	cg05213896	*IL4I1*	Chr.19:g.49890396	TGCGCTTCAC[CG]CCGGTCCGCT
	F-праймер (5' > 3'): AACTTAAAAACACCTTAATA R-праймер (5' > 3'): AATTTGAAGGTGTTGAAGGT SBE-олигонуклеотид (5' > 3'): $(T)_n$-ACAACAACAACAAAAAAATAATAC			
4	cg08128734	*RASSF5*	Chr.1:g.206512090	GATTTAGCCC[CG]TGTTTGGCTT
	F-праймер (5' > 3'): TTTATTTTGGGTATTTGGAAGGT R-праймер (5' > 3'): CCATTTCTATACTTCATTTTCT SBE-олигонуклеотид (5' > 3'): $(T)_n$-CTTAAACTCTAAAACCAAACAC			
5	cg08468401	–	Chr.3:g.14261631	TGAGATGGTT[CG]AGGTGCCAGG
	F-праймер (5' > 3'): TTTTAGTTGTTTGTGGTGTT R-праймер (5' > 3'): TTTTCTTTAAACTTAATCTCATTC SBE-олигонуклеотид (5' > 3'): $(T)_n$-ATAACATAACTCCTAACACCTC			
6	cg10501210	*MIR29B2CHG*	Chr.1:g.207823675	CACCCATCGC[CG]CCTGCGCAGA
	F-праймер (5' > 3'): GAAAAGAAATTAAAGTGGGAA R-праймер (5' > 3'): ACCAAATTCTAAAACATTCRAAAC SBE-олигонуклеотид (5' > 3'): $(T)_n$-CCAAAATTTAAATCTACRCAAAC			
7	cg10836509	–	Chr.1:g.226111902	CTCCCATGGC[CG]CAGGCAGCAG
	F-праймер (5' > 3'): GTTAGGTTTGGTGATGTAGT R-праймер (5' > 3'): CAAAAACTTAAACCTCCAAA SBE-олигонуклеотид (5' > 3'): $(T)_n$-AAACRCRTCTCCTACTACCTAC			
8	cg13014709	*CDK4*	Chr.12:g.57754606	TCAGACAATC[CG]ATTCCATTCC
	F-праймер (5' > 3'): TTATATATTATAATTATAGG R-праймер (5' > 3'): ACTCAAACAATATACACTTTAC SBE-олигонуклеотид (5' > 3'): $(T)_n$-CAAATTTCAAAATTTCC			
9	cg13837679	*ALX4*	Chr.11:g.44308365	AGGCCTCCGC[CG]TGTCCCAGCC
	F-праймер (5' > 3'): GTTTGTGAAAGAAGGAATTT R-праймер (5' > 3'): AATCCAAACTAATAAAAAAACAC SBE-олигонуклеотид (5' > 3'): $(T)_n$-ACCTAAAAACAAAAACTAAAACAC			
10	cg19283806	*CCDC102B*	Chr.18:g.68722183	TGTTTGCACTCATCC[CG]CATCCTTTTTTGGCT
	F-праймер (5' > 3'): GGGGAATGTTTGTATTTATTT R-праймер (5' > 3'): TAAAAAAATTACCCTTCTCT SBE-олигонуклеотид (5' > 3'): $(T)_n$-ACAAAAATAACAACCAAAAAAAAATAC			

Таблица 1. Окончание

№	CpG-динуклеотид	Ген	Хромосомная позиция*	Нуклеотидная последовательность
11	cg21843517	*SNRNP35*	Chr.12:g.123465627	CACACGACCG[CG]CGGTCTGGAG
	F-праймер (5' > 3'): AAGGAGTATGATTTATTTAAAGT R-праймер (5' > 3'): AAAAAAATCTCCTATAACACC SBE-олигонуклеотид (5' > 3'): $(T)_n$-CAACATTACCCTCCAAACRAC			
12	cg22454769	*FHL2*	Chr.2:g.105399310	TCCGGCGTGGGCTCT[CG]GGCGCGAGTTTCGG
	F-праймер (5' > 3'): TGTTAGGGTTTTTTTTTTYGTGT R-праймер (5' > 3'): TAAAACTTCTCCAATCTCCC SBE-олигонуклеотид (5' > 3'): $(T)_n$-CAAACCTCRTCCRAAACTC			
13	cg24079702	*FHL2*	Chr.2:g.105399314	CGTGGGCTCTCGGG[CG]CGAGTTTCGGACG
	F-праймер (5' > 3'): TGTTAGGGTTTTTTTTTTYGTGT R-праймер (5' > 3'): TAAAACTTCTCCAATCTCCC SBE-олигонуклеотид (5' > 3'): $(T)_n$-CAAACCTCRTCCRAAACTC			
14	cg24958325	*ANKMY1*	Chr.2:g.240558804	TGCCTCCACA[CG]CCAGTGCCTG
	F-праймер (5' > 3'): TAAGGTTGGGTTAGGATTTT R-праймер (5' > 3'): CAATTCTATAAATCATAATCAAC SBE-олигонуклеотид (5' > 3'): $(T)_n$-ATAAACTCAACAAACAAACACTAAC			

* GRCh38.p13 (GCF_000001405.39).

ной дисперсии *R*^2 составил не менее 91.8% (рис. 1,*а*). Возраст с точностью до 3 лет корректно предсказан для 42.9% индивидов, с точностью до 5 лет – для 72.0%, с точностью до 7 лет – для 86.9%, для 13.1% ошибка предсказания возраста составила более 7 лет.

Значение MAD (без разделения по полу) для модели предсказания возраста по образцам *буккального эпителия* составило 4.6 года, RSME – 3.4 года. При разделении по полу для мужчин MAD – 4.5 года, RSME – 3.4 года; для женщин MAD – 4.7 года, RSME – 3.5 года. Скорректированный процент объясненной дисперсии *R*^2 составил не менее 83.6% (рис. 1,*б*). Возраст с точностью до трех лет корректно предсказан для 33.9% индивидов, с точностью до пяти лет – для 54.1%, с точностью до семи лет – для 70.5%, для 29.5% ошибка предсказания возраста составила более семи лет.

Значение MAD для модели предсказания возраста по образцам *спермы* составило 3.0 года, RSME – 2.3 года. Скорректированный процент объясненной дисперсии *R*^2 составил не менее 46.6% (рис. 1,*в*). Возраст с точностью до трех лет корректно предсказан для 58.7% индивидов, с точностью до пяти лет – для 80.5%, для 19.5% индивидов ошибка предсказания возраста составила более пяти лет.

Таким образом, наилучшая точность предсказания возраста индивида (без указания половой принадлежности) характерна для образцов крови – при анализе 12 CpG-динуклеотидов ошибка составила не более 3.4 года (для мужчин не более 3.3 года). Полученные нами результаты по всем трем типам биологических образцов сопоставимы с аналогичными исследованиями [5–19]. Для отобранных нами в процессе статистического анализа CpG-динуклеотидов показаны высокие значения корреляции с хронологическим возрастом.

Вероятность отнесения образца к возрастной группе

Для криминалистов в судебной практике при выяснении предположительного возраста неизвестного индивида может быть важна информация не только о конкретном возрасте, но и об отнесении данного субъекта к определенной возрастной группе, например: “моложе 40 лет” или “старше 40 лет” и т.п. Точность отнесения неизвестного индивида к конкретной группе по результатам анализа метилирования CpG-динуклеотидов будет выше, чем при ответе на вопрос о приближенном к истинному значению хронологическом возрасте, так как число групп в анализе с использованием, например, логистической регрессии будет равняться двум.

Таблица 2. Коэффициенты корреляции (*R*) с возрастом для CpG-динуклеотидов (кровь и буккальный эпителий)

Кровь

CpG-динуклеотид	мужчины ($n = 119$)				женщины ($n = 156$)			
	НГ 95%	*R*	ВГ 95%	*p*-уровень	НГ 95%	*R*	ВГ 95%	*p*-уровень
cg00481951	–0.010	0.182	0.344	4.80E-02	0.013	0.161	0.316	4.51E-02
cg02872426	–0.287	–0.423	–0.558	1.42E-06	–0.318	–0.485	–0.596	1.12E-06
cg05213896	0.355	0.515	0.636	2.12E-09	0.452	0.572	0.673	6.29E-15
cg06784991	0.157	0.334	0.495	2.03E-04	0.312	0.454	0.583	2.60E-09
cg06874016	–0.551	–0.383	–0.192	1.74E-05	–0.510	–0.375	–0.223	1.40E-06
cg07547549	0.267	0.441	0.579	5.33E-07	0.384	0.508	0.632	1.35E-11
cg07553761	0.488	0.629	0.732	1.91E-14	0.571	0.675	0.764	4.28E-22
cg08128734	–0.775	–0.682	–0.555	1.27E-17	–0.799	–0.730	–0.642	2.90E-27
cg08468401	–0.663	–0.530	–0.349	5.51E-10	–0.795	–0.713	–0.610	1.48E-25
cg08928145*	–0.111	0.078	0.251	0.398	–0.127	0.046	0.216	0.568
cg10501210	–0.568	–0.430	–0.269	1.09E-06	–0.608	–0.496	–0.366	4.64E-11
cg11807280	–0.715	–0.605	–0.458	3.15E-13	–0.718	–0.619	–0.501	7.51E-18
cg12757011	0.149	0.336	0.486	1.88E-04	0.203	0.363	0.506	3.25E-06
cg14361627	0.679	0.776	0.845	3.55E-25	0.665	0.751	0.813	1.56E-29
cg16054275	–0.574	–0.417	–0.245	2.37E-06	–0.496	–0.370	–0.225	2.05E-06
cg16867657	0.456	0.610	0.724	1.72E-13	0.628	0.716	0.791	7.70E-26
cg18384097	–0.486	–0.308	–0.114	6.47E-04	–0.426	–0.300	–0.143	1.42E-04
cg18473521	0.344	0.503	0.626	5.62E-09	0.300	0.450	0.574	3.88E-09
cg19283806	–0.858	–0.789	–0.676	1.65E-26	–0.871	–0.820	–0.751	3.23E-39
cg19671120	0.275	0.431	0.576	1.01E-06	0.076	0.243	0.388	2.21E-03
cg22454769	0.410	0.571	0.690	1.19E-11	0.653	0.734	0.794	1.29E-27
cg24079702	–0.087	0.106	0.299	0.252	0.078	0.238	0.386	2.73E-03
cg25410668	0.314	0.484	0.620	2.40E-08	0.579	0.686	0.767	4.88E-23
Chr.3:g.51707136	0.031	0.234	0.422	1.04E-02	0.337	0.465	0.590	9.58E-10
Chr.4:g.16573797	–0.627	–0.508	–0.354	3.55E-09	–0.600	–0.482	–0.334	1.92E-10
Chr.7:g.130734357	0.300	0.470	0.608	6.81E-08	0.322	0.466	0.584	8.98E-10
Chr.20:g.63980491	–0.599	–0.466	–0.308	9.48E-08	–0.703	–0.608	–0.493	3.99E-17

Буккальный эпителий

CpG-динуклеотид	мужчины ($n = 262$)				женщины ($n = 204$)			
cg00481951	0.654	0.722	0.779	1.89E-43	0.638	0.717	0.778	1.58E-33
cg07547549	0.714	0.770	0.819	1.23E-52	0.710	0.780	0.833	5.43E-43
cg08928145	0.467	0.555	0.634	1.56E-22	0.579	0.671	0.743	5.04E-28
cg12757011	0.551	0.633	0.702	8.56E-31	0.519	0.613	0.687	1.89E-22
cg14361627	0.694	0.753	0.796	3.68E-49	0.700	0.770	0.826	2.30E-41
cg18384097*	–0.230	–0.112	–0.001	0.071	–0.310	–0.184	–0.058	0.009
cg19671120	0.139	0.254	0.359	3.13E-05	0.080	0.210	0.328	0.003

Примечание. НГ 95% – нижняя граница 95%-ного доверительного интервала; ВГ 95% – верхняя граница 95%-ного доверительного интервала; * исключенные из моделирования CpG-динуклеотиды; *p*-уровень – уровень статистической значимости.

Например, при разбиении всего массива образцов на две категории (без разделения по полу) и использовании логистической регрессии на основании молекулярно-генетического анализа 12 CpG-динуклеотидов в образцах крови с вероятностью 92.0% имеется возможность сделать заключение, что возраст неизвестного индивида составляет меньше 40 лет, с вероятностью 98.2% – больше 40 лет, с вероятностью 89.7% – меньше 50 лет, с вероятностью 99.0% – больше 50 лет; с вероятностью 88.2% – меньше 60 лет, с вероятностью 89.9% – больше 60 лет. Также при анализе уровня метилирования 6 CpG-динуклеотидов в образцах буккального эпителия с вероятностью 93.8% имеется возможность сделать заключение, что возраст неизвестного индивида составляет меньше 40 лет, с вероятностью 89.8% – больше 40 лет. На основании анализа 5 CpG-динуклеотидов в образцах спермы с вероятностью 84.4% можно сделать заключение, что возраст неизвестного индивида составляет меньше 35 лет.

Таким образом, для полученных нами результатов характерен баланс точности предсказания возраста с использованием небольшого количества CpG-динуклеотидов – от 5 для спермы до 12 для крови. Мультиплексирование при молекулярно-генетическом тестировании позволяет проводить анализ в генетическом анализаторе за 1–2 реакции. Для технологии микросеквенирования SNaPshot характерны высокие значения точности и воспроизводимости. Тем не менее экспертам-криминалистам при использовании подобного подхода для предсказания возраста по биологическим следам необходимо будет учитывать ряд факторов, которые позволят получить достоверный результат: получать на этапе конверсии ДНК достаточное количество материала для ПЦР, так как повторная конверсия может оказать существенное влияние на воспроизводимость; проводить исследование метилирования CpG-динуклеотидов не менее чем в двух проворностях, чтобы усреднить результат. Выполнение данных рекомендаций позволит максимально нивелировать методические ошибки.

Полученные нами данные будут положены в основу при разработке калькуляторов предсказания возраста по биологическим следам для экспертов-криминалистов.

На основании данных, представленных в открытом доступе на платформе GEO NCBI для проектов по определению полногеномного профиля метилирования ДНК, сформирован перечень CpG-динуклеотидов с высокими значениями линейной зависимости между уровнем метилирования и хронологическим возрастом. Для молекулярно-генетического анализа отобран 41 CpG-динуклеотид с высоким прогностическим потенциалом для создания моделей предсказания возраста человека по биологическим образцам. Для данных CpG-динуклеотидов определен уровень метилирования для 1208 биологических образцов индивидов из Республики Беларусь, рассчитаны коэффициенты корреляции R с возрастом. С использованием множественной линейной регрессии созданы модели для предсказания возраста по образцам крови, буккального эпителия и спермы.

Таблица 3. Коэффициенты корреляции (R) с возрастом для CpG-динуклеотидов (сперма)

CpG-динуклеотиды	НГ 95%	R	ВГ 95%	p-уровень
cg00881487	–0.301	–0.389	–0.222	2.93E-11
cg03762081	–0.261	–0.348	–0.181	9.84E-09
cg04123357	0.361	0.276	0.439	7.88E-16
cg06304190	–0.369	–0.446	–0.282	1.50E-16
cg06979108	0.326	0.242	0.402	5.07E-13
cg10836509	–0.266	–0.351	–0.175	5.16E-09
cg12837463	–0.373	–0.452	–0.289	6.81E-17
cg13014709	–0.422	–0.497	–0.331	1.28E-21
cg13030797	–0.199	–0.287	–0.113	1.48E-05
cg13837679*	–0.071	–0.156	0.021	0.124
cg14750551	–0.256	–0.334	–0.176	2.02E-08
cg21843517	0.244	0.15	0.331	9.20E-08
cg23488376	–0.294	–0.368	–0.216	9.71E-11
cg24958325	–0.181	–0.269	–0.091	8.17E-05

Примечание. НГ 95% – нижняя граница 95%-ного доверительного интервала, ВГ 95% – верхняя граница 95%-ного доверительного интервала, * – исключенные из моделирования CpG-динуклеотиды, p-уровень – уровень статистической значимости.

Предложенные нами модели предсказания возраста на основании профиля метилирования CpG-динуклеотидов крови, буккального эпителия и спермы отличаются относительной простотой, так как в анализе используется небольшое количество маркеров и методика, разработанная с их использованием, может применяться в криминалистических лабораториях молекулярно-генетической направленности. Ошибка предсказания возраста для рассчитанных нами моделей соответствует аналогичным исследованиям [6, 15, 24].

Разрабатываемая нами методика определения возраста по биологическим образцам на основе анализа профиля метилирования ДНК основывается на результатах отечественных научно-технических исследований. В результате данный под-

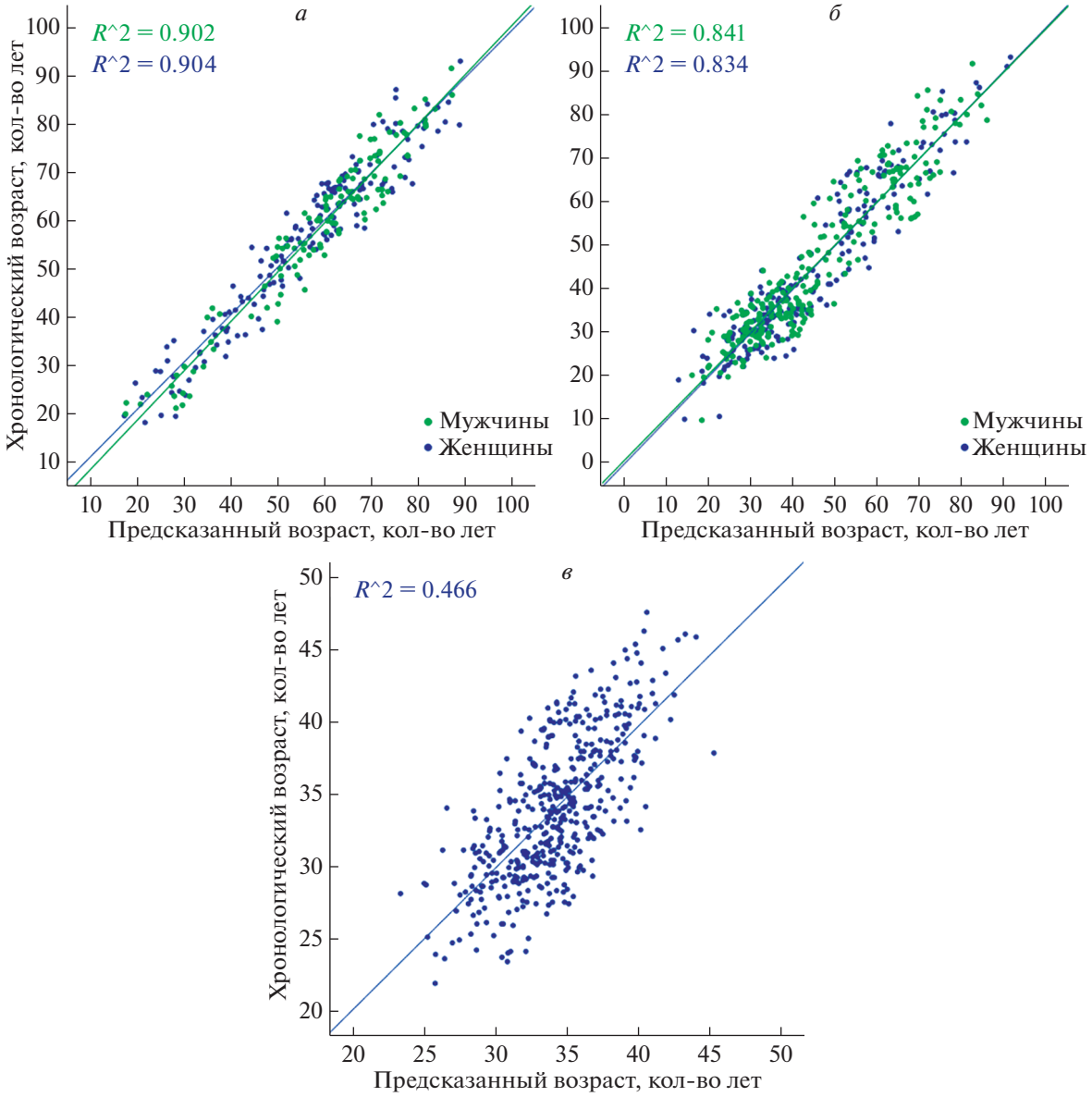

Рис. 1. Точность предсказания возраста индивида по образцам крови (*а*), буккального эпителия (*б*) и спермы (*в*).

ход обеспечит использование актуальных знаний в области прикладной генетики при решении криминалистических задач на территории Республики Беларусь и Российской Федерации.

Работа выполнена в рамках мероприятия 2 “Разработка методики определения вероятного возраста индивида по характеристике его ДНК” научно-технической программы Союзного государства “Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства” (ДНК-идентификация).

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 об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Золотаренко А.Д., Чекалин Е.В., Брускин С.А.* Современные молекулярно-генетические методы определения возраста в криминалистике // Генетика. 2019. Т. 55. № 12. С. 1397–1409. https://doi.org/10.1134/S0016675819120154
2. *Кильчевский А., Моссэ И., Шаптуренко М., Буракова А.* Генетика – судебной экспертизе Беларуси // Наука и инновации. 2020. № 10(212). С. 22–28.
3. *Hannum G., Guinney J., Zhao L. et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates // Mol. Cell. 2013. V. 24. P. 359–367. https://doi.org/10.1016/j.molcel.2012.10.016
4. *Kurdyukov S., Bullock M.* DNA methylation analysis: Choosing the right method // Biology (Basel). 2016. V. 5. № 1. P. e1–e21. https://doi.org/10.3390/biology5010003
5. *Bocklandt S., Lin W., Sehl M.E. et al.* Epigenetic predictor of age // PLoS One. 2011. V. 6. P. 1–6. https://doi.org/10.1371/journal.pone.0014821
6. *Naue J., Hoefsloot H.C.J., Mook O.R.F. et al.* Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression // Forensic Sci. Int. Genet. 2017. V. 31. P. 19–28. https://doi.org/10.1016/j.fsigen.2017.07.015
7. *Park J.L., Kim J.H., Seo E. et al.* Identification and evaluation of age-correlated DNA methylation markers for forensic use // Forensic Sci. Int. Genet. 2016. V. 23. P. 64–70. https://doi.org/10.1016/j.fsigen.2016.03.005
8. *Vidaki A., Ballard D., Aliferi A. et al.* Syndercombe Court D. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing // Forensic Sci. Int. Genet. 2017. V. 28. P. 225–236. https://doi.org/10.1016/j.fsigen.2017.02.009
9. *Zbiec-Piekarska R., Spolnicka M., Kupiec T. et al.* Development of a forensically useful age prediction method based on DNA methylation analysis // Forensic Sci. Int. Genet. 2015. V. 17. P. 173–179. https://doi.org/10.1016/j.fsigen.2015.05.001
10. *Eipel M., Mayer F., Arent T. et al.* Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures // Aging (Albany NY). 2016. V. 8. № 5. P. 1034–1048. https://doi.org/10.18632/aging.100972
11. *McEwen L.M., O'Donnell K.J., McGill M.G. et al.* The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells // PNAS. 2020. V. 117. № 38. P. 23329–23335. https://doi.org/10.1073/pnas.1820843116
12. *Koop B.E., Mayer F., Gündüz T. et al.* Postmortem age estimation via DNA methylation analysis in buccal swabs from corpses in different stages of decomposition-a "proof of principle" study // Int. J. Legal Med. 2021. V. 135. № 1. P. 167–173. https://doi.org/10.1007/s00414-020-02360-7
13. *Dongen J., Ehli E.A., Jansen R. et al.* Genome-wide analysis of DNA methylation in buccal cells: a study of monozygotic twins and mQTLs // Epigenetics Chromatin. 2018. V. 11. № 1. P. e1–e14. https://doi.org/10.1186/s13072-018-0225-x
14. *Wozniak A., Heidegger A., Piniewska-Rog D. et al.* Development of the VISAGE enhanced tool and statistical models for epigenetic age estimation in blood, buccal cells and bones // Aging (Albany NY). 2021. V. 13. № 5. P. 6459–6484. https://doi.org/10.18632/aging.202783
15. *Lee H.Y., Jung S.E., Oh Y.N. et al.* Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study // Forensic Sci. Int. Genet. 2015. V. 19. P. 28–34. https://doi.org/10.1016/j.fsigen.2015.05.014
16. *Alsaleh H., McCallum N.A., Halligan D.L., Haddrill P.R.* A multi-tissue age prediction model based on DNA methylation analysis // Forensic Sci. Int. Genet. Suppl. Ser. 2017. V. 6. P. 62–64. https://doi.org/10.1016/j.fsigss.2017.09.056
17. *Li L., Song F., Huang Y., Zhu H., Hou Y.* Age-associated DNA methylation determination of semen by pyrosequencing in Chinese Han population // Forensic Sci. Int. Genet. 2017. V. 6. P. e99–e100. https://doi.org/10.1016/j.fsigss.2017.09.042
18. *Jenkins T.G., James E.R., Alonso D.F. et al.* Cigarette smoking significantly alters sperm DNA methylation patterns // Andrology. 2017. V. 5. № 6. P. 1089–1099. https://doi.org/10.1111/andr.12416
19. *Lee J.W., Choung C.M., Jung J.Y. et al.* A validation study of DNA methylation-based age prediction using semen in forensic casework samples // Leg. Med. (Tokyo). 2018. V. 31. P. 74–77. https://doi.org/10.1016/j.legalmed.2018.01.005
20. *Fleckhaus J., Freire-Aradas A., Rothschild M.A., Schneider P.M.* Impact of genetic ancestry on chronological age prediction using DNA methylation analysis // Forensic Sci. Int. Genet. Suppl. Ser. 2017. V. 6. P. e399–e400. https://doi.org/10.1016/j.fsigss.2017.09.162
21. *Donkin I., Barres R.* Sperm epigenetics and influence of environmental factors // Mol. Metab. 2018. V. 14. P. 1–11. https://doi.org/10.1016/j.molmet.2018.02.00
22. *Кипень В.Н., Богданова М.В., Буракова А.А. и др.* Обоснование минимального объема выборки для предсказания хронологического возраста человека // Мол. и прикл. генетика. 2021. Т. 30. С. 39–48. https://doi.org/10.47612/1999-9127-2021-30-39-48
23. *Кипень В.Н., Богданова М.В., Буракова А.А. и др.* Предсказательный потенциал CpG-маркеров для определения хронологического возраста человека // Мол. и прикл. генетика. 2020. Т. 28. С. 70–80.
24. *Hong S.R., Jung S., Lee E.H. et al.* DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers // Forensic Sci. Int. Genet. 2017. V. 29. P. 118–125. https://doi.org/10.1016/j.fsigen.2017.04.006

Determination of Human Chronological Age from Biological Samples Based on the Analysis of CpG-Dinucleotides Methylation

V. A. Lemesh[a], *, V. N. Kipen[a], **, M. V. Bahdanava[a], A. A. Burakova[a], A. G. Bulgak[b], A. V. Bayda[c], S. A. Bruskin[d], O. V. Zotova[b], and O. I. Dobysh[a]

[a] *Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, 220072 Belarus*
[b] *State Institution "Republican Scientific and Practical Centre "Cardiology", Minsk, 220089 Belarus*
[c] *Belarusian Medical Academy of Postgraduate Education, Minsk, 220013 Belarus*
[d] *Vavilov Institute of General Genetics RAS, Moscow, 119991 Russia*
**e-mail: v.lemesh@igc.by*
***e-mail: v.kipen@igc.by*

Based on bioinformatics and statistical analysis of GEO-projects to determine the genome-wide profile of human DNA methylation, a list of 41 CpG dinucleotides with high predictive potential was formed to create models for predicting human age from biological samples. The methylation level was determined for 1208 samples of individuals from the Republic of Belarus (275 – blood, 466 – buccal epithelium, 467 – sperm). The correlation coefficients *R* were calculated, and mathematical models for determining the age of individual were constructed. The average value of the accuracy of age prediction from blood samples using 12 CpG-dinucleotides was 3.4 years (for men – 3.3, for women – 3.5); for buccal epithelium samples using 6 CpG-dinucleotides – 4.6 years (for men – 4.5, for women – 4.7); for sperm samples using 5 CpG-dinucleotides – 3.0 years. The results obtained will be used as a basis for the development of calculators for predicting the age of an individual based on traces of biological character for forensic experts.

Keywords: forensic, age determination, blood, buccal epithelium, semen, DNA methylation, human methylation 450K BeadChip.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1386–1395

УДК 575.174

ФИЛОГЕНЕТИЧЕСКИЙ АНАЛИЗ ЮЖНОСИБИРСКОЙ ГАПЛОГРУППЫ Q-YP1102 ПО ДАННЫМ О Y-SNP- И Y-STR-МАРКЕРАХ У ТУВИНЦЕВ И ОКРУЖАЮЩИХ ПОПУЛЯЦИЙ

© 2021 г. А. Т. Агджоян[1, 2, *], Л. Д. Дамба[4], В. М. Гурьянов[1], В. В. Запорожченко[1, 2], О. П. Балановский[1, 2, 3]

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

[2]Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, 115522 Россия

[3]АНО "Биобанк Северной Евразии", Москва, 115201 Россия

[4]Научно-исследовательский институт медико-социальных проблем и управления Республики Тыва, Кызыл, 667003 Россия

**e-mail: aagdzhoyan@gmail.com*

Поступила в редакцию 31.05.2021 г.
После доработки 18.06.2021 г.
Принята к публикации 26.07.2021 г.

Гаплогруппа Q Y-хромосомы составляет основу генофонда американских индейцев, а в Евразии наибольших частот достигает в популяциях Центральной и Южной Сибири. Проведенное генотипирование обширной панели SNP-маркеров, выделяющих около 60 ветвей в составе гаплогруппы Q, позволило провести углубленный филогеографический анализ в популяциях Южной Сибири и смежных регионов. Изучение шести популяций тувинцев и тоджинцев выявило три ветви гаплогруппы Q, восходящие к одному корню – YP1102. Наиболее частая из них линия YP1691 достигает максимума в генофонде северо-восточных тувинцев-тоджинцев (51%), понижаясь в частоте с востока на запад до 4%. Подобному тренду следует распространение и базальной линии YP1102*. Третья линия BZ99* редка у тувинцев (2%), но заметна в генофонде хакасов (9%). Филогенетический анализ STR-гаплотипов показал распространение одних и тех же кластеров не только у разных групп тувинцев, но и среди окружающих популяций. Структура основного кластера свидетельствует об увеличении численности носителей линии YP1691 в период позднего железного века. Карта частоты линии YP1691 отражает последствия активных миграционных процессов в течение нашей эры на территории Южной и Западной Сибири. Среди древних популяций Сибири вариант YP1102* найден у пяти индивидов: четырех представителей бронзового века Прибайкалья и одного представителя культуры скифо-сибирского типа на границе Монголии и Тувы. Данные древней ДНК и более высокое генетическое разнообразие в населении Южной Сибири позволяют предположить происхождение гаплогруппы Q-YP1102 в данном регионе и ее дальнейшее проникновение на запад и север Сибири, в том числе, возможно, с предками популяций енисейской лингвистической семьи.

Ключевые слова: Южная Сибирь, тувинцы, тоджинцы, енисейская лингвистическая семья, генофонд, Y-хромосома, гаплогруппа Q, гаплотип, филогенетический анализ, картографический анализ.

DOI: 10.31857/S001667582112002X

Филогеографические исследования Y-хромосомы широко используются в популяционной генетике и в смежных дисциплинах (археологии, генетической генеалогии, криминалистике) для изучения миграции древнего населения, в том числе известных из небиологических источников (лингвистики, археологии, ранней истории), а также для анализа происхождения отдельных генеалогических линий. При изучении генофонда Сибири основное внимание уделяется двум гаплогруппам Y-хромосомы, которые отражают наследие древнего населения и связи между крупными частями Сибири: паттерн распространения гаплогруппы N связывает Южную и Западную Сибирь, а гаплогруппы Q – Южную и Центральную Сибирь. Особая роль населения Южной Сибири, и в первую очередь Алтае-Саянского нагорья, в формировании общей структуры населения, лингвистического и генетического разнообразия Сибири хорошо известна. В предыдущих работах мы изучили генофонд популяций Саян (тувинцев и тоджинцев) по панели основных гаплогрупп Y-хромосомы [1–3], в том числе определены частоты гаплогруппы Q. Предметом данного исследования является углубленный анализ филогенетики и филогеографии гаплогруппы Q у тувинцев и

других популяций Южной Сибири и смежных регионов на основе глубокого субтипирования нескольких десятков SNP-маркеров, определяющих отдельные линии внутри гаплогруппы Q.

Ввиду значительного возраста гаплогруппы Q-M242 (около 30 тыс. лет) и разного паттерна распространения ее ветвей в Евразии и Америке (она охватывает практически весь генофонд американских индейцев и является маркером первичного заселения человеком Нового Света) сохраняется широкий интерес исследователей и к истории этой гаплогруппы на ее "прародине" – в Сибири. В связи с этим продолжаются работы и попытки изучения филогеографии гаплогруппы Q-M242 на современном уровне – с использованием данных NGS-анализа в современных и древних популяциях. Например, работа [4] является практически исчерпывающей по охвату ветвей гаплогруппы Q в популяциях Евразии и Америки, не фокусируясь на какой-либо из них. Основная задача исследования состояла в изучении доисторических миграций между континентами на основе подробного филогенетического дерева гаплогруппы Q. При этом вопросы распространения конкретных ветвей внутри отдельных регионов, включая Сибирь, практически не рассматривались.

В других недавних работах анализ гаплогруппы Q не был основным предметом исследования, а являлся частью более широкой задачи – изучения генофонда Северо-Восточной Азии [5] или в целом Восточной Азии [6] на основе полногеномных данных о современных и древних популяциях. В работе [5] основные выводы связаны с динамикой населения в Сибири в период от позднего палеолита до неолита на основе полногеномных панелей SNP-маркеров, но также включают и результаты анализа Y-хромосомы у некоторых древних индивидов (обнаружены только гаплогруппы Q или N). Исследование [6] охватывает территории Монголии (в том числе приграничье с Тувой), Северного Китая, Приамурья и Японии, включает результаты анализа образцов древней ДНК за последние 8 тыс. лет, среди которых встречена и гаплогруппа Q-M242.

В более локальном географическом (территория Тувы) и хронологическом (последние 2.5 тыс. лет) масштабе изучены древние популяции скифского времени в работе [7]. Показано, что сообщества сибирских скифов были организованы как патрилокальные и патрилинейные (в мужских захоронениях обнаружены близкие родственники), по структуре генофонда эти группы оказались более близки к населению Азии, чем к западным группам скифов. В данной работе у пяти индивидов захоронения Аржан (раннескифского времени) обнаружена гаплогруппа Y-хромосомы Q-L330 (на филогенетическом дереве – общий корень линий YP1102 и YP771), однако сравнение полученных данных проведено преимущественно в масштабе древнего населения других регионов и эпох.

Пожалуй, единственным примером изучения линий высокого разрешения внутри гаплогруппы Q-M242 в Южной Сибири является работа исследователей из Томского НИМЦ, представленная на одной из конференций [8]. В.Н. Харьковым и коллегами было изучено распространение двух относительно недавних линий (Q-B31 и Q-B33) в Западной и Южной Сибири. Исследователи пришли к заключению, что территория Тувы являлась источником происхождения данных линий Y-хромосомы, причем этот общий генетический компонент находит параллели и в распространении аутосомных маркеров.

Мы поддерживаем гипотезу южносибирского происхождения тех ветвей гаплогруппы Q, которые распространены одновременно и в Южной, и в Центральной, и в Западной Сибири. Для более глубокого изучения этого вопроса необходимо подробное изучение распространения различных ветвей гаплогруппы Q в популяциях Южной Сибири. До настоящего времени данных генотипирования популяций Южной Сибири по панели маркеров, основанной на подробной филогении гаплогруппы Q-M242, в научной литературе не представлено. Поэтому целью данного исследования было изучение филогенетической структуры и филогеографического паттерна распространения различных линий гаплогруппы Q-M242 в популяциях Южной Сибири и смежных регионов, а средством достижения этой цели стало генотипирование более полусотни SNP-маркеров, отражающих глубокую филогенетическую структуру гаплогруппы, в сопоставлении с данными по STR-маркерам.

МАТЕРИАЛЫ И МЕТОДЫ

Изучены 545 образцов ДНК из пяти популяций Саянского нагорья, представляющих разные географические группы: западных, центральных, южных и юго-восточных тувинцев, а также тоджинцев, выборки охарактеризованы в табл. 1. Образцы были собраны Л.Д. Дамба под руководством Е.В. Балановской согласно правилам Биобанка Северной Евразии [9] в экспедиционных обследованиях 2015–2017 гг. как описано ранее [1–3, 10, 11]. В выборки включены только неродственные между собой индивиды, все предки которых на протяжении не менее трех поколений относили себя к данному этносу и родились на данной территории. Сбор образцов сопровождался письменным информированным согласием обследуемых под контролем Этической комиссии ФГБНУ "МГНЦ".

Частоты основных гаплогрупп Y-хромосомы в популяциях тувинцев были представлены ранее [2], в том числе было показано, что гаплогруппа

Таблица 1. Частоты гаплогруппы Q-M242 и трех линий, обнаруженных в популяциях тувинцев

Популяция	*N*	Q-M242	YP1102			Другие гаплогруппы
				BZ99		
			YP1102*	BZ99*	YP1691	
Тувинцы западные (Барун-Хемчикский к.)	89	0.06	0.00	0.00	0.06	0.94
Тувинцы южные (Эрзинский к.)	92	0.12	0.04	0.00	0.08	0.88
Тувинцы центральные (Тандынский к.)	90	0.04	0.00	0.00	0.04	0.96
Тувинцы центральные (Чаа-Хольский к.)	94	0.20	0.02	0.02	0.16	0.80
Тувинцы юго-восточные (Тере-Хольский к.)	90	0.26	0.08	0.00	0.17	0.74
Тувинцы-тоджинцы (Тоджинский к.)	90	0.51	0.00	0.00	0.51	0.49
Средняя частота		*0.20*	*0.02*	*0.00*	*0.17*	*0.80*
Итого	*545*					

Примечание. По данным Всероссийской переписи населения 2010 г. численность тувинцев в Республике Тыва составляет около 247 тыс. человек, из них тувинцев-тоджинцев – около 1.9 тыс. человек.

Q-M242 составляет заметную часть генофонда саянских популяций. В данной работе проведено углубленное изучение гаплогруппы Q у тувинцев и окружающих популяций Сибири на основе генотипирования 58 SNP-маркеров, маркирующих разные филогенетические линии внутри гаплогруппы Q-M242. Для субгаплогруппы YP1102, выявленной у тувинцев, проведено генотипирование той же панели 58 маркеров у окружающих популяций, суммарная проанализированная выборка носителей Q-YP1102 составила 394 образца. Из них 136 образцов были дополнительно генотипированы по панели STR-маркеров.

Материалом для исследования послужила геномная ДНК, выделенная из образцов венозной крови методом фенол-хлороформной экстракции и с помощью автоматизированной системы QiaSymphony.

Принадлежность образцов ДНК к линиям внутри гаплогруппы Q-M242 Y-хромосомы определялась по данным анализа 58 SNP-маркеров: B143, B280, B31, BZ1000, BZ1030, BZ1499, BZ1520, BZ3140, BZ321, BZ3908, BZ427, BZ431, BZ664, BZ99, F1893, L275, L330, L332, L53, L54, L712, L713, L804, L940, M25, M3, M378, SK1995, Y1230, Y12452, Y18091, Y2137, Y2200, Y2211, Y2659, Y2677, Y2990, Y5186, Y521, Y558, Y661, YP1095, YP1102, YP1226, YP1669, YP1690, YP1691, YP3952, YP4000, YP4004, YP4055, YP4385, YP4387, YP745, YP771, YP812, Z35990, Z36048. Генотипирование проведено с помощью технологии OpenArray на приборе QuantStudio 12KFlex, позволяющей проводить генотипирование методом ПЦР в реальном времени для большого числа образцов по всем SNP-маркерам. Все перечисленные SNP были включены в кастомную OpenArray панель.

Фрагментный анализ 17 STR-локусов проведен на приборе ABI 3130xl (Applied Biosystems) с использованием набора Y-filer PCR Amplification Kit (Applied Biosystems). Для сравнения привлечены гаплотипы южных алтайцев, хакасов, изученные в [12], а также наши данные для сибирских татар и народов Монголии.

Филогенетическая сеть построена в программе Network 10.2.0.0 [13] по данным о 15 STR-маркерах (DYS389I, DYS389II, DYS390, DYS456, DYS19, DYS458, DYS437, DYS438, DYS448, GATAH4, DYS391, DYS392, DYS393, DYS439, DYS635) и по трем SNP-маркерам (YP1102, BZ99, YP1691; использовались с более высоким значением веса), оформлена в программе Network Publisher v 2.1.2.5 [13].

Возраст кластеров на основе STR-маркеров оценен по числу мутационных шагов (ρ-статистика) [14] с использованием длины поколения 31 год [15] и двух скоростей мутирования: “генеалогической” и “комбинированной”. “Генеалогическая” скорость (0.027 мутаций на локус за поколение) учитывает накопление мутаций в среднем для разных STR-локусов Y-хромосомы, ее значе-

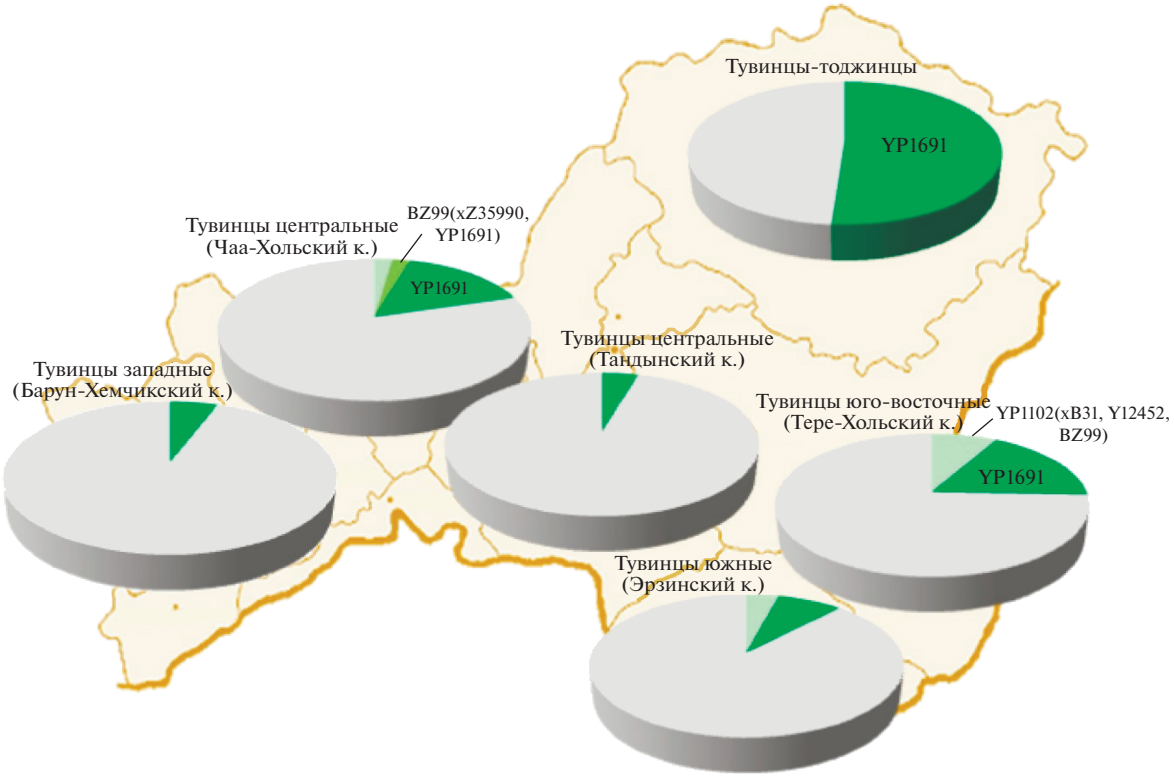

Рис. 1. Спектр ветвей гаплогруппы Q-M242 у территориальных групп тувинцев.

ние было получено в исследованиях при прямом подсчете мутаций в известных родословных [16]. “Комбинированная” скорость (0.042 мутаций на локус за поколение) представляет собой сумму скоростей мутирования для каждого из проанализированных 15 STR-локусов, данные получены с ресурса YHRD.org (Available at: https://yhrd.org/pages/resources/mutation_rates. Accessed: 25.08.2020). Датировки для линий YP1102, BZ99, YP1691 приводятся по данным компании Yfull (рассчитаны при анализе полных последовательностей Y-хромосом).

Также был проведен анализ опубликованных данных о древней ДНК с гаплогруппой Q-M242. Были проанализированы BAM файлы из четырех работ [5, 6, 17, 18] и был установлен итоговый генотип до уровня маркера YP1691. Для двух индивидов из работ [5, 6] принадлежность к линии YP1102(xB31, Y12452, BZ99) показана в [19].

Карты распространения линий гаплогруппы Q-M242 построены с помощью оригинального картографического пакета “GeneGeo” [20] методом средневзвешенной интерполяции с радиусом влияния 1000 км и значением весовой функции, равным 3 (подробное описание технологии и процесса создания карт приведено в [21]).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Разнообразие ветвей гаплогруппы Q-M242 в популяциях Саян

Результаты анализа 58 SNP-маркеров внутри гаплогруппы Q-M242 выявили в генофонде тувинцев и тоджинцев только три линии – YP1102(xB31, Y12452, BZ99), BZ99(xZ35990, YP1691) и YP1691, восходящие к одному корню – варианту YP1102 (табл. 1). Время появления SNP-мутации YP1102, оцененное по полногеномным данным, составляет около 4.6 тыс. лет назад [19].

Наиболее частой из трех является линия YP1691: она охватывает 17% генофонда тувинцев в целом, при этом достигая максимума на северо-востоке у тоджинцев (51%), сохраняя частоту, близкую к средней, у юго-восточной и центральной групп (Тере-Хольского и Чаа-Хольского кожуунов Тувы) и снижаясь в юго-западном направлении (до 8% – у южных тувинцев, 6% – у западных и до 4% – у популяции центральных тувинцев Тандынского кожууна; рис. 1).

Наиболее редкой из трех является ветвь BZ99(xZ35990, YP1691), которая вместе с YP1691 восходит к общему SNP-маркеру BZ99. Линия BZ99(xZ35990, YP1691) встречена только у двух представителей центральной группы тувинцев Чаа-Хольского кожууна (2%), она может включать в себя один или два пока неизвестных вари-

анта (запись “BZ99(xZ35990, YP1691)” обозначает генотип, относящийся к “материнской” линии BZ99, но исключающий принадлежность к “дочерним” линиям Z35990 и YP1691; далее по тексту упоминается в виде BZ99*).

Наконец, третья из обнаруженных ветвей – линия YP1102(xB31, Y12452, BZ99), далее по тексту упоминается как YP1102* – встречается у трех из шести популяций тувинцев: центральных Чаа-Хольского кожууна (2%), южных (4%) и юго-восточных (8%, рис. 1). Частота данной линии сравнительно невелика, но при этом, как и для гаплогруппы Q-M242 в целом, прослеживается понижение доли YP1102* у тувинцев в направлении с востока на запад.

Наличие двух линий гаплогруппы Q-M242 на разных концах ареала тувинцев может указывать на их наследование от одной прапопуляции, в которой могли быть представлены и предковая ветвь YP1102, и разные варианты внутри нее. В таком случае присутствие только одной линии YP1691 у тувинцев-тоджинцев (где гаплогруппа Q-M242 охватывает половину генофонда) можно полагать следствием дрейфа генов в изолированной популяции, издревле проживающей в горно-таежной местности. В пользу этого свидетельствует наличие двух ветвей гаплогруппы Q (YP1102* и YP1691) у южных соседей тоджинцев – юго-восточных и южных тувинцев. Иными словами, у древнего населения восточных областей Тувы, на основе которого сформировались тоджинцы, юго-восточные и южные тувинцы, в генофонде могли быть представлены обе линии гаплогруппы Q, но впоследствии у тоджинцев в результате дрейфа генов сохранилась (и значительно увеличилась в частоте) только одна – линия YP1691. Однако подобное отразилось бы на внутренней структуре линии YP1691 в популяции тоджинцев, например привело бы к снижению разнообразия STR-гаплотипов и увеличению числа носителей одного гаплотипа. На представленной далее филогенетической сети рассмотрено разнообразие STR-гаплотипов Y-хромосомы в пределах линии YP1102 не только у разных групп тувинцев, но среди и носителей этого маркера из других популяций Сибири.

Филогенетический анализ Q-YP1102 по STR-маркерам

Популяционный скрининг маркеров внутри гаплогруппы Q-M242 в популяциях Северной Евразии позволил обнаружить линию YP1102 (и три ветви внутри нее, встреченные у тувинцев) среди других народов Сибири. Результаты филогенетического анализа для носителей линий YP1102*, BZ99* и YP1691 (рис. 2) представили возможность изучить их разнообразие на основе другой системы генетических маркеров – “быстромутирующих” STR Y-хромосомы.

Филогенетическая сеть (рис. 2) представлена в двух вариантах: в верхнем блоке (рис. 2,*а*) разными цветами показаны гаплотипы носителей каждой из трех линий гаплогруппы Q-M242, в нижнем блоке (рис. 2,*б*) разные цвета обозначают популяции, из которых происходят носители гаплотипов. На данной сети (рис. 2,*а*) выделяются два основных кластера, соответствующих двум ветвям гаплогруппы Q-M242: более обширный YP1691 и небольшой по числу носителей BZ99*. Гаплотипы носителей третьей линии YP1102* рассеяны на филогенетической сети и отчетливой кластеризации для них не наблюдается.

Кластер BZ99* включает преимущественно гаплотипы хакасов-койбалов, но также – и двух тувинцев центральной группы из Чаа-Хольского кожууна (рис. 2,*б*). Возвращаясь к ранее выдвинутому вопросу, можно сделать вывод: линия BZ99* у тувинцев – это одна ранее неизвестная ветвь (не две), и встречается она также у хакасов. Время формирования данного кластера составляет около 1.5 тыс. лет (средняя двух оценок, табл. 2), а время возникновения самой SNP-мутации BZ99 – около 3.9 тыс. лет назад (на основе полногеномных данных компании Yfull [19]). Возможно, что эта ветвь проникла в генофонд тувинцев при контактах с хакасами, либо же является наследием древней популяции, проживавшей в ареалах обоих народов.

Более обширный кластер YP1691 (рис. 2,*б*) кроме тувинцев включает гаплотипы сибирских татар, южных алтайцев, хакасов, народов северо-запада Монголии – урянхайцев, халха, захчин. Данный кластер звездчатой формы, с выраженным гаплотипом основателя (довольно частым: 39 носителей, представлен почти у всех популяций из кластера). Датировка кластера YP1691, рассчитанная на основе данных об STR-маркерах, составляет в среднем около 1 тыс. лет (средняя двух оценок, табл. 2), а датировка появления SNP-мутации YP1691 – около 2.3 тыс. лет назад (на основе полногеномных данных компании Yfull [19https://www.yfull.com/tree/Q-YP1102/]).

Более половины гаплотипов кластера YP1691 (14 из 26, в том числе – и гаплотип основателя) встречено у тувинцев-тоджинцев (рис. 2,*б*). При этом заметное накопление образцов в пределах одного гаплотипа прослеживается только для двух групп (обозначены цифрами 1 и 2 на рис. 2,*б*). Гаплотип 1 ($N = 9$) встречен у двух тувинцев из центральной и юго-восточной групп, а среди тувинцев-тоджинцев – у семи индивидов, четыре из которых представляют род *бараан*, другие носители – рода *маады*, *чооду*, *ак*. Иными словами, накопление этого гаплотипа у тувинцев-тоджинцев связано именно с принадлежностью большинства

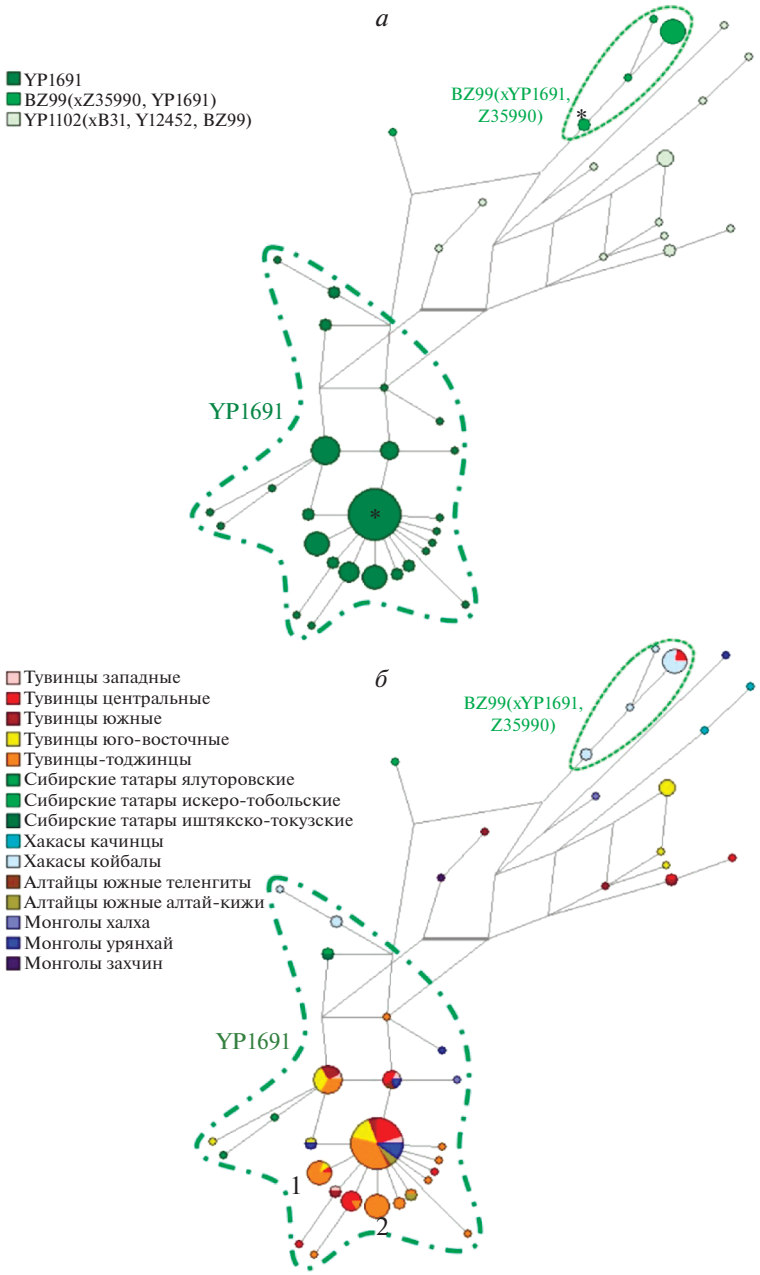

Рис. 2. Филогенетическая сеть гаплогруппы Q-YP1102 в населении Сибири. *а* – принадлежность гаплотипов к субтипам гаплогруппы Q (согласно легенде слева); *б* – принадлежность носителей гаплотипов к популяциям Сибири (согласно легенде слева).

носителей к одной родовой группе. Гаплотип 2 (N = = 9) встречен только у тувинцев-тоджинцев, в пяти родовых группах: *кол* (N = 3), *хойук* (N = 2), *ак* (N = 2), единичные – *бараан* и *чооду*. В данном случае некоторое увеличение числа носителей также наблюдается в пределах первых трех родовых групп.

Отметим, что и в других популяциях тувинцев из кластера YP1691 встречаются носители нескольких разных гаплотипов. Например, у западных тувинцев, где общая частота линии YP1691 составляет лишь 6% (пять носителей в популяции), встречено четыре разных гаплотипа (рис. 2,*б*); аналогичная закономерность прослеживается для юго-восточных, южных и центральных групп.

Иными словами, во всех изученных популяциях тувинцев сохраняется заметное разнообразие гаплотипов и отчетливого указания на последствия дрейфа генов (как причины повышения частоты линии) не наблюдается. Напротив, структура кластера YP1691 (много производных гаплотипов, отходящих от основателя) отражает скорее послед-

Таблица 2. Датировки кластеров STR-гаплотипов на филогенетической сети

Кластер	Состав кластера	Количество образцов в кластере	Количество гаплотипов в кластере	Специфичность кластера (Sp)	Возраст, число мутаций ($\rho \pm \sigma$)	Возраст кластера, лет (по STR-гаплотипам, "генеалогическая" скорость)	Возраст кластера, лет (по STR-гаплотипам, "комбинированная" скорость)	Возраст SNP-линии, лет (по Yfull)
YP1691	Тувинцы	106	26	81%	1.04 ± 0.35	1191 ± 400	771 ± 260	2300 ± 800
BZ99*	Хакасы	13	4	85%	1.62 ± 1.10	1855 ± 1260	1200 ± 815	3900 ± 900

ствия демографического роста в прапопуляции, из которой данная линия гаплогруппы Q проникла в генофонды не только разных территориальных групп тувинцев, но и других народов Сибири. Например, самый частый гаплотип (он же – "основатель" кластера YP1691) встречается, кроме тувинцев, у южных алтайцев (двух алтай-кижи, одного теленгита) и урянхайцев Монголии ($N = 4$, таблица Приложения), а производные от основателя гаплотипы распространены еще шире (среди хакасов и сибирских татар).

Носители линии YP1691 из популяций сибирских татар и хакасов, находясь в составе одноименного кластера, тем не менее отдалены от гаплотипа основателя и ближайших к нему (рис. 2,*б*). Возможно, что при увеличении числа образцов из популяций Западной Сибири на филогенетической сети может проявиться отдельный – специфичный для них – кластер.

Филогеографический анализ распространения Q-YP1102 в Сибири

Карты распространения рассматриваемых ветвей гаплогруппы Q-M242 в Сибири представлены на рис. 3: линии YP1102 в целом (рис. 3,*а*) и ее ветвей (рис. 3,*б–г*).

Самая недавняя по датировке (2.3 тыс. лет; табл. 2) и частая у тувинцев ветвь YP1691 выявлена нами и у многих других народов (рис. 3): на северо-западе Монголии (урянхайцы, частота 18%), в Южной Сибири (алтай-кижи – 8%, теленгиты – 1.5%, хакасы – 2.5%) и Западной Сибири (тоболо-иртышские и барабинские сибирские татары – до 3.5%, ханты – 13%).

Ветвь BZ99* в Сибири обнаруживается только на отдельных территориях, не образующих непрерывного ареала (рис. 3): с наибольшей частотой встречена у хакасов (9%), у тоболо-иртышских татар составляет 3%, у хантов – 1%. Напомним, что BZ99* – самая редкая линия гаплогруппы Q у тувинцев: она встречена только в центральной популяции Чаа-Хальского кожууна у двух индивидов (рис. 1,*а*).

Базальная ветвь YP1102* за пределами ареала тувинцев встречается с низкой частотой: у народов Монголии (захчин, урянхайцы, халха) – с частотой 1–3%, у хакасов – 2%, у барабинских сибирских татар – 3% (рис. 3).

Итак, паттерн трех карт достаточно близок: хотя бы с низкими частотами, но каждая ветвь представлена и в Южной, и в Западной Сибири. Ареал ветвей YP1102* и BZ99* входит в область распространения самой молодой ветви YP1691, поэтому обобщенная карта распространения всей линии YP1102 незначительно отличается от таковой для YP1691 (рис. 3).

В научной литературе о географии линии YP1102 и ее ветвей представлены только фрагментарные данные. О присутствии линии YP1102 в генофонде кетов и селькупов – коренного населения междуречья Оби и Енисея – известно из работы [4] по данным полногеномного анализа. О присутствии линии B33 (ветвь внутри линии YP1691 [ISOGG]) в генофонде кетов, селькупов и ряда популяций Южной Сибири и ее корреляции с "енисейским" генетическим компонентом изложено в сообщении В.Н. Харькова и соавт. [8]. Таким образом, представленные в данном исследовании карты и сведения из научной литературы указывают на распространение и ветви YP1691, и всей линии YP1102 на территории также Центральной Сибири. Принимая во внимание датировку ветви YP1691 (около 2.3 тыс. лет), следует предполагать, что ее носители расселились столь широко (территории Южной, Западной и Центральной Сибири) в течение последних двух тысячелетий. Менее очевидно, какая территория явилась источником распространения этой ветви. Присутствие в генофонде и тувинцев, и хакасов всех трех ветвей (YP1102*, BZ99*, YP1691) может указывать на Южную Сибирь как прародину всей линии YP1102.

Данные NGS-анализа Y-хромосом из древних популяций Сибири позволили обнаружить базальную ветвь YP1102* у пяти мужчин: четырех представителей бронзового века Прибайкалья (образцы DA339, DA343, DA356 из работ [17, 18],

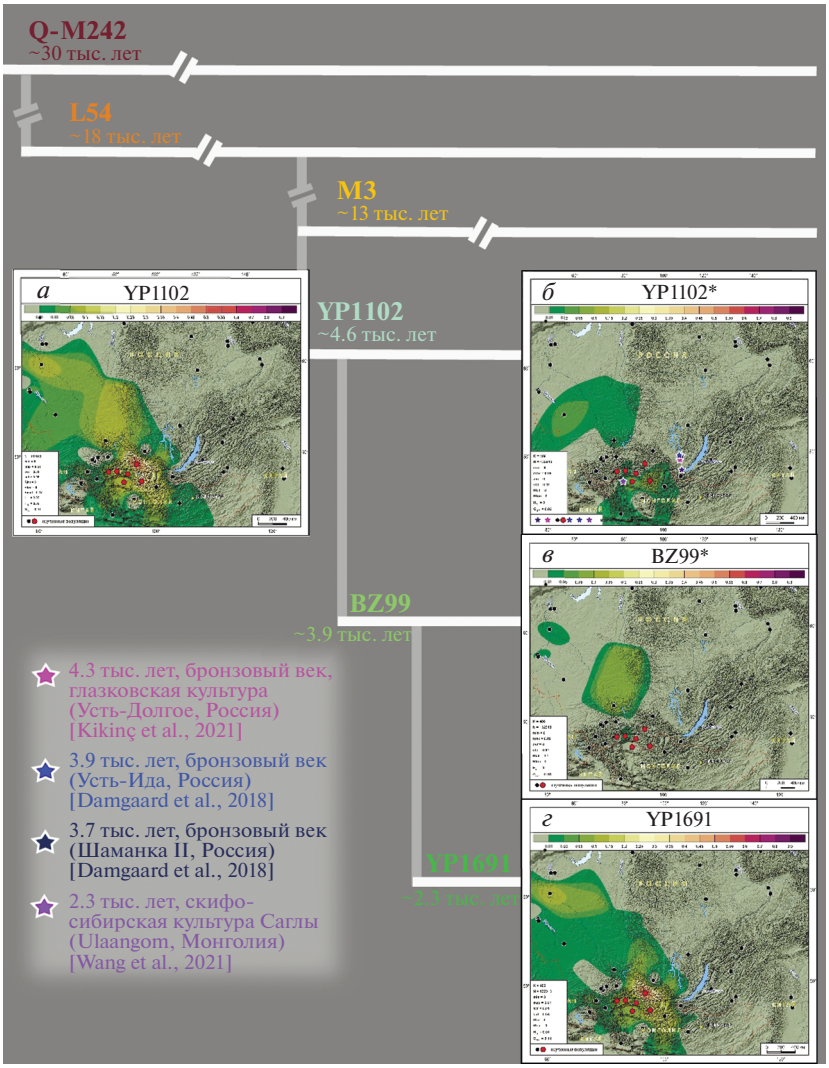

Рис. 3. Филогенетическая схема гаплогруппы Q-M242 с указанием датировок (по данным Yfull) и карт распространения в Сибири трех изученных линий. *а* – карта суммарной частоты линии YP1102; *б* – карта распространения линии YP1102*; *в* – карта распространения линии BZ99*; *г* – карта распространения линии YP1691. На всех картах изученные популяции тувинцев показаны красными точками, все группы сравнения (современное население), для которых проведен скрининг той же панели маркеров, – черными точками. Значение частоты генотипа на карте соответствует цветовой шкале в верхнем поле каждой карты. На карте линии YP1102* звездами разного цвета обозначены древние популяции, у представителей которых найдена данная ветвь гаплогруппы Q, расшифровка данных обозначений приведена на рисунке слева (датировки и периоды, названия археологических культур и стоянок, ссылки на источник данных).

образец irk022 из [5]) и одного индивида из могильника культуры Саглы скифо-сибирского типа на границе Монголии и Тувы [6]. На карте ветви YP1102* обозначены места археологических стоянок, где были обнаружены носители данной линии гаплогруппы Q-M242 (рис. 2,*б*), и датировки (округленные значения радиоуглеродных дат, приведенных в указанных публикациях). Иными словами, линия YP1102* присутствовала среди древнего населения Сибири разного географического положения и разных периодов времени.

Можно сделать вывод, что большинство известных древних представителей ветви YP1102* проживало хотя и за пределами современного распространения данной линии Y-хромосомы, но в регионе, непосредственно примыкающем к ее ареалу – на территории Прибайкалья, около 4 тыс. лет назад (рис. 3,*б*). Три из четырех индивидов (DA343, DA356, irk022) происходят из населения глазковской археологической культуры [5, 18].

Пятый носитель линии YP1102* – представитель скифского мира и населения, проживавшего почти в пределах современного ареала тувинцев более 2 тыс. лет назад (рис. 2,*б*). О наличии линии YP1102* на данной территории еще раньше – в

бронзовом веке – пока можно только предполагать.

Присутствие в населении Южной Сибири базальной линии YP1102* в течение последних 4 тыс. лет (документированной на этой территории и около 4 тыс. лет назад, и около 2 тыс. лет назад, и в настоящее время), широкая представленность ветвей BZ99* и YP1691 в современных популяциях региона (преимущественно у тувинцев) могут указывать на вероятное происхождение всей линии YP1102 из данного ареала. В таком случае наличие линии YP1102 (и ее ветви YP1691) в генофонде современных популяций кетов и селькупов может указывать на происхождение части их предков из древнего населения Южной Сибири. Это согласуется с лингвистическими данными о территориально более южном происхождении народов енисейской языковой семьи [22], единственным современным представителем которой являются кеты.

Структура кластера YP1691 на филогенетической сети свидетельствует о росте численности населения Южной Сибири позднего железного века. Широкий ареал линии YP1691, вероятно, является следствием активных миграционных процессов в течение последних двух тысяч лет, которые привели к ее распространению из Южной Сибири на север и запад.

Авторы выражают благодарность всем донорам образцов, которые принимали участие в данном исследовании, а также АНО “Биобанк Северной Евразии” за предоставление коллекций ДНК.

Исследование выполнено при финансовой поддержке Министерства науки и образования РФ (Госконтракт № 011-17 от 26.09.2017) в рамках научно-технической программы Союзного государства “ДНК-идентификация” (работы по генотипированию панелей SNP-маркеров, картографический анализ), Государственного задания Министерства науки и высшего образования РФ для Медико-генетического научного центра им. акад. Н.П. Бочкова (работы по генотипированию STR-маркеров, филогенетический анализ), Государственного задания Министерства науки и высшего образования РФ для Института общей генетики им. Н.И. Вавилова РАН (интерпретация результатов, написание текста статьи).

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

1. *Дамба Л.Д., Балановская Е.В., Агджоян А.Т. и др.* Генофонд трех восточных родов тувинцев по данным полиморфизма Y-хромосомы // Вестник Моск. ун-та. Серия XXIII. Антропология. 2019. № 1. С. 74–85. https://doi.org/10.32521/2074-8132.2019.1.074-085
2. *Балановская Е.В., Дамба Л.Д., Агджоян А.Т. и др.* Генофонд охотников-оленеводов Южной Сибири: тофалары и тоджинцы // Вестник Моск. ун-та. Серия XXIII. Антропология. 2019. № 4. С. 67–80. https://doi.org/10.32521/2074-8132.2019.4.067-080
3. *Жабагин М.К., Дамба Л.Д., Короткова Н.А. и др.* Анализ родоплеменной структуры тувинцев по маркерам Y-хромосомы // Генетика. 2020. Т. 56. № 6. С. 732–738. https://doi.org/10.31857/S0016675820060132
4. *Grugni V., Raveane A., Ongaro L. et al.* Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas // BMC Biol. 2019. V. 17. № 3. https://doi.org/10.1186/s12915-018-0622-4
5. *Kılınç G.M., Kashuba N., Koptekin D. et al.* Human population dynamics and *Yersinia pestis* in ancient northeast Asia // Sci. Adv. 2021. V. 7(2):eabc4587. https://doi.org/10.1126/sciadv.abc4587
6. *Wang C.C., Yeh H.Y., Popov A.N. et al.* Genomic insights into the formation of human populations in East Asia // Nature. 2021. V. 591. P. 413–419. https://doi.org/10.1038/s41586-021-03336-2
7. *Mary L., Zvénigorosky V., Kovalev A. et al.* Genetic kinship and admixture in Iron Age Scytho-Siberians // Hum. Genet. 2019. V. 138. P. 411–423. https://doi.org/10.1007/s00439-019-02002-y
8. *Харьков В.Н., Новикова Л.М., Хитринская И.Ю., Степанов В.А.* Енисейский генетический субстрат в генофонде коренного населения Южной и Западной Сибири // VII Съезд Вавиловского общества генетиков и селекционеров (ВОГиС), посвященный 100-летию кафедры генетики СПбГУ, и ассоциированные симпозиумы: медж. конгресс (Санкт-Петербург, 18–22 июня 2019). Санкт-Петербург: Изд-во ВВМ, 2019. С. 337.
9. *Балановская Е.В., Жабагин М.К., Агджоян А.Т. и др.* Популяционные биобанки: принципы организации и перспективы применения в геногеографии и персонализированной медицине // Генетика. 2016. Т. 52. № 12. С. 1371–1387. https://doi.org/10.7868/S001667581612002X
10. *Дамба Л.Д., Айыжы Е.В., Монгуш Б.Б. и др.* Комплексный подход в изучении родовой структуры тувинцев Республики Тыва на примере родов монгуш и ооржак // Вестник Тув. гос. ун-та. Вып. 2. Естеств. и сельскохоз. науки. 2018. № 2(37). С. 37–44.
11. *Дамба Л.Д., Айыжы Е.В., Монгуш Б.Б. и др.* Краткий обзор некоторых родоплеменных групп тувинцев по данным междисциплинарных исследований // Вестник Тув. гос. ун-та. Вып. 2. Естеств. и сельскохоз. науки. 2019. № 2(45). С. 19–30. https://doi.org/10.24411/2077-5326-2019-10003
12. *Балаганская О.А.* Полиморфизм Y-хромосомы у тюркоязычного населения Алтая, Саян, Тянь-Ша-

ня и Памира в контексте взаимодействия генофондов Западной и Восточной Евразии: Автореф. дис. ... канд. биол. наук. М.: МГНЦ РАМН, 2011. 27 с.

13. Электронный ресурс: http://www.fluxus-engineering.com (дата обращения – 28.12.2020).
14. *Forster P., Harding R., Torroni A., Bandelt H.J.* Origin and evolution of Native American mtDNA variation: a reappraisal // Am. J. Hum. Genet. 1996. V. 59. № 4. P. 935–945.
15. *Fenner J.N.* Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies // Am. J. Phys. Anthropol. 2005. V. 128. P. 415–423. https://doi.org/10.1002/ajpa.20188
16. *Балановский О.П., Запорожченко В.В.* Хромосома-летописец: датировки генетики, события истории, соблазн ДНК-генеалогии // Генетика. 2016. Т. 52. № 7. С. 810–830. https://doi.org/10.7868/S0016675816070043
17. *Damgaard P.D.B., Marchi N., Rasmussen S. et al.* 137 ancient human genomes from across the Eurasian steppes // Nature. 2018. V. 557. P. 369–374. https://doi.org/10.1038/s41586-018-0094-2
18. *Damgaard P.D.B., Martiniano R., Kamm J. et al.* The first horse herders and the impact of early Bronze Age steppe expansions into Asia // Science. 2018. V. 360(6396). https://doi.org/10.1126/science.aar7711
19. Электронный ресурс: Филогенетическое дерево линии Q-YP1102 на сайте компании Yfull https://www.yfull.com/tree/Q-YP1102/ (дата обращения – 28.05.2021).
20. *Кошель С.М.* Геоинформационные технологии в геногеографии // Современная географическая картография. М.: Дата+, 2012. С. 158–166.
21. *Балановский О.П.* Генофонд Европы. М.: Тов-во науч. изд. КМК, 2015. 354 с.
22. *Vajda E.J.* Yeniseian Peoples and Languages: A History of Their Study with an Annotated Bibliography and a Source Guide. England: Curzon Press, 2001. 389 p.

Phylogenetic Analysis of the South Siberian Haplogroup Q-YP1102 from Y-SNP and Y-STR-Data on Tuvinian and Surrounding Populations

A. T. Agdzhoyan[a, b, *], L. D. Damba[d], V. M. Gurianov[a], V. V. Zaporozhchenko[a, b], and O. P. Balanovsky[a, b, c]

[a] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia*

[b] *Research Centre for Medical Genetics, Moscow, 115522 Russia*

[c] *Biobank of North Eurasia, Moscow, 115201 Russia*

[d] *Research Institute of Medical and Social Problems and Control, Kyzyl, 667003 Russia*

**e-mail: aagdzhoyan@gmail.com*

Haplogroup Q-M242 forms the majority of the Native American Y chromosomes, reaching its highest frequency in Eurasia in the populations of Central and Southern Siberia. We performed genotyping with a panel of 60 SNP markers within haplogroup Q and carried out an in-depth phylogeographic analysis in the populations of Southern Siberia and adjacent regions. The study of 6 populations of Tuvans and Tojins revealed three branches of haplogroup Q emerging from a single root (YP1102). The most common of them, the lineage YP1691, reaches its maximum in the northeastern Tuvans (51%) and decreases in frequency from east to west to 4%. A similar trend is followed by the distribution of the basal lineage YP1102*. The third branch BZ99* is rare in Tuvans (2%), but is noticeable in the Khakas people (9%). Phylogenetic analysis of STR haplotypes showed the distribution of the same clusters not only in different groups of Tuvans, but also among the surrounding populations. The structure of the main cluster indicates an increase in the number of YP1691 carriers during the Late Iron Age. The frequency map of the lineage YP1691 reflects the results of active migrations in Southern and Western Siberia during the last millenia. In the ancient populations of Siberia, the YP1102* variant was found in five individuals: four of them are inhabitants of the Bronze Age Baikal region and one is a representative of a Scythian type culture on the border of Mongolia and Tuva. Ancient DNA data and a higher genetic diversity in the population of Southern Siberia suggest the origin of haplogroup Q-YP1102 in this region and its further penetration to the west and north of Siberia, including possibly the route of the Yeniseian-speaking populations.

Keywords: Southern Siberia, Tuvans, Tojins, Yeniseian languages, gene pool, Y-chromosome, haplogroup Q, haplotype, phylogenetic analysis, cartographic analysis.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1396–1402

УДК 575.174:599.9

ГЕНЕТИЧЕСКОЕ РАЗНООБРАЗИЕ 21 АУТОСОМНОГО STR-МАРКЕРА СИСТЕМЫ CODIS В ПОПУЛЯЦИЯХ ВОСТОЧНОЙ ЕВРОПЫ[1]

© 2021 г. В. Н. Харьков[1, *], С. А. Котова[2], Н. А. Колесников[1], Е. А. Спивак[2], К. В. Вагайцева[1], Т. В. Забавская[2], В. И. Рыбакова[2], А. С. Парфенова[2], А. Н. Верчук[2], Н. К. Янковский[3, 4], В. А. Степанов[1]

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

[2]Научно-практический центр Государственного комитета судебных экспертиз Республики Беларусь, Минск, 220114 Беларусь

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

[4]Московский государственный университет им. М.В. Ломоносова, Москва, 119991 Россия

**e-mail: vladimir.kharkov@medgenetics.ru*

Поступила в редакцию 07.06.2021 г.
После доработки 15.06.2021 г.
Принята к публикации 17.06.2021 г.

Впервые получены генетические характеристики 15 локальных популяций восьми народов Восточной Европы по расширенной панели, включающей 21 аутосомный STR-маркер, применяемый в судебно-экспертной практике для ДНК-идентификации и установления родства. Проведена оценка полиморфизма маркеров в российских, белорусских и молдавских популяциях, охарактеризованы уровни внутри- и межпопуляционной генетической дифференциации исследованных групп населения. Рассчитаны параметры информационной значимости системы аутосомных микросателлитов для экспертной ДНК-идентификации и установления родства. Создана база данных частот аллелей аутосомных STR-локусов, которая может быть использована для проведения вероятностно-статистических расчетов при оценке уровня достоверности экспертного исследования в странах Союзного государства. Обоснована необходимость учета популяционно-генетической структуры населения при вероятностной оценке достоверности результатов судебно-экспертного ДНК-анализа.

Ключевые слова: микросателлитный локус, частоты аллелей, популяция, генетическая дифференциация, ДНК-идентификация, криминалистические параметры.

DOI: 10.31857/S0016675821120055

ВВЕДЕНИЕ

Каждый индивидуум несет уникальный геном, который характеризуется неповторяющимся сочетанием аллелей множества полиморфных ДНК-маркеров. Основным методом криминалистической идентификации на базе анализа ДНК является исследование аутосомных коротких тандемных повторов (STR), обладающих высоким уровнем аллельного полиморфизма. Генотипирование аллельного состояния 15–20 локусов таких повторов обладает достаточным уровнем информативности для установления индивидуального тождества. Для анализа аутосомных STR-локусов используются наборы реагентов для генотипирования, основными производителями которых являются компании Thermo Fisher Scientific (США) и Promega Corporation (США). В основе наборов реагентов указанных производителей лежит панель маркеров стандарта CODIS (англ. Combined DNA Index System), предложенная в 2015 г. и утвержденная к использованию для криминалистического ДНК-анализа в 2017 г. Федеральным Бюро Расследований Министерства юстиции США. Несмотря на то что система CODIS разработана и функционирует на территории США, многие другие страны, в том числе Республика Беларусь и Российская Федерация, опираются на стандарт данной системы при проведении ДНК-идентификации личности в судебной экспертизе. В 2017 г. панель маркеров стандарта CODIS, до этого включавшая 13 микросателлитов, была расширена. В настоящее время в состав так называемого ядра системы входят 20 аутосомных STR-маркеров: CSF1PO, FGA, THO1, TPOX, VWA, D3S1358, D5S818, D7S820,

[1] Дополнительная информация для этой статьи доступна по doi 10.31857/S0016675821120055 для авторизованных пользователей.

D8S1179, D13S317, D16S539, D18S51, D21S11, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, D22S1045. Эти изменения привели к необходимости обновления референсных баз данных, содержащих информацию, необходимую для вероятностно-статистической оценки результатов экспертного исследования, а также актуализации информации об уровне генетической дифференциации локальных популяций по маркерам расширенной системы CODIS и информационной значимости нового ядра системы для ДНК-идентификации в ряде локальных популяций Восточной Европы. Большинство публикаций, посвященных этому вопросу, на настоящий момент содержат информацию только по 16 STR-маркерам [1–14].

В настоящей работе мы охарактеризовали частоты аллелей, уровень генетической дифференциации популяций и оценили идентификационный потенциал системы из 21 аутосомного STR-маркера (D1S1656, D2S441, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D21S11, D22S1045, CSF1PO, FGA, Penta D, Penta E, TH01, TPOX, vWA) в популяционных выборках восьми этнических групп Восточной Европы – у русских, белорусов, молдаван, гагаузов, поляков, коми, марийцев и украинцев.

МАТЕРИАЛЫ И МЕТОДЫ

Материал исследования составили образцы ДНК из 15 различных локальных популяций восьми народов Восточной Европы. Общая численность выборок составила 1093 индивида. Образцы ДНК были взяты у неродственных лиц из следующих популяций, проживающих в России (русских – 205 образцов, коми – 83, марийцев – 76), Молдавии (молдаван – 117, гагаузов – 57), Беларуси (русских – 57, белорусов – 413, поляков – 32, украинцев – 110). Материал из популяций Беларуси содержит шесть выборок из историко-этнографических регионов республики: Восточного Полесья ($N = 41$), Западного Полесья ($N = 56$), Центрального региона ($N = 135$), Поднепровья ($N = 80$), Поозерья ($N = 55$), Понеманья ($N = 46$). Коми представлены двумя выборками: коми-зыряне ($N = 43$), коми-ижемцы ($N = 40$).

Забор первичного биологического материала (венозной крови) у доноров проводили с соблюдением процедуры письменного информированного согласия на проведение исследования. На каждого донора составлялась анкета с его родословной, указанием этнической принадлежности и мест рождения предков. В исследование включены только образцы ДНК доноров, по результатам анкетирования отрицавших факт метисации с представителями других этносов минимум в трех поколениях. Индивида относили к данной этнической группе на основании его собственной этнической идентификации, его родителей и места рождения.

Генотипирование образцов по аутосомным STR-маркерам проводили методом мультиплексной полимеразной цепной реакции с последующим анализом продуктов амплификации на автоматическом генетическом анализаторе как описано ранее [10]. Экспериментальные исследования проведены на базе Центра коллективного пользования научно-исследовательским оборудованием “Медицинская геномика” Томского НИМЦРАН и НПЦ Государственного комитета судебных экспертиз Республики Беларусь.

Оценку соответствия распределений генотипов равновесию Харди–Вайнберга, генетической вариабельности STR-локусов и межпопуляционное сравнение проводили с использованием программного средства Arlequin v. 3.11 [15]. Для анализа генетической дифференциации популяций применяли расчет попарных значений F_{st} и анализ молекулярной дисперсии (AMOVA), используя в качестве матрицы генетических расстояний матрицу среднеквадратичных различий в числе повторов R_{st}.

Дискриминационный потенциал системы из 21 STR-маркера оценивали с помощью стандартных криминалистических показателей, а именно: вероятности случайного совпадения генотипов (MP, matching probability), вероятности дискриминации неродственных индивидов (PD, power of discrimination), исключающей способности (PE, power of exclusion), и индекса отцовства (PI, paternity index) [16]. Частота аллелей и криминалистические параметры рассчитывали с помощью программного пакета MS Excel.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Генетическая вариабельность 21 STR

Частоты аллелей 21 STR демонстрируют существенное разнообразие в исследованных популяциях. Наиболее частые аллели по большинству локусов во всех исследованных популяциях совпадают. Наибольшее суммарное число аллелей по 21 аутосомному маркеру было обнаружено в белорусской популяции (250), чуть меньшее число аллелей показано в российской и украинской популяциях (228, 214 соответственно), 204 аллеля было обнаружено в популяции молдаван. В популяционных выборках гагаузов (195), марийцев (192) и поляков (184) показано менее 200 аллелей. Наименьшее количество аллелей было обнаружено в коми-ижемской популяции (163).

После применения коррекции Бонферрони ($p = 0.002$) не наблюдалось существенных отклонений от равновесия Харди–Вайнберга, за исключением четырех маркеров в трех популяциях. Маркер FGA показал отклонения в российской

популяции ($p = 0.00082 \pm 0.00002$). Маркер Penta E показал отклонения в общей выборке белорусов ($p = 0.00175 \pm 0.00004$). В подвыборке Восточного Полесья отклонение от равновесия было показано по локусу D1S1656 ($p = 0.00095 \pm 0.00003$). Наконец, D16S539 ($p = 0.00095 \pm 0.00003$) показал отклонения в выборке коми-ижемцев.

Анализ генетических расстояний для шести выборок белорусов не показал между ними статистически значимой генетической дистанции. Таким образом, для данной популяции нет необходимости учитывать территориальное происхождение индивида внутри страны при идентификационном анализе. Тем не менее в дальнейших расчетах мы приводим результаты как для тотальной выборки белорусов, так и для территориальных выборок.

Обратная ситуация показана для выборок коми. Анализ генетических расстояний по Нею обнаружил наличие значимой генетической дистанции между коми-зырянами и коми-ижемцами (0.0054, $p = 0.004$). Эти результаты дают основание для рассмотрения этих выборок отдельно друг от друга при формировании референсных популяционных данных при проведении генетической экспертизы.

Также значимые генетические расстояния по 21 анализируемому STR-маркеру отделяют популяции коми-зырян, коми-ижемцов, марийцев, молдаван друг от друга и от остальных анализируемых популяций. Напротив, между русскими, белорусами, поляками и украинцами значительная генетическая дистанция не показана (табл. 1).

Стоит отметить, что маркером с наибольшим числом аллелей является локус D12S391, для которого показано 18 аллельных вариантов в белорусской популяции и 15 аллелей в польской популяции. В остальных анализируемых популяциях наибольшее количество аллелей отмечено для локуса Penta E. Данные о частотах аллелей доступны по запросу у авторов.

Генетическая дифференциация популяций

Оценку генетической дифференциации исследованных выборок по совокупности всех STR-маркеров проводили с помощью анализа молекулярной дисперсии. При сравнении всего массива выборок различия между популяциями составили 0.35%. В случае объединения выборок до уровня этнических групп дифференциация на уровне этносов возросла до 0.39%. При сравнении двух выборок русских и белорусов различия полностью отсутствовали. Сравнение популяций русских с коми и марийцами продемонстрировало небольшой отрицательный уровень генетической дифференциации (0.03%), свидетельствующий о несколько большем генетическом сходстве индивидов из различных популяций по сравнению с индивидами из той же популяции. Дифференциация популяций русских, молдаван и гагаузов составила 0.13%.

При анализе частот аллелей для всех популяций из Республики Беларусь (всех выборок белорусов, русских, поляков и украинцев) доля различий между ними составила 0.1%. При сравнении выборок белорусов из разных историко-этнографических регионов республики уровень дифференциации между ними был 0.1%. При сравнении группы белорусов с русскими, поляками и украинцами различие достигло значения 0.8%. Эти результаты вполне ожидаемы в силу мультиаллельности STR-маркеров и их относительно высокой степени мутирования и рекомбинации. Также закономерно увеличение показателей генетической дифференциации при сравнении различных по происхождению этнических групп. Результаты дисперсионного анализа показывают отсутствие значительных генетических различий между русскими и белорусами, и славянскими популяциями в целом, а также очень низкий уровень различий между русскими, коми и марийцами по STR-генотипам. Это означает, что, несмотря на имеющуюся информацию о значительной антропологической дифференциации и существующие современные различия между исследованными этнотерриториальными группами, часть их генофонда по аутосомным STR-маркерам является очень мало дифференцированной.

Генетические взаимоотношения между популяциями

Визуализация матрицы генетических расстояний между популяционными выборками методом многомерного шкалирования (рис. 1) показывает, что исследованные выборки распределены в соответствии с их этнической близостью и географической локализацией. Наиболее генетически близкими друг другу являются выборки русских из России и Беларуси. Максимально удалены друг от друга наиболее территориально разобщенные выборки коми-ижемцев и поляков. Коми-зыряне и коми-ижемцы демонстрируют значительные различия по первому измерению, что вероятно связано с относительно сильным сдвигом частот аллелей в выборке коми-ижемцев в связи с их относительно недавним расселением на эту территорию и наличием эффекта основателя в этой субэтнической группе. В соответствии с географической локализацией близки друг другу и выборки белорусов из Восточного и Западного Полесья, Центральной Белоруссии, Поднепровья и Понеманья. Выбивается только выборка Поозерья, которая является самой северной, но при этом группируется с южными выборками из Полесья. Анализ подчерки-

Таблица 1. Генетические расстояния по Нею

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Русские (1)	*													
Гагаузы (2)	0.0010 (0.125)	*												
Коми-ижемцы (3)	0.0075 (0.000)	0.0103 (0.000)	*											
Коми-зыряне (4)	0.0020 (0.042)	0.0051 (0.001)	0.0054 (0.004)	*										
Марийцы (5)	0.0156 (0.000)	0.0186 (0.000)	0.0254 (0.000)	0.0148 (0.000)	*									
Молдаване (6)	0.0017 (0.002)	0.0008 (0.182)	0.0107 (0.000)	0.0032 (0.004)	0.0169 (0.000)	*								
BELvp (7)	–0.0001 (0.505)	0.0002 (0.378)	0.0095 (0.000)	0.0018 (0.118)	0.0168 (0.000)	–0.0006 (0.594)	*							
BELzp (8)	0.0016 (0.030)	0.0029 (0.025)	0.0097 (0.000)	0.0027 (0.044)	0.0189 (0.000)	0.0032 (0.001)	–0.0017 (0.860)	*						
BELp (9)	0.0003 (0.363)	0.0002 (0.406)	0.0073 (0.000)	0.0014 (0.211)	0.0154 (0.000)	0.0008 (0.190)	–0.0007 (0.637)	–0.0011 (0.804)	*					
BELpd (10)	0.0007 (0.149)	0.0017 (0.052)	0.0099 (0.000)	0.0042 (0.003)	0.0215 (0.000)	0.0015 (0.023)	–0.0005 (0.628)	–0.0001 (0.537)	0.0003 (0.404)	*				
BELpz (11)	0.0001 (0.453)	0.0020 (0.088)	0.0093 (0.001)	0.0031 (0.037)	0.0159 (0.000)	–0.0001 (0.464)	–0.0009 (0.698)	0.0008 (0.269)	0.0012 (0.218)	0.0002 (0.460)	*			
BELc (12)	–0.0004 (0.811)	0.0017 (0.040)	0.0098 (0.000)	0.0023 (0.033)	0.0165 (0.000)	0.0017 (0.004)	0.0004 (0.337)	0.0016 (0.057)	–0.0000 (0.502)	0.0002 (0.418)	–0.0002 (0.603)	*		
Поляки (13)	0.0006 (0.289)	0.0034 (0.028)	0.0075 (0.001)	0.0050 (0.011)	0.0205 (0.000)	0.0026 (0.035)	–0.0011 (0.627)	0.0002 (0.408)	–0.0010 (0.663)	0.0001 (0.462)	–0.0012 (0.748)	–0.0004 (0.583)	*	
Белорусы (14)	0.0001 (0.327)	0.0012 (0.065)	0.0092 (0.000)	0.0027 (0.006)	0.0173 (0.000)	0.0011 (0.007)	–0.0017 (0.982)	–0.0010 (0.928)	–0.0013 (0.950)	–0.0012 (0.993)	–0.0011 (0.953)	–0.0010 (0.998)	–0.0005 (0.659)	*
Украинцы (15)	–0.0006 (0.845)	0.0019 (0.030)	0.0077 (0.000)	0.0020 (0.044)	0.0188 (0.000)	0.0011 (0.030)	–0.0000 (0.439)	–0.0001 (0.509)	–0.0006 (0.668)	–0.0002 (0.543)	–0.0006 (0.727)	–0.0007 (0.921)	0.0002 (0.393)	–0.0005 (0.901)

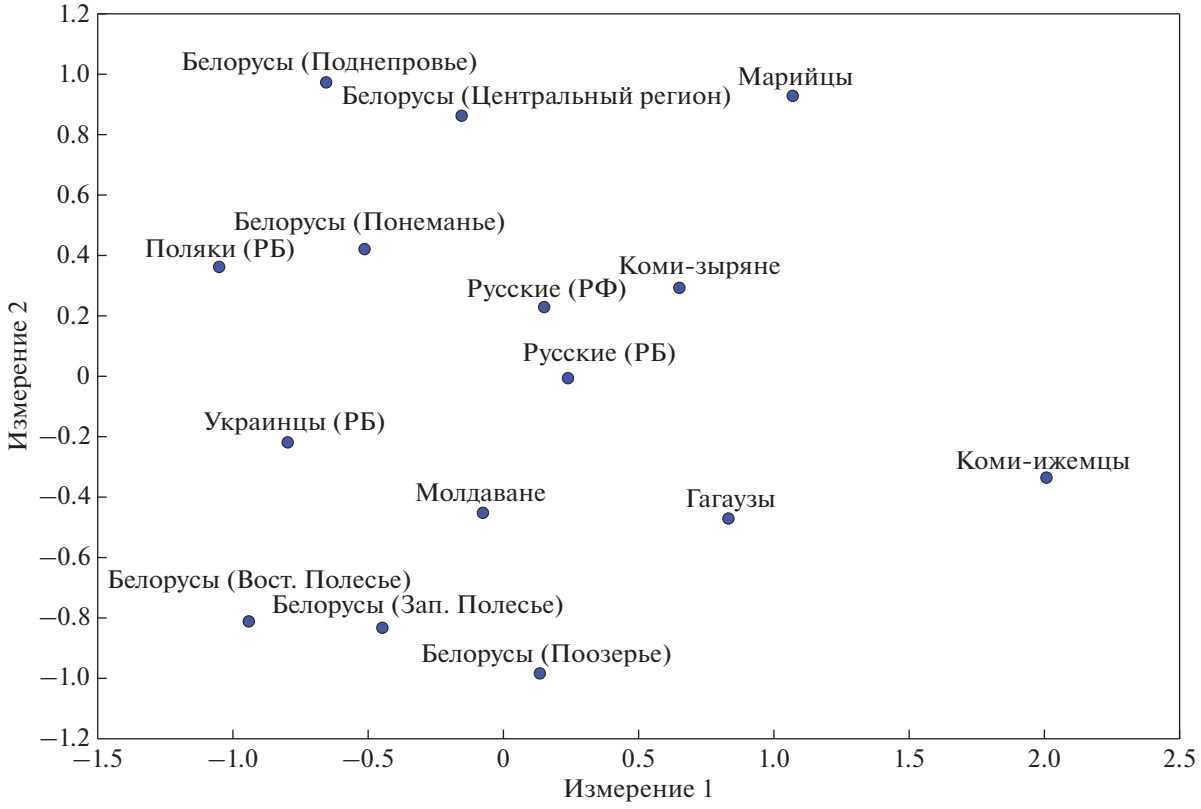

Рис. 1. Положение исследованных популяций в пространстве MDS.

вает генетическую близость между собой молдаван и гагаузов.

Идентификационный потенциал системы аутосомных STR-локусов

Для оценки возможности использования системы маркеров при ДНК-идентификации в судебной экспертизе, рассчитывали стандартные криминалистические параметры информативвности: вероятность случайного совпадения генотипов (MP, matching probability), вероятность дискриминации неродственных индивидов (PD, power of discrimination), исключающая способность (PE, power of exclusion), индекс отцовства (PI, paternity index). Показатели MP и PD используются при ДНК-идентификации личности, а значения PE и PI рассчитываются при определении отцовства.

Криминалистические параметры для каждого локуса представлены в дополнительных материалах. Наибольшая информативность расширенного набора маркеров для ДНК-идентификации показана для выборок коми (коми-зыряне, коми-ижемцы) (табл. 2). В популяции русских значение вероятности дискриминации неродственных индивидов (PD) возросло до 0.999999999994978, а показатель исключающей способности (PE) изменился от 0.99999999989 до 0.999999998679. Наиболее низкая идентификационная информативность набора из 21 STR показана для выборки поляков. Показатели информативности для идентификации и определения отцовства во всех популяциях на несколько порядков превышают значения, установленные нормативными актами, действующими на территории Российской Федерации. Данные о частотах аллелей 21 STR-локуса в популяциях, исследованных в настоящей работе, включены в разработанную авторами базу данных. Представленные частоты могут быть референсными (для соответствующей популяции или этнической группы) для вероятностно-статистической оценки результатов экспертного исследования при идентификации личности, установлении родства и др. Кроме того, данные могут использоваться для популяционно-генетических сравнительных исследований.

Таким образом, были установлены частоты аллелей и криминалистических параметров 21 аутосомного STR-локуса для 15 популяций восьми народов Восточной Европы. Показан значимый дискриминирующий потенциал используемого набора для ДНК-идентификации и его эффективность при определении родства. Оценка генетической дифференциации различных популяций выявила необходимость учета популяционной принадлежности при создании баз данных

Таблица 2. Идентификационная информативность совокупности 21 аутосомного STR-локуса в популяциях Восточной Европы

Популяция	PD	PE
Гагаузы	0.999999999954539	0.999999997608839
Коми-зыряне	1.0	0.999999999658466
Коми-ижемцы	1.0	0.999999999137624
Марийцы	0.999999999972582	0.999999998174319
Молдаване	0.999999999994137	0.999999999509155
Поляки	0.999999999789335	0.999999999974522
Русские	0.999999999994978	0.999999998679004
Украинцы	0.999999999996165	0.999999999893009
Белорусы	0.999999999997410	0.999999999263556
Восточное Полесье	0.999999999952167	0.999999999989964
Днепр	0.999999999952167	0.999999999133377
Центральная	0.999999999996428	0.999999999233081
Озерное плато	0.999999999969731	0.999999998645556
Неман	0.999999999961487	0.999999999765122
Западное Полесье	0.999999999980485	0.999999999810011

генотипов для сравнения референсных частот аллелей.

Работа выполнена в рамках Научно-технической программы Союзного государства “Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства” (“ДНК-идентификация”), Государственный контракт № 011-17 от 26.09.2017.

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

1. *Szczerkowska Z., Kapińska E., Wysocka J., Cybulska L.* Northern Polish population data and forensic usefulness of 15 autosomal STR loci // Forensic Sci. Int. 2004. V. 144. № 1. P. 69–71. https://doi.org/10.1016/j.forsciint.2004.02.020
2. *Jacewicz R., Jedrzejczyk M., Ludwikowska M., Berent J.* Population database on 15 autosomal STR loci in 1000 unrelated individuals from the Lodz region of Poland // Forensic Sci Int Genet. 2008. V. 2. № 1. P. e1–e3. https://doi.org/10.1016/j.fsigen.2007.08.002
3. *Малярчук Б.А., Wozniak M., Czarny J. и др.* Вариабельность 15 аутосомных микросателлитных локусов ДНК в русской популяции // Мол. биология. 2007. Т. 41. № 1. С. 3–7.
4. *Piatek J., Jacewicz R., Ossowski A. et al.* Population genetics of 15 autosomal STR loci in the population of Pomorze Zachodnie (NW Poland) // Forensic Sci. Int. Genet. 2008. V. 2. № 3. P. e41–e43. https://doi.org/10.1016/j.fsigen.2007.09.002
5. *Rebała K., Wysocka J., Kapińska E. et al.* Belarusian population genetic database for 15 autosomal STR loci // Forensic Sci. Int. 2007. V. 173. № 2–3. P. 235–237. https://doi.org/10.1016/j.forsciint.2007.02.002
6. *Pepiński W., Skawrońska M., Niemcunowicz-Janica A. et al.* Polymorphism of 11 autosomal STRS in a population sample of Belarussian minority residing in the region of Podlasie (Northeastern Poland) – an extension of the STR core set // Arch. Med. Sadowej Kryminol. 2008. V. 58. № 2–3. P. 65–68.
7. *Zhivotovsky L.A., Malyarchuk B.A., Derenko M.V. et al.* Developing STR databases on structured populations: The native South Siberian population versus the Russian population // Forensic Sci. Int. Genet. 2009. V. 3. № 4. P. e111–e116. https://doi.org/10.1016/j.fsigen.2008.08.001
8. *Parys-Proszek A., Kupiec T., Wolańska-Nowak P., Branicki W.* Genetic variation of 15 autosomal STR loci in a population sample from Poland // Leg. Med. (Tokyo). 2010. V. 12. № 5. P. 246–248. https://doi.org/10.1016/j.legalmed.2010.05.002
9. *Stepanov V.A., Melnikov A.V., Lash-Zavada A.Y. et al.* Genetic variability of 15 autosomal STR loci in Russian populations // Leg. Med. (Tokyo). 2010. V. 12. № 5. P. 256–258. https://doi.org/10.1016/j.legalmed.2010.05.006
10. *Степанов В.А., Балановский О.П., Мельников А.В. и др.* Характеристика популяций Российской Фе-

дерации по панели пятнадцати локусов, используемых для ДНК-идентификации и в судебно-медицинской экспертизе // Acta Naturae. 2011. Т. 3. № 2. С. 59–71.

11. *Soltyszewski I., Pepinski W., Wolanska-Nowak P. et al.* Polish population data on 15 autosomal STRs of AmpFlSTR NGM PCR kit // Forensic Sci. Int. Genet. 2014. V. 9. P. 142–149. https://doi.org/10.1016/j.fsigen.2013.12.002
12. *Цыбовский И.С., Вермейчик В.М., Котова С.А. и др.* Создание судебной референтной базы данных по 18 аутосомным STR для ДНК-идентификации в Республике Беларусь // Генетика. 2017. Т. 53. № 2. С. 249–258.
13. *Benvisto A., Messina F., Finocchio A. et al.* A genetic portrait of the South-Eastern Carpathians based on autosomal short tandem repeats loci used in forensics // Am. J. Hum. Biol. 2018. V. 30. № 5. P. e23139. https://doi.org/10.1002/ajhb.23139
14. *Rębała K., Veselinović I., Siváková D. et al.* Northern Slavs from Serbia do not show a founder effect at autosomal and Y-chromosomal STRs and retain their paternal genetic heritage // Forensic Sci. Int. Genet. 2009. V. 3. № 4. P. e133–e136. https://doi.org/10.1016/j.fsigen.2008.11.001
15. *Excoffier L., Laval G., Schneider S.* Arlequin ver. 3.0: An integrated software package for population genetics data analysis // Evol. Bioinformatics Online. 2005. V. 1. P. 47–50.
16. *Desmarais D., Zhong Y., Chakraborty R. et al.* Development of a highly polymorphicSTR marker for identity testing purposes at the humanandrogen receptor gene (HUMARA) // J. Forensic Sci. 1998. V. 43. P. 1046–1049.

Genetic Diversity of 21 Autosomic STR-Markers of the Codis System in Populations of Eastern Europe

V. N. Kharkov[a, *], S. A. Kotova[b], N. A. Kolesnikov[a], E. A. Spivak[b], K. V. Vagaitseva[a], T. V. Zabavskaya[b], V. I. Rybakova[b], A. S. Parfenova[b], A. N. Verchuk[b], N. K. Yankovsky[c, d], and V. A. Stepanov[a]

[a]Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, 634050 Russia

[b]Scientific Practical Center of the State Forensic Expertise Committee of the Republic of Belarus, Minsk, 220114 Belarus

[c]Vavilov Institute of General Genetics, Moscow, 119991 Russia

[d]Lomonosov Moscow State University, Moscow, 119991 Russia

**e-mail: vladimir.kharkov@medgenetics.ru*

For the first time, genetic characteristics of 15 local populations of 8 ethnic groups of Eastern Europe were obtained using an expanded panel that included twenty-one autosomal STR markers used in forensic practice for DNA identification and kinship establishment. The assessment of polymorphism of markers in the Russian, Belarusian and Moldovan populations was carried out, the levels of intra- and interpopulation genetic differentiation of the studied population groups were characterized. The parameters of the informational significance of the system of autosomal microsatellites for expert DNA identification and establishment of kinship were estimated. A database of allele frequencies of autosomal STR loci has been developed, which can be used to carry out probabilistic and statistical calculations when assessing the level of reliability of an expert study in the countries of the Union State. It was shawn thet population genetic structure should been taking into account in forensic DNA analysis.

Keywords: microsatellite locus, allele frequencies, population, genetic differentiation, DNA identification, forensic parameters.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1403–1414

УДК 575.174:599.9

РЕКОНСТРУКЦИЯ ПРОИСХОЖДЕНИЯ ГЫДАНСКИХ НЕНЦЕВ НА ОСНОВЕ ГЕНЕТИЧЕСКОГО АНАЛИЗА ИХ РОДОВОЙ СТРУКТУРЫ С ПОМОЩЬЮ НОВОГО НАБОРА YSTR-МАРКЕРОВ

© 2021 г. В. Н. Харьков[1, *], Л. В. Валихова[1], Е. Л. Яковлева[1], В. Н. Сереброва[1], Н. А. Колесников[1], Т. И. Петелина[2], И. Ю. Хитринская[1], В. А. Степанов[1]

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

[2]*Тюменский кардиологический научный центр, Томский национальный исследовательский медицинский центр Российской академии наук, Тюмень, 625026 Россия*

**e-mail: vladimir.kharkov@medgenetics.ru*

Поступила в редакцию 20.05.2021 г.
После доработки 17.06.2021 г.
Принята к публикации 28.06.2021 г.

Исследована генетическая структура ненецких родов по маркерам Y-хромосомы с использованием разработанного набора YSTR для определения популяционной принадлежности неизвестного индивида. Результаты анализа частот гаплогрупп и YSTR-гаплотипов свидетельствуют, что ненецкие рода являются родственными объединениями, в большинстве случаев имеющими одного предка по мужской линии. Показано, что Y-хромосомный генофонд ненцев структурирован по родовому принципу. Для подавляющего большинства образцов выявлена тесная генетическая близость представителей одного рода и относительно недавний эффект основателя по мужской линии. Структура гаплотипов гаплогрупп Y-хромосомы в составе ненецкого генофонда демонстрирует генетическую близость ненцев с хантами и энцами. Выявлены значительные различия между двумя ненецкими фратриями. Оценка возраста происхождения отдельных сублиний Y-хромосомы и родовых кластеров гаплотипов полностью согласуется с данными этнографии о формировании родов и фратрий в составе гыданских ненцев. Результаты сравнительного анализа спектра гаплотипов убедительно свидетельствуют о специфичности сублиний Y-хромосомы и кластеров гаплотипов не только на уровне этносов и популяций, но и на уровне фратрий, родов и отдельных фамильных групп. За единичными исключениями каждая фратрия и род имеют свой специфичный кластер гаплотипов, равноудаленных друг от друга.

Ключевые слова: Y-хромосома, генофонд, популяция, генетическое разнообразие, ненцы, род, ДНК-идентификация.

DOI: 10.31857/S0016675821120067

Ненцы – народ, проживающий в европейской части России и севере Западной Сибири, являются одной из многочисленных групп коренных малочисленных народов. Они расселены на огромной территории, простирающейся с запада на восток на несколько тысяч километров, и делятся на европейских и азиатских (сибирских). Европейские ненцы проживают в Ненецком автономном округе Архангельской области, сибирские – в Ямало-Ненецком автономном округе Тюменской области и в Долгано-Ненецком Таймырском муниципальном районе Красноярского края. Небольшие группы ненцев проживают в Ханты-Мансийском автономном округе, в Мурманской области, а также в Республике Коми. Наряду с нганасанским, энецким и селькупским язык ненцев входит в самодийскую группу уральской языковой семьи. По антропологическим признакам ненцы относятся к монголоидам.

По данным Всероссийской переписи 2010 г. в составе населения Ямало-Ненецкого автономного округа их насчитывалось 29772 человека (5.7% от общей численности населения) [1]. Самая восточная часть ареала ненцев расположена на территории между Обью и Енисеем. Ненцев, проживающих в Тазовском районе Ямало-Ненецкого автономного округа, можно объединить общим названием “гыданские ненцы”, так как большая часть района расположена на территории Гыданского полуострова, они относятся к группе тундровых [2]. В Тазовском районе находятся национальные поселки – Гыда, Антипаюта, Находка и районный центр Тазовский.

Как и для большинства южносибирских и западносибирских народов, отличительной чертой ненцев является наличие родов, где счет родства ведется по мужской линии. Родовая организация ненцев достаточно хорошо изучена антропологами и этнографами [1]. Браки у ненцев всегда были строго экзогамными. У гыданских, ямальских и тазовских ненцев рода объединялись в две экзогамные фратрии, одна из которых – Харючи ("настоящие ненцы"). Она состояла из родов самодийского происхождения, вторая (Вануйто, "иноплеменники") – из родов, восходящих к аборигенному генетическому субстрату, а также энецкого и хантыйского происхождения [3]. Иноэтничные группы родов включены ненцами в свою дуально-фратриальную организацию и их представители заключают браки в соответствии с их экзогамными нормами. Со второй половины XVIII в. ненцы стали наращивать поголовье домашних оленей, что привело к постепенному увеличению численности населения и дроблению больших патриархальных родов на малые рода-патронимии [4]. Гыданская группа ненцев является популяцией, наиболее сохранившей традиционную социальную структуру, так как была наименее подвержена изменениям в социально-экономической и культурной сферах.

В XX в. в состав гыданских ненцев влились отдельные представители лесных и европейских ненцев, а также несколько русских и татар. На 2001 год на территории Тазовского района проживали представители 43 ненецких родов и восьми оненечившихся фамилий. На Гыдане проживают представители обеих фратрий. К родам фратрии Харючи относятся Адер, Евай, Лапсуй, Ненянг, Няруй, Окотэтто, Сусой, Сэротэтто, Сюгней, Тогой, Тэсида, Хабдю, Харючи, Хороля, Худи, Хэно, Ядне, Яндо, Яптунай. Во фратрию Вануйто входят рода Вануйто, Вэнго, Ламдо, Пуйко, Сабы, Яр, Яптик, Яунгад. Основу гыданского этнического массива составили ненцы родов Харючи, Яр, Сюгней, Айваседа, Адер.

В формировании восточных ненцев приняли участие также энецкие и хантыйские по происхождению рода, которые вошли в состав фратрии Вануйто. К энецким родам принадлежат Марьик, Аседа, Паровых, Тёр, Оковай. К хантыйским относятся Тибичи, Вэхо, Неркыхы, Няданги, Пурунгуй, Салиндер [5].

Родовую принадлежность у ненцев в большинстве случаев возможно установить по фамилиям, так как практически все современные ненецкие фамилии повторяют название рода либо полностью в исходной форме, либо с небольшими изменениями. Таким образом, подробные данные об этнографии и родовой структуре гыданских ненцев дают богатый материал для исследования генетической структуры родов и фратрий.

Большое число информативных SNP и STR-маркеров Y-хромосомы, используемых в настоящее время для изучения популяционно-генетической структуры различных этносов, позволяют проверить, является ли тот или иной род кровнородственным объединением, имеющим одного родоначальника по мужской линии, или осознание такой общности лишь условное и род является общностью людей, проживающих на одной территории, но не связанных генетическим родством по отцовским генеалогическим линиям. Ранее по маркерам Y-хромосомы уже была охарактеризована генетическая структура родов у казахов [6, 7], башкир [8], хакасов и шорцев [9].

В ходе выполнения проекта по разработке различных тест-систем для ДНК-идентификации и определения популяционной принадлежности неизвестного индивида по образцу его ДНК было показано, что гаплотипы, построенные с помощью генотипирования специально отобранных по уровню их информативности и темпу мутирования STR-маркеров Y-хромосомы, обладают значительной этнической и территориальной специфичностью. Была установлена очень высокая межпопуляционная дифференциация полученных YSTR-гаплотипов [10], что позволило создать эффективно работающий набор YSTR-маркеров для целей ДНК-идентификации.

Цель исследования – охарактеризовать генетическую структуру родов гыданских ненцев с использованием маркеров Y-хромосомы, определить их генетическое родство по мужской линии и установить межэтническое генетическое родство отдельных родов, входящих в состав различных фратрий. В качестве инструмента исследования мы использовали разработанный нами набор YSTR для определения популяционной принадлежности индивидов по образцу их ДНК, решая параллельно задачу тестирования возможностей набора для уточнения степени точности и детализации привязки индивидов не только к этносам и популяциям, но и к конкретному роду и фамилии.

МАТЕРИАЛЫ И МЕТОДЫ

Материал исследования составили образцы ДНК мужчин из различных локальных популяций гыданских ненцев ($N = 322$). Выборку гыданских ненцев представляет население поселка Тазовский, села Антипаюта, села Гыда, Тазовского района Ямало-Ненецкого автономного округа Российской Федерации. Материал получен в ходе совместных научно-практических медицинских экспедиций в 2019 г. и депонирован в биоресурсную коллекцию "Биобанк населения Северной Евразии". Забор первичного биологического материала (венозной крови) у доноров проводили с соблюдением процедуры письменного информированного согласия на проведение исследования.

На каждого донора составлялась анкета с его родословной, указанием этнической принадлежности и мест рождения предков. В исследование включены только образцы ДНК доноров-мужчин, по результатам анкетирования отрицавших факт метисации по отцовской линии с представителями других этносов минимум в трех поколениях. Индивида относили к данной этнической группе на основании его собственной этнической идентификации, его родителей, места рождения и принадлежности к определенному роду.

Среди всех мужчин, включенных в настоящее исследование, оказались представлены следующие рода (в алфавитном порядке): Адер, Вануйто, Вэлло, Вэнго, Евай, Лапсуй, Лар, Марьик, Ненянг, Няч, Оковай, Окотэтто, Пуйко, Пурунгуй, Салиндер, Сусой, Тёр, Тибичи, Тогой, Тэсида, Худи, Ядне, Яндо, Яптик, Яптунай, Яр, Яунгат.

Для изучения состава и структуры гаплогрупп Y-хромосомы в исследование были включены две системы генетических маркеров: диаллельные локусы, представленные SNP, и полиаллельные высоковариабельные микросателлиты (YSTR). С помощью 103 SNP-маркеров определяли принадлежность образцов к той или иной гаплогруппе. Классификация гаплогрупп дана в соответствии с предложенной Консорциумом по исследованию Y-хромосомы [11], с последующими изменениями [12]. Анализ STR-гаплотипов внутри гаплогрупп проводили с использованием 45 микросателлитных маркеров нерекомбинирующей части Y-хромосомы (YSTR) (DYS19, 385a, 385b, 388, 389I, 389II, 390, 391, 392, 393, 426, 434, 435, 436, 437, 438, 439, 442, 444, 445, 448, 449, 456, 458, 460, 461, 481, 504, 505, 518, 525, 531, 533, 537, 552, 570, 576, 635, 643, YCAIIa, YCAIIb, GATA H4.1, Y-GATA-A10, GGAAT1B07). Генотипирование SNP-маркеров проводили с помощью ПЦР и последующего анализа фрагментов ДНК с применением ПДРФ как описано ранее [10]. STR-маркеры генотипировали с помощью капиллярного электрофореза на генетическом анализаторе ABI Prism 3730 и программного обеспечения GeneMapper как описано ранее [10]. Экспериментальные исследования проведены на базе Центра коллективного пользования научно-исследовательским оборудованием “Медицинская геномика” Томского НИМЦ РАН). Построение медианных сетей гаплотипов Y-хромосомы проводили с использованием программы Network v 10.0.0.0 (Fluxus Technology Ltd.; www.fluxus-engineering.com) по методу медианных сетей Бандельта [13]. Оценку возраста генерации наблюдаемого разнообразия гаплотипов в гаплогруппах проводили методом ASD [14], на основании средних квадратичных отличий в числе повторов между всеми маркерами. Длина поколения принималась равной 30 годам, темп мутирования 0.0021 на локус на поколение. При расчете возраста родов по отдельным гаплогруппам исключались единичные образцы, которые значительно выбивались из общего кластера гаплотипов. Расчеты проводили для родовых групп численностью не менее пяти образцов.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В результате анализа частоты встречаемости использованных SNP-маркеров в исследуемой выборке гыданских ненцев было выявлено шесть гаплогрупп Y-хромосомы. За исключением единичных образцов, практически все исследованные образцы относятся к Y-хромосомной макрогруппе N. Все шесть оставшихся образцов относятся к различным западно-евразийским по происхождению гаплогруппам, что, возможно, свидетельствует о единичных фактах недавней метисации или включения в состав родов отдельных представителей пришлого населения (табл. 1) (рис. 1).

Фратрия Харючи

В формировании ненцев приняли участие в основном два компонента: южносибирский самодийский и местный аборигенный. В результате их взаимодействия, по мнению Б. Долгих [15], сформировались две фратрии ненцев. Одна восходит к самодийцам – Харючи, а другая к аборигенам – Вануйта (или Вануйто).

Практически все прогенотипированные образцы, относящиеся к фратрии Харючи, относятся (за единичными исключениями) к гаплогруппе N1a2b1b1-B170,B171,Z35101(xB172). Это представители родов Адер, Евай (Явай), Лапсуй, Ненянг, Окотэтто, Сусой, Тогой, Тэсида, Худи, Ядне, Ядно и Яптунай. Все представители рода Вэлло, который относится к лесным ненцам, также принадлежат к этой гаплогруппе, хотя он не входит в состав фратрии Харючи (рис. 1).

Медианная сеть YSTR-гаплотипов демонстрирует ярко выраженную звездообразную филогению гаплогруппы N1a2b1b1-B170, свидетельствующую о сильном эффекте основателя и общем предке для всех родов по мужской линии. Исключением являются все представители рода Евай, которые формируют отдельный кластер гаплотипов. Этому возможно два объяснения: во-первых, род Евай может иметь более древнего общего предка с остальными родами фратрии Харючи и являться потомками параллельно существовавшей с ней группы мужчин; во-вторых, возможно, что различия по числу тандемных повторов сразу в нескольких маркерах являются результатом нескольких одновременных мутаций при гаметогенезе конкретного индивида, который и был родоначальником этого рода. Существование таких выбивающихся из общей сети гаплотипов отдельных субветвей в различных гаплогруппах было

Таблица 1. Частоты встречаемости гаплогрупп Y-хромосомы у гыданских ненцев

Род	N1a2b1b1-B170,B171, Z35101 (xB172)	N1a2b1b1b-B172, Z35108	N1a2b 2a-E-Y3185	N1a2b1-B478, Z35080 (xVL65)	N1a1a1a1 a2a1c-Y13850	N1a1a1 a1a2-Z1936	Другие	Суммарное число
Адер	3 (100%)	–	–	–	–	–	–	3
Вануйто	–	46 (100%)	–	–	–	–	–	46
Вэлло	6 (100%)	–	–	–	–	–	–	6
Вэнго	1 (14.29%)	–	4 (57.14%)	–	–	–	2 (28.57%)	7
Евай (Явай)	9 (100%)	–	–	–	–	–	–	9
Лапсуй	37 (94.87%)	–	–	–	2 (5.13%)	–	–	39
Лар	–	–	–	–	2 (100%)	–	–	2
Марьик	–	–	–	3 (100%)	–	–	–	3
Ненянг	4 (100%)	–	–	–	–	–	–	4
Няч	–	–	–	6 (100%)	–	–	–	6
Оковай	–	–	–	1 (100%)	–	–	–	1
Окотэтто	2 (66.67%)	1 (33.33%)	–	–	–	–	–	3
Пуйко	–	2 (100%)	–	–	–	–	–	2
Пурунгуй	–	2 (66.67%)	–	–	–	1 (33.33%)	–	3
Салиндер	–	–	–	–	34 (89.47%)	–	4 (10.53%)	38
Сусой	7 (100%)	–	–	–	–	–	–	7
Тёр	–	–	–	1 (100%)	–	–	–	1
Тибичи	2 (25%)	–	–	–	6 (75%)	–	–	8
Тогой	4 (66.67%)	–	–	–	2 (33.33%)	–	–	6
Тэсида	11 (100%)	–	–	–	–	–	–	11
Худи	3 (100%)	–	–	–	–	–	–	3
Ядне	35 (87.5%)	2 (5%)	–	–	3 (7.5%)	–	–	40
Яндо	14 (100%)	–	–	–	–	–	–	14

Таблица 1. Окончание

Род	N1a2b1b1-B170,B171, Z35101 (xB172)	N1a2b1b1b-B172, Z35108	N1a2b 2a-E-Y3185	N1a2b1-B478, Z35080 (xVL65)	N1a1a1a1 a2a1c-Y13850	N1a1a1 a1a2-Z1936	Другие	Суммарное число
Яптик	1 (25%)	–	–	3 (75%)	–	–	–	4
Яптунай	20 (100%)	–	–	–	–	–	–	20
Яр	7 (20%)	1 (2.86%)	11 (31.43%)	–	2 (5.71%)	14 (40%)	–	35
Яунгат	–	1 (100%)	–	–	–	–	–	1
	166	55	15	14	51	15	6	322

показано и ранее, при анализе большого массива данных по различным популяциям и гаплогруппам [16]. С учетом небольшой дисперсии числа повторов в этом кластере гаплотипов у рода Евай наиболее вероятен именно второй вариант. Согласно данным этнографов род Евай выделился из рода Адер в первой половине XX в. [2, 5]. Ветвь гаплотипов рода Евай на медианной сети наиболее близка именно к кластеру, состоящему из представителей родов Адер, Сусой, Вэнго и Ненянг.

Модальный гаплотип этой гаплогруппы обнаружен у представителей семи ненецких родов. При этом все модальные гаплотипы всех крупных родов отличаются от модального гаплотипа гаплогруппы N1a2b1b1-B170 как минимум на один мутационный шаг. Только шесть гаплотипов кроме общего модального повторяются более чем у одного рода. Все они отличаются от модального на один мутационный шаг, и в силу его большой доли от общего числа гаплотипов могут быть результатом повторных параллельных или обратных мутаций отдельных STR. За единичными исключениями каждый род характеризуется собственным спектром гаплотипов, в среднем равноудаленных от модального. Каждый отдельный род также имеет ярко выраженную генетическую близость принадлежащих к нему индивидов и звездообразную

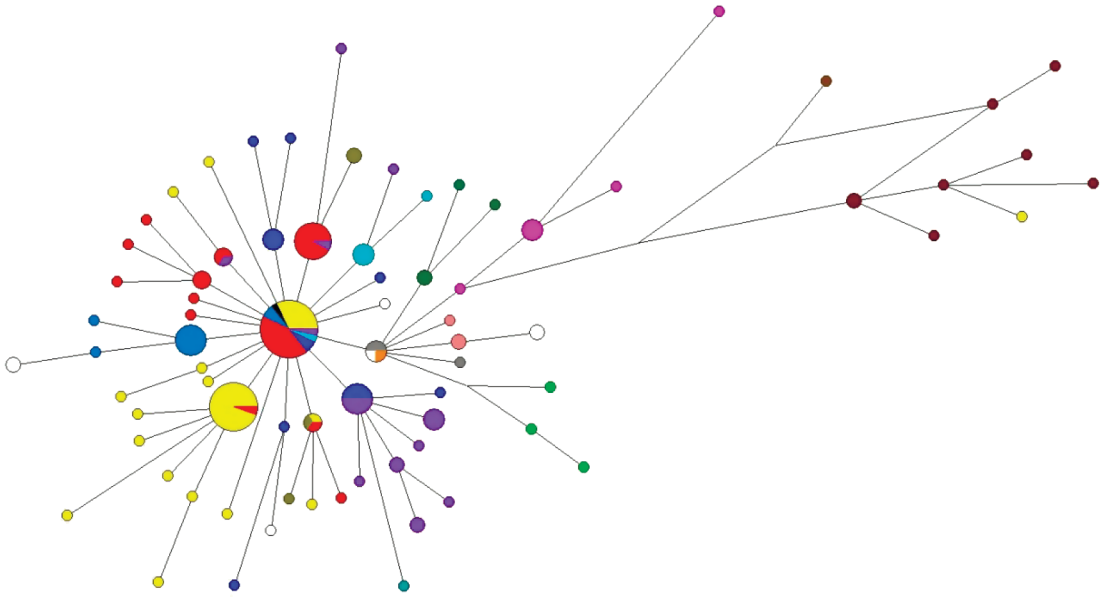

Рис. 1. Медианная сеть гаплогруппы N1a2b1b1-B170,B171,Z35101 (xB172). Желтым цветом обозначен род Лапсуй, красным – Ядне, светло-голубым – Яр, темно-голубым – Яптик, синим – Тэсида, темно-синим – Яндо, фиолетовым – Яптунай, светло-зеленым – Адер, темно-зеленым – Ненянг, болотным – Тогой, белым – Окотэтто, серым – Худи, светло-розовым – Вэлло, розовым – Сусой, коричневым – Евай, оранжевым – Вэнго.

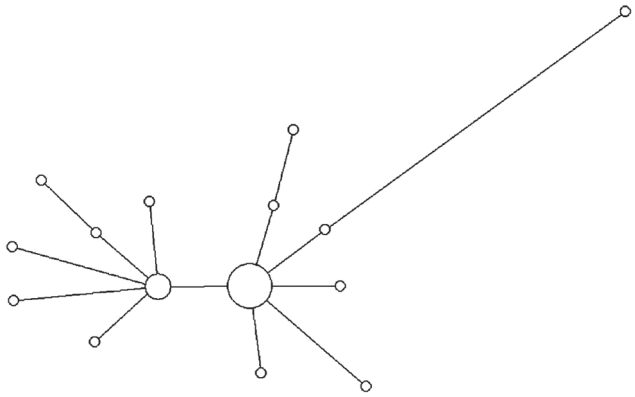

Рис. 2. Медианная сеть гаплогруппы N1a2b1b1-B170. Род Лапсуй.

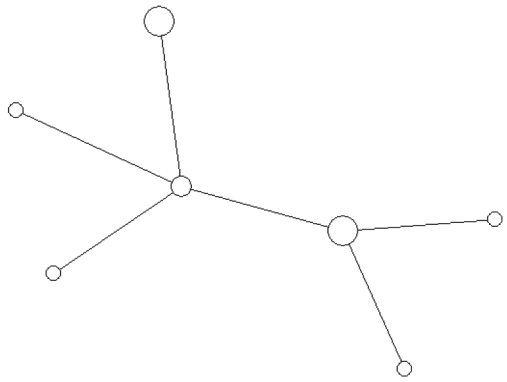

Рис. 3. Медианная сеть гаплогруппы N1a2b1b1-B170. Род Яндо.

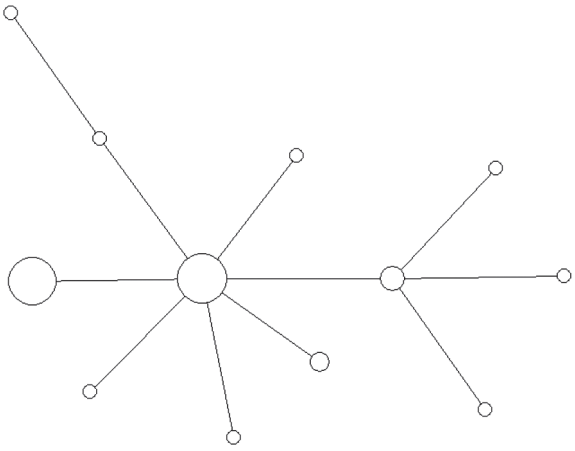

Рис. 4. Медианная сеть гаплогруппы N1a2b1b1-B170. Род Ядне.

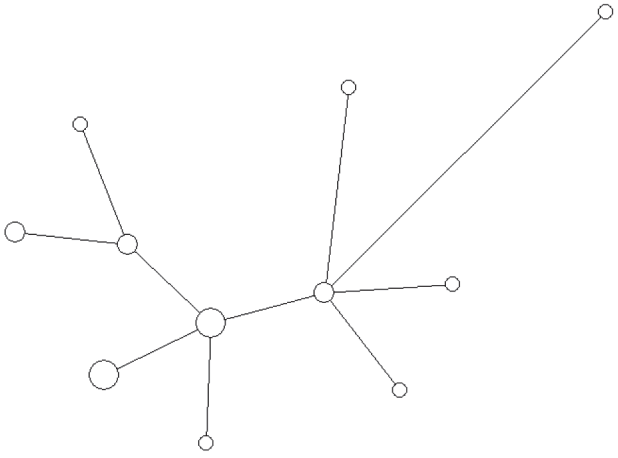

Рис. 5. Медианная сеть гаплогруппы N1a2b1b1-B170. Род Яптунай.

филогению гаплотипов на медианной сети, что подтверждает биологическое родство мужчин данного рода по отцовской линии (рис. 2–5).

Таким образом, подавляющее число мужчин родов фратрии Харючи принадлежит к одной гаплогруппе и является потомками одного общего предка, что подтверждает данные этнографии о родстве различных родов и выделении их из общих предковых групп. Рода этой фратрии являются патрилинейно родственными друг другу, и в ходе увеличения численности и расселения по новым территориям, за единичными исключениями, не ассимилировали и не включали в свой состав мужчин из других родов. Полученные результаты полностью соответствуют данным о доминировании самодийского исходного компонента во фратрии Харючи и данным этнографии о генеалогической близости всех входящих в нее родов.

Общий возраст для всего массива гаплотипов основной гаплогруппы N1a2b1b1-B170 составил 724 года (SD = 167 лет), для рода Лапсуй – 395 лет (SD = 202 года), для рода Вэлло – 805 лет (SD = = 313 лет), для рода Евай – 814 лет (SD = 266 лет).

Для многих родов, входящих во фратрию Харючи, по различным документам (переписям населения, спискам крестьян и т.д.) известно примерное время их отделения от рода Харючи. Для рода Сусой время генерации гаплотипов составляет 104 года (SD = 73 года). Род Сусой является подразделением рода Окотэтто и по данным этнографии он обособился в первой половине XX в. [5]. Возраст рода Ядне – 285 лет (SD = 146 лет). Родоначальником Ядне является Яур Удолин, который упоминается в ревизских документах в составе рода Харючи начиная с 1763 г. [2, 5]. Возраст рода Яндо – 366 лет (SD = 189 лет). Этот род впервые был отмечен переписью 1897 г. До этого времени сведений о нем не имеется. Однако представители этого рода утверждают, что Яндо является одним из самых древних ненецких родов. Возраст рода Яптунай – 476 лет (SD = 207 лет). Название

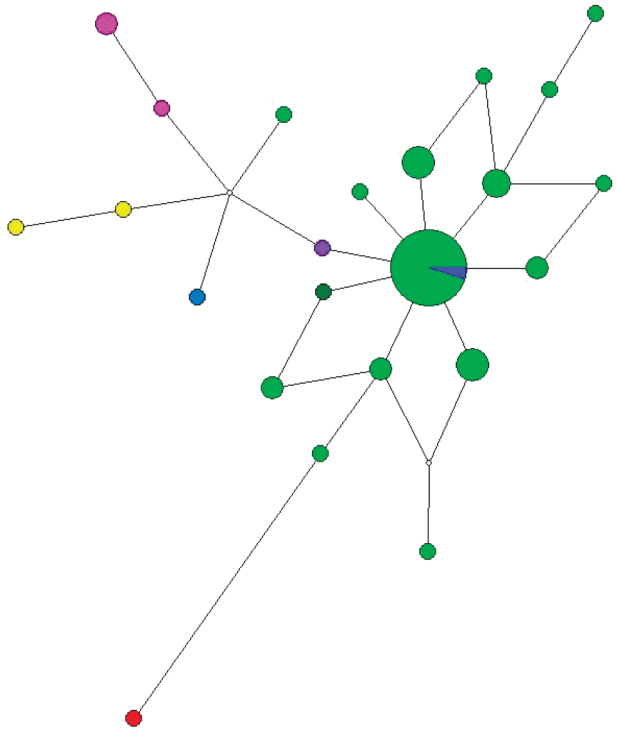

Рис. 6. Медианная сеть гаплогруппы N1a2b1b1b-B172. Зеленым цветом обозначен род Вануйто, фиолетовым – Яр, розовым – Пуйко, синим – Вэлло, желтым – Пурунгуй, красным – Вэнго, темно-зеленым – Яунгат.

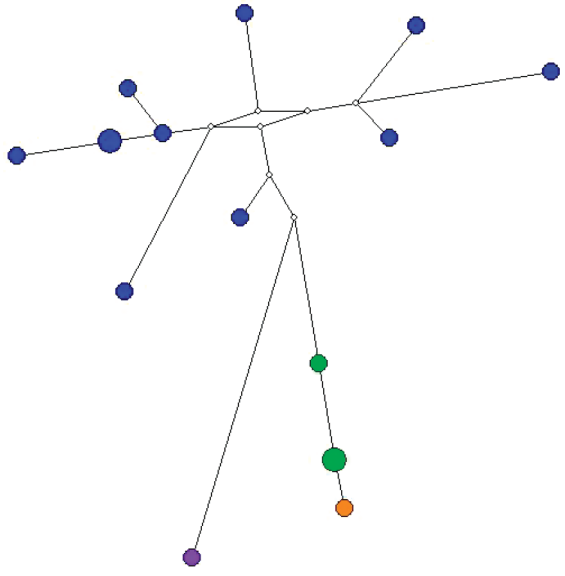

Рис. 7. Медианная сеть гаплогруппы N1a2b2a-E-Y3185. Синим цветом отмечен род Яр, зеленым – Вэнго, оранжевым – Тайбери, фиолетовым – Тибичи.

этого рода впервые упоминается в списках крестьян в конце 30-х гг. XIX в. Возраст рода Тэсида – 100 лет (SD = 74 года). В официальных документах этот род впервые отмечен во второй половине XIX в. Таким образом, результаты оценки возраста генерации гаплотипического разнообразия для перечисленных родов хорошо совпадают с данными этнографии и записей о их обособлении.

Общий возраст гаплогруппы N1a2b1b1-B170 близок по времени к приходу основной миграционной волны самодийцев на территорию северной части Западной Сибири.

Фратрия Вануйто

Во фратрии Вануйто наблюдается совершенно другой состав гаплогрупп и значительная их специфика у разных родов по сравнению с Харючи. В то время как все до одного представителя родов Вануйто ($N = 46$), Пуйко ($N = 2$) и Яунгат ($N = 1$) принадлежат к гаплогруппе N1a2b1b1b-B172, Z35108 (рис. 6), большинство членов рода Вэнго ($N = 4$) относятся к другой линии – N1a2b2a-E-Y3185 (рис. 7), а рода Яптик – к гаплогруппе N1a2b1-B478-Z35080xVL65 ($N = 3$) (рис. 8) и N1a2b1b1-B170 ($N = 1$) (рис. 1).

Досамодийский аборигенный элемент в составе ненцев представлен главным образом в рамках фратрии Вануйто. Ее название Б. Долгих [15] связывал со словом ванг – яма и предполагал, что оно распространялось на “ямных жителей” – жителей землянок, т.е. на аборигенов, вошедших затем в состав фратрии, дав ей свое имя. Сначала

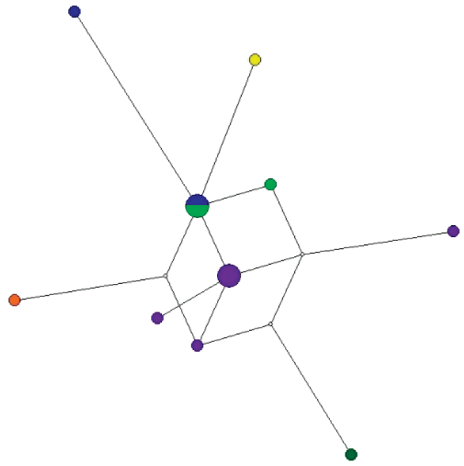

Рис. 8. Медианная сеть гаплогруппы N1a2b1-B478,Z35080(xVL65). Фиолетовым цветом обозначен род Няч, синим – Яптик, зеленым – Марьик, оранжевым – Тёр, желтым – Оковай.

название закрепилось за аборигенами, позже став именем фратрии, восходящей по происхождению к аборигенному пласту. Принадлежность большинства представителей части родов этой фратрии к гаплогруппе N1a2b1b1b-B172, а не к параллельной ей N1a2b1b1-B170 хорошо согласуется с этим предположением. При этом наличие других гаплогрупп свидетельствует о большем генетическом разнообразии фратрии Вануйто по сравнению с Харючи и возможно большей гетерогенности аборигенного компонента, вошедшего в состав современных гыданских ненцев.

Для основного рода Вануйто возраст составил 308 лет (SD = 94 года). Общий возраст для всего массива гаплотипов основной гаплогруппы N1a2b1b1b-B172,Z35108 составил 811 лет (SD = = 211 лет), что практически совпадает с возрастом N1a2b1b1-B170 и вероятно свидетельствует о начале демографического роста ненцев после ассимиляции пришлыми самодийцами аборигенного населения Севера. При этом, судя по всему, не происходило вытеснения коренных мужчин и их вклад в генофонд современных гыданских ненцев весьма значителен.

Род Яр. К фратрии Вануйто этнографы также относят род Яр – единственный род тазовских ненцев, из числа самых ранних переселенцев в низовья Таза. Сведений о его происхождении по данным этнографии нет. Из всех исследованных в рамках настоящей работы родов Яр является самым гетерогенным по составу. Представители этого рода относятся к пяти гаплогруппам Y-хромосомы: N1a1a1a1a2-Z1936 (N = 14) (рис. 10), N1a2b2a-E-Y3185 (N = 11) (рис. 7), N1a2b1b1-B170 (N = 7) (рис. 1), N1a1a1a1a2a1c-Y13850 (N = 2) (рис. 9) и N1a2b1b1b-B172 (N = 1) (рис. 6). Даже если исключить единичные образцы двух последних гаплогрупп, то этот род разделяется на три различных по происхождению генетических компонента. Внутри каждой гаплогруппы наблюдается значительное разнообразие гаплотипов и характерная для всех других родов звездообразная филогения медианной сети, что свидетельствует об отсутствии недавней метисации с другими родами или о включении в состав отдельных пришлых мальчиков или мужчин среди представителей этих кластеров гаплотипов. Возможно, что такая гетерогенность объясняется одинаковым наименованием разных по происхождению родов по топонимам, приписыванию общего названия для родов, проживающих на одной территории, или ассимиляции и включении в свой состав мужчин из местного населения при расселении на тазовской территории исходного рода Яр. Для гаплогруппы N1a1a1a1a2-Z1936 возраст составил 551 год (SD = 196 лет). Для гаплогруппы N1a2b1b1-B170 возраст составил 419 лет (SD = 283 года).

Хантыйские рода. В начале XX в. стали селиться на Тазу и представители хантыйских по происхождению родов. Они сформировались в XVIII–XIX вв. в низовьях Оби в результате развития ненецкого крупнотабунного оленеводства и вовлечения в него части северных хантов. К этим родам относятся Лар, Няданги, Неркыхы, Пурунгуй, Салиндер и Тибичи [17].

Мужчины, представляющие самый многочисленный род Салиндер (N = 38), полностью принадлежат к совершенно другой гаплогруппе N1a1a1a1a2a1c-Y13850. К этой же гаплогруппе относятся все представители рода Лар (N = 2) (рис. 9). Род Тибичи разделяется на N1a1a1a1a2a1c-Y13850 (N = 6) и N1a2b1b1-B170 (N = 2). Лишь в роде Пу-

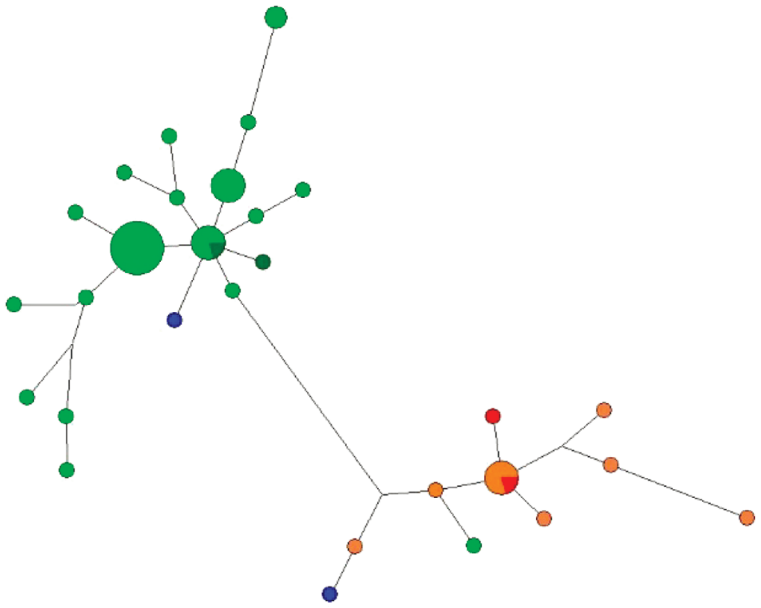

Рис. 9. Медианная сеть гаплогруппы N1a1a1a1a2a1c-Y13850. Оранжевым цветом обозначен род Тибичи, светло-зеленым – Салиндер, темно-зеленым – Лар, синим – Ядне, красным – Яр.

рунгуй не выявлена эта основная для хантыйских родов гаплогруппа: два образца принадлежат к гаплогруппе N1a2b1b1b-B172 (рис. 6) и один к N1a1a1a1a2-Z1936 (рис. 10). Гаплогруппа N1a1a1a1a2a1c-Y13850 также обнаружена нами и в выборке казымских хантов, из северной части их ареала.

Общий возраст для всего массива гаплотипов основной гаплогруппы N1a1a1a1a2a1c-Y13850 составил 1752 года (SD = 517 лет), для рода Салиндер – 843 года (SD = 383 года), для рода Тибичи – 607 лет (SD = 342 года). Общий возраст этой гаплогруппы у гыданских ненцев может быть несколько завышен, поскольку между разными родами наблюдаются значительные различия размера аллелей по некоторым YSTR-маркерам, которые могли возникнуть в результате нескольких совместных мутаций у отдельного индивида или накопиться за два–три поколения его потомков.

Данные по гаплогруппам Y-хромосомы полностью подтверждают хантыйское происхождение этих родов, их отличие от фратрии Харючи, других родов фратрии Вануйто и генетическое родство друг другу.

Энецкие рода. Марьик, Няч, Оковай и Тёр, относящиеся к родам энецкого происхождения, полностью состоят из представителей отдельной гаплогруппы N1a2b1-B478,Z35080(xVL65) (рис. 8). Гаплотипы этой линии также образуют звездообразную филогению, с относительно небольшой дисперсией аллелей, что свидетельствует об их тесном генетическом родстве и относительно недавнем общем происхождении.

Вытеснение энцев ненцами из междуречья Таза и Енисея началось с двух “больших” военных столкновений между ними во второй половине XVII в. В следующем столетии ненцы продолжали проникновение на восток для освоения завоеванных территорий, а к XIX в. окончательно закрепились в низовьях Таза и Енисея. Оставшиеся на Тазу энцы были ассимилированы ненцами, а энецкие рода включены в ненецкую двухфратриальную систему. Они стали относиться к фратрии Вануйто, в которую входят потомки аборигенного населения и различные иноплеменники, в отличие от фратрии Харючи, состоящей из потомков самодийцев, выходцев из Южной Сибири [4].

Полученные результаты позволяют более точно взглянуть на историю формирования гыданских ненцев, с одной стороны, на самом современном уровне генотипирования Y-хромосомных маркеров, с другой стороны, полностью учитывая родовую структуру народа, анализируя межэтнические взаимосвязи и родовые различия.

По результатам генотипирования удалось выявить в подавляющем большинстве родов основную гаплогруппу, представителей которой среди исследованной выборки образцов этого рода абсолютное большинство. Генотипирование широкого набора YSTR-маркеров позволило получить достаточно детальные и специфичные кластеры гаплотипов, что позволяет с высокой степенью достоверности определить не только этническую

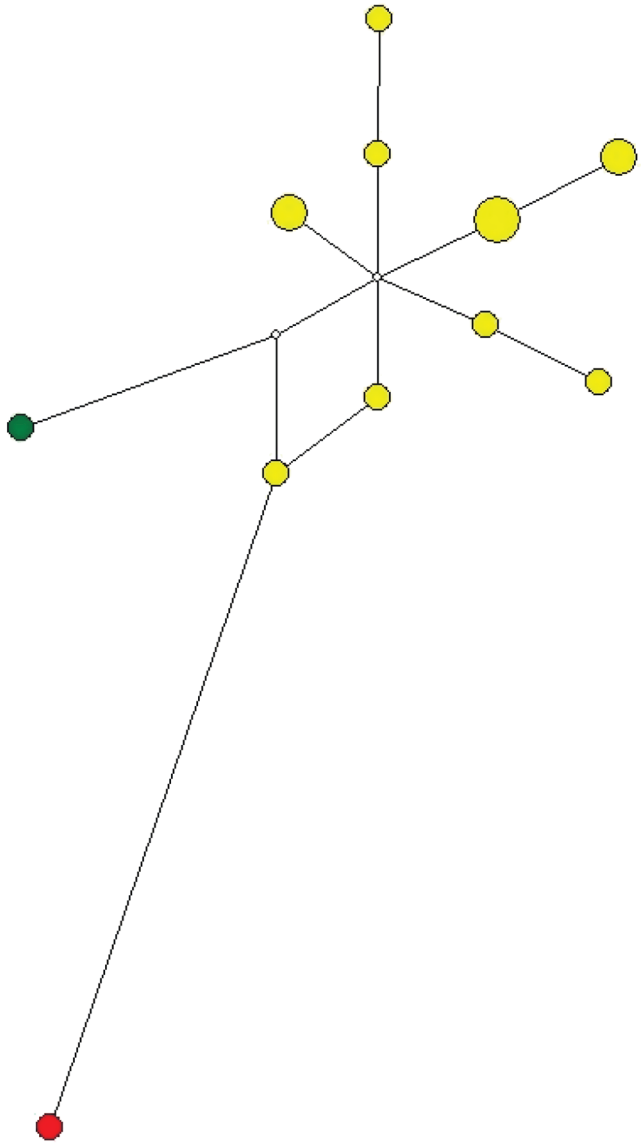

Рис. 10. Медианная сеть гаплогруппы N1a1a1a1a2-Z1936. Желтым цветом обозначен род Яр, зеленым – Пурунгуй, красным – Аседа.

и субпопуляционную принадлежность образцов, но и в большинстве случаев определить их принадлежность к конкретным роду и фамилии. Практически каждый род гыданских ненцев характеризуется полным доминированием одной гаплогруппы со специфичным спектром гаплотипов, подчеркивающим недавний эффект основателя. Анализ гаплотипов убедительно показал, что представители одного рода являются родственниками по мужской линии и имеют родоначальника, жившего в относительно недалеком прошлом. Таким образом, генофонд представителей восточной группы ненецкого этноса, а точнее часть, маркируемая гаплогруппами Y-хромосомы, действительно структурирована прежде всего по родовому принципу. Именно этот уровень организации генофонда гыданских ненцев как единой системы является наиболее точно характеризующим его популяционную структуру. Две фратрии гыданских ненцев значительно различаются по генетической структуре родов по маркерам Y-хромосомы, что подтверждает их формирование на основе различных предковых компонент. Ненецкие рода хантыйского и энецкого происхождения по составу Y-хромосомных гаплогрупп полностью отличаются друг от друга и от самодийской по происхождению фратрии Харючи и аборигенной Вануйто.

Тестирование разработанного набора YSTR для определения этнической принадлежности на

примере ненцев продемонстрировало его высокую эффективность и достоверность не только при определении популяционной принадлежности, но и более точную детализацию при работе с широкими и репрезентативными выборками для отдельных родовых и фамильных групп. Это позволяет точнее и эффективнее устанавливать привязку образцов мужской ДНК и открывает значительные перспективы для применения этого набора YSTR в реальной практике.

Работа выполнена в рамках Научно-технической программы Союзного государства "Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства" ("ДНК-идентификация"), Государственный контракт № 011-17 от 26.09.2017, а также поддержана грантом РФФИ № 18-29-13045 "Популяционная геномика и транскриптомика человека: поиск сигналов не-нейтральной эволюции".

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

1. *Волжанина Е.А.* Демографический облик ненцев ямала по материалам всероссийских переписей XXI в. // Вестник археологии, антропологии и этнографии. 2017. № 3(38). С. 120–130.
2. *Квашин Ю.Н.* Этнические и демографические процессы у тазовских ненцев в начале XXI в. // Вестник археологии, антропологии и этнографии. 2012. № 3(18). С. 141–152.
3. *Хомич Л.В.* Проблемы этногенеза и этнической истории ненцев. Л.: Наука, Ленингр. отд-е, 1976. 187 с.
4. *Васильев В.И.* Проблема формирования енисейских ненцев (к вопросу об этнической природе этнографических групп в составе современных народностей Севера) // Этногенез и этническая история народов Севера. М.: Наука, 1975. 89 с.
5. *Квашин Ю.Н.* Основные элементы родовой структуры гыданских ненцев // Вестник археологии, антропологии и этнографии. 2001. № 3. С. 1–4.
6. *Abilev S., Malyarchuk B., Derenko M. et al.* The Y-chromosome C3*Star-Cluster attributed to Genghis Khan's Descendants is present at hight frequency in the Kerey Clan from Kazakhstan // Hum. Biol. 2012. V. 84. № 1. P. 79–89. https://doi.org/10.3378/027.084.0106
7. *Zhabagin M.K., Dibizova H.D., Frolova S.A. et al.* The relation between the Y-chromosomal variation and the clan structure: The geen pool of the steppe aristocracy and the steppe clergy of the Kazakhs // Moscow Univ. Anthropol. Bul. 2014. V. 1. P. 96–101.
8. *Балановская Е.В., Юсупов Ю.М., Схаляхо Р.А. и др.* Генетические портреты семи кланов северо-западных башкир: вклад финно-угорского компонента в генофонд башкир // Вестник Моск. ун-та. Серия XXIII. Антропология. 2017. № 3. С. 94–103.
9. *Харьков В.Н., Новикова Л.М., Штыгашева О.В. и др.* Генофонд хакасов и шорцев по маркерам Y-хромосомы: общие компоненты и генетическая структура родов // Генетика. 2020. Т. 56. № 7. С. 826–833. https://doi.org/10.31857/S0016675820070073
10. *Харьков В.Н., Зарубин А.А., Вагайцева К.В. и др.* Y-хромосома как инструмент для ДНК-идентификации и определения популяционной принадлежности // Генетика. 2020. Т. 56. № 9. С. 1065–1074. https://doi.org/10.31857/S0016675820090118
11. The Y-Chromosome Consortium. A nomenclature system for the tree of human Y-chromosmal binaryhaplogroups // Genome Res. 2002. V. 12. P. 339–348.
12. *Karafet T.M., Mendez F.L., Meilerman M.B. et al.* Newbinary polymorphisms reshape and increase resolutionof the human Y chromosomal haplogroup tree // Genome Res. 2008. V. 18. P. 830–838.
13. *Bandelt H.J.* Median-joining networks for inferring intraspecific phylogenies // Mol. Biol. Evol. 1999. V. 16. P. 37–48.
14. *Sengupta S., Zhivotovsky L.A., King R. et al.* Polarity and temporality of high resolution Y-chromosome distributions in India identify both indigenous and exogenous-expansions and reveal minor genetic influence of central Asian pastoralists // Am. J. Hum. Genet. 2006. V. 78. P. 202–221. https://doi.org/10.1086/499411
15. *Долгих Б.О.* Очерки по этнической истории ненцев и энцев. М.: Наука, 1970. 268 с.
16. *Харьков В.Н., Хамина К.В., Медведева О.Ф. и др.* Генофонд бурят: клинальная изменчивость и территориальная подразделенность по маркерам Y-хромосомы // Генетика. 2014. Т. 50. № 2. С. 203–213. https://doi.org/10.7868/S0016675813110088
17. *Квашин Ю.Н.* Гыданские ненцы: история формирования современной родовой структуры (XVIII–XX вв.). М.: ИЭА РАН, 2003. 186 с.

Reconstruction of the Origin of the Gydan Nents Based on Genetic Analysis of Their Tribal Structure Using a New Set of YSTR-Markers

V. N. Kharkov[a, *], L. V. Valikhova[a], E. L. Yakovleva[a], V. N. Serebrova[a], N. A. Kolesnikov[a], T. I. Petelina[b], I. Yu. Khitrinskaya[a], and V. A. Stepanov[a]

[a] Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, 634050 Russia

[b] Tyumen Cardiological Research Center, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tyumen, 625026 Russia

**e-mail: vladimir.kharkov@medgenetics.ru*

The genetic structure of the Nenets clans was investigated by Y-chromosome markers using the developed YSTR set to determine the population origin of an unknown individual. The analysis of the frequencies of haplogroups and YSTR-haplotypes indicate that the Nenets clans are related associations, in most cases having one ancestor in the male lineage. It is shown that the Y-chromosomal gene pool of the Nenets is structured according to the tribal principle. For the vast majority of samples, a close genetic affinity of representatives of the same tribe and a relatively recent founder effect in the male lineage were revealed. The structure of the haplotypes of the Y-chromosome haplogroups in the Nenets gene pool demonstrates the genetic affinity of the Nenets with the Khanty and Enets. significant differences between the two Nenets phratries were revealed. Estimation of the age of origin of individual Y-chromosome sublines and tribal clusters of haplotypes is fully consistent with ethnographic data on the formation of tribes and phratries within the Gydan Nenets. The results of a comparative analysis of the spectrum of haplotypes convincingly indicate the specificity of Y-chromosome sublines and clusters of haplotypes not only at the level of ethnic groups and populations, but also at the level of phratries, tribes, and individual family groups. With a few exceptions, each phratry and tribe has its own specific cluster of haplotypes, equidistant from each other.

Keywords: Y-chromosome, gene pool, population, genetic diversity, Nenets, genus, DNA identification.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1415–1422

УДК 575.1:618.39-021.3

АССОЦИАЦИЯ МЕТИЛИРОВАНИЯ ГЕНОВ НЕЙРОМЕДИАТОРНЫХ СИСТЕМ МОЗГА С ПСИХОЭМОЦИОНАЛЬНЫМИ ХАРАКТЕРИСТИКАМИ ЧЕЛОВЕКА

© 2021 г. И. Б. Моссэ[1], *, Н. Г. Седляр[1], А. С. Бабенко[2], К. А. Моссэ[1], Р. С. Шулинский[1], А. В. Кильчевский[1]

[1]*Институт генетики и цитологии Национальной академии наук Беларуси, Минск, 220072 Республика Беларусь*

[2]*Белорусский государственный медицинский университет, Минск, 220116 Республика Беларусь*

**e-mail: I.Mosse@igc.by*

Поступила в редакцию 22.06.2021 г.
После доработки 28.06.2021 г.
Принята к публикации 29.06.2021 г.

Эпигенетическая регуляция активности генов вовлечена в развитие ряда психических нарушений, включая нарушения развития головного мозга, личностные расстройства и девиантное поведение. Проведено исследование уровня метилирования 24 избранных CpG-сайтов в группах пациентов с перечисленными нарушениями и контрольной группе. Показано, что CpG-сайты в регуляторных участках трех генов нейромедиаторных систем человека – *BDNF*, *COMT* и *CACNA1C* – гиперметилированы, что может приводить к снижению экспрессии указанных генов. Полученные результаты могут быть основой для разработки диагностических тест-систем и поиска фармакологических мишеней для выявления и коррекции указанных состояний.

Ключевые слова: метилирование CpG-сайтов, эпигенетическая регуляция, высокопроизводительное секвенирование, психоэмоциональный статус.

DOI: 10.31857/S0016675821120110

Интерес к эпигенетическим механизмам резко возрос после того, как появилось большое количество данных о роли эпигенетических процессов в основополагающих биологических явлениях таких как онтогенез, старение, а также в развитии ряда заболеваний (ожирение, наркомания, алкоголизм, шизофрения, личностные расстройства, депрессивные состояния и т.п.). Изменения экспрессии генов, возникшие под влиянием стрессорных средовых факторов, приобретают устойчивый долговременный характер, т.е. остаются относительно неизменными в течение длительного времени, часто на всю оставшуюся жизнь (поэтому кратковременные адаптивные изменения, связанные с изменением генной экспрессии, под понятие эпигенетических сдвигов не подпадают).

Возникшие изменения активности могут сохраняться на протяжении двух–трех поколений, особенно при продолжении воздействия стрессовых средовых факторов [1]. На существующую генетическую программу “накладываются” эпигенетические метки, индуцированные неблагоприятными жизненными обстоятельствами, в том числе психосоциальными факторами.

Эпигенетические механизмы играют важную роль в стресс-индуцированных нарушениях поведения и психики. Фактически можно говорить о психиатрической и поведенческой эпигенетике как о самостоятельном разделе исследований [2]. Стресс при участии эпигенетических механизмов выступает как один из основных факторов, провоцирующих ряд расстройств психики.

Оценка статуса метилирования специфических локусов геномной ДНК является эффективным способом отслеживания, изучения и контроля уникальных процессов, регулирующих жизнедеятельность человека [3, 4]. Эпигенетические стресс-индуцированные феномены при этом становятся ведущим фактором нарушений психического здоровья, прежде всего депрессии и тревоги, но также личностных расстройств, зависимостей и асоциального поведения [5, 6]. Выявление ключевых генетическихи эпигенетических маркеров риска развития психических расстройств и агрессивного поведения позволяет оценить вероятность наличия анализируемых признаков у индивида по его ДНК в целях диагностики, выявления фармакологических мишеней и возможного применения в криминалистике.

Наиболее изученный вид эпигенетической модификации ДНК у людей – метилирование цитозина в положении С5 в CpG-динуклеотидах. Более 28 миллионов сайтов CpG распределены по всему геному человека и 70–80% из них могут быть метилированы. В большинстве случаев гиперметилирование промоторных областей приводит к снижению экспрессия генов [7].

Интерес ученых к статусу метилирования генов, ответственных за психоэмоциональный статус человека, возрастает. Анализ современного состояния проблемы показал, что необходима оценка целого спектра генетических и эпигенетических локусов генома человека. Для решения таких задач активно применяются технологии секвенирования нового поколения (**NGS**) и разработка небольших таргетных панелей. Преимуществом такого подхода является, с одной стороны, возможность исследования десятков и сотен локусов в ходе анализа одного образца, с другой – имеется возможность протестировать в рамках одного эксперимента сотни образцов, что сильно снижает стоимость одного анализа для конечного потребителя, делая его более доступным в рутинной практике [9].

В представленном исследовании выявлены гены нейромедиаторных систем мозга, уровень метилирования CpG-сайтов в регуляторных участках которых ассоциирован с нарушениями психического здоровья человека.

МАТЕРИАЛЫ И МЕТОДЫ

Исследование проведено на двух группах представителей белорусской популяции: контрольная группа условно здоровых добровольцев (**К**) численностью 30 человек и пациенты РНПЦ психического здоровья с органическими нарушениями головного мозга и умственной отсталостью, страдающие личностными расстройствами и девиантным поведением (**П**) – 61 человек.

От каждого участника было получено информированное согласие на включение в исследование. Дизайн исследования утвержден Биоэтическим комитетом Института генетики и цитологии НАН Беларуси.

В качестве биологического материала использовали буккальный эпителий. Выделение ДНК проводили наборами “Нуклеосорб”, модификация А (ОДО Праймтех).

Метилирование ДНК определяли методом бисульфитной конверсии, при которой каждый неметилированный цитозин превращается в урацил посредством дезаминирования. Бисульфитную конверсию геномной ДНК в количестве 1 мкг выполняли с использованием набора реагентов EpiTect Fast DNA Bisulfite Kit (Qiagen) согласно инструкции производителя. Набор оптимизирован для получения после конверсии участков геномной ДНК длиной 300–500 пн и в некотором количестве более 500 пн.

Для анализа уровня метилирования целевых генов использовали разработанную нами панель специфических олигонуклеотидных праймеров, фланкирующих выбранные для исследования локусы [8]. Амплификацию конвертированной ДНК проводили для 17 целевых участков исследуемых генов, которые содержали 24 специфических CpG-сайта метилирования.

Для контроля качества и определения концентрации полученных фрагментов ДНК использовали прибор для флуориметрического анализа Glomax Explorer (Promega) с набором реагентов QuantiFluor® ONE dsDNA System (Promega).

Нормализацию и пулирование образцов выполняли с помощью системы автоматического дозирования Eppendorf epMotion. Для приготовления NGS-библиотек использовали наборы реагентов QIAseq Methyl Library Kit (Qiagen), позволяющие работать с фрагментированными участками ДНК после бисульфитной конверсии. Конечная концентрация анализируемой библиотеки составила 8 пмоль, концентрация контрольной библиотеки PhiX – 10%.

Высокопроизводительное секвенирование выполнялось одновременно для 24 образцов на приборе IlluminaMiseq. Для запуска использован набор v3 2 × 250 (500 cycles) Illumina – MiSeq Reagent Kit v3.

Оценку качества прочтений проводили в программе FastQC, с последующим триммингом некачественных оснований в Trimmomatic. Локальное картирование прочтений на референс (GRCh37 [GCF_000001405.13]) проводили по алгоритму Burrows-Wheeler-Aligner при помощи программного модуля Bowtie2 в утилите Bismark.

Фильтрацию с последующим преобразованием карты выравнивания прочтений (**SAM**), также сортировку и индексирование проводили в Samtools. Вызов сайтов метилирования на “изначально верхней цепи” (original top strand) в CpG-островах проведено в пайплайне Bismark. Подсчет отношения деметилированных сайтов к метилированным был произведен на языке программирования Python с использованием библиотеки Pandas.

Сравнение уровней метилирования отдельных CpG-сайтов в группах исследования проводилось с использованием критериев Манна–Уитни. Различия принимались как статистически значимые при $p < 0.05$.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

На первом этапе исследования анализ результатов высокопроизводительного секвенирования позволил получить данные об уровне метилиро-

вания 24 целевых CpG-сайтов в образцах ДНК от 61 пациента и 30 человек из группы контроля.

Наибольшие различия между группой пациентов и контрольной группой были зафиксированы для шести CpG-сайтов, расположенных в промоторных областях трех генов – *BDNF*, *COMT* и *CACNA1C*.

Ген *BDNF*, локализованный в хромосомном локусе 11p1, кодирует мозговой нейротрофический фактор – цитокин, представляющий собой гомодимерный белок с молекулярной массой 27 кДа. *BDNF* оказывает уникальное действие на ЦНС, стимулируя в нейронах репаративные процессы, повышая синаптическую пластичность и устойчивость клеток ЦНС к повреждающим факторам [9, 10]. Дефицит *BDNF* снижает пластичность нейронов, нарушает память и обучаемость, когнитивные способности. Экспрессия *BDNF* значительно уменьшается у пациентов с депрессией, аффективными расстройствами, при стрессе [11, 12]. Хронический стресс приводит к снижению продукции *BDNF*, которая участвует во многих клеточных процессах в нервной ткани, особенно связанных с долговременной памятью, путем активизации метилирования гистонов и изменения степени компактизации хроматина.

BDNF – один из наиболее изученных генов у пациентов с психоэмоциональными нарушениями и в том числе с биполярным расстройством (**БР**). В частности, повышенное метилирование области промотора *BDNF* (экзон 1) было обнаружено в мононуклеарных клетках периферической крови пациентов с БР второго типа, у которых уровни экспрессии гена *BDNF* также были значительно снижены по сравнению с контрольной группой [13].

В другом исследовании степень метилирования ДНК различалась у пациентов с БР и контрольной группы для 11 из 36 CpG-сайтов, проанализированных в промоторах 3 и 5 гена *BDNF* [14]. Несколько дифференциально метилированных сайтов перекрывались с участками связывания факторов транскрипции в промоторе 5 или непосредственно рядом с ними, предполагая возможную роль метилирования ДНК в модуляции экспрессии гена *BDNF*.

Ген *COMT* расположен на участке 22q11.21, кодирует фермент катехол-О-метилтрансферазу, участвующую в обмене различных катехоламинов (адреналин, норадреналин, дофамин) [15, 16]. Физиологическая роль фермента заключается в элиминации катехоламинов из синаптической щели путем реакции метилирования. Динамика катехоламинов в разных участках мозга регулирует социальное поведение. Катехол-О-метилтрансфераза регулирует передачу нервного импульса, влияет на особенности эмоциональных реакций, участвует в метаболизме эстрогенов [16]. Существуют две изоформы данного фермента: S-COMT (растворимая изоформа, обнаружена в крови и периферических тканях) и MB-COMT (мембрансвязанный протеин, основная часть которого локализована в префронтальной коре) [17]. Известно около 300 полиморфных сайтов [17], для некоторых из которых показано клиническое значение [18].

Гиперметилирование сайта COMT_19949901 (cg04856117) было также зафиксировано при изучении степени метилирования ДНК 41-ного сайта CpG в гене *COMT* у 185 человек с посттравматическим стрессовым расстройством и контрольной группы [19]. По результатам исследования увеличение уровня метилирования только двух сайтов, расположенных в промоторе гена, включая cg04856117, было независимо связано с данным состоянием.

Ген *CACNA1C* расположен на участке хромосомы 12p13.3, кодирует белок главного потенциалзависимого кальциевого канала L-типа, Cav1.2 (субъединица альфа-1С). Каналы Cav1.2 являются критическими модуляторами ряда клеточных процессов, которые участвуют в развитии психопатологии [20]. Полногеномные исследования выявили ассоциацию гена *CACNA1C* с множественными психоневрологическими расстройствами, включая большое депрессивное расстройство, биполярное расстройство и шизофрению, которые проявляются депрессивными симптомами [21].

Метилирование гена *CACNA1C* также изучалось у пациентов с БР – как одна из возможных причин данного состояния. Среди 5-CpG-островков, расположенных в различных частях гена, вариации в уровнях метилирования были зафиксированы только для сайтов, расположенных в третьем интроне гена. При этом пять из шести сайтов показали значительное гиперметилирование в ДНК пациентов с БР по сравнению с контролем. Также обнаружено, что на уровень метилирования влияет пол и также близлежащие полиморфные сайты [22].

Распределение уровней метилирования шести сайтов, расположенных в промоторных областях генов *BDNF*, *COMT* и *CACNA1C*, в исследованных группах представлено в виде графиков box-plot на рис. 1–3. Для всех шести сайтов выявленные различия между группами были статистически достоверны ($p < 0.05$). Результаты представлены в табл. 1.

Для того, чтобы оценить, насколько изменения в уровне метилирования целевых сайтов у пациентов с личностными расстройствами и асоциальным поведением отражают общие изменения в метилировании промоторных областей генов *BDNF*, *COMT* и *CACNA1C*, мы дополнительно проанализировали все CpG-сайты, выявленные во фрагментах ДНК генов в исследованных ам-

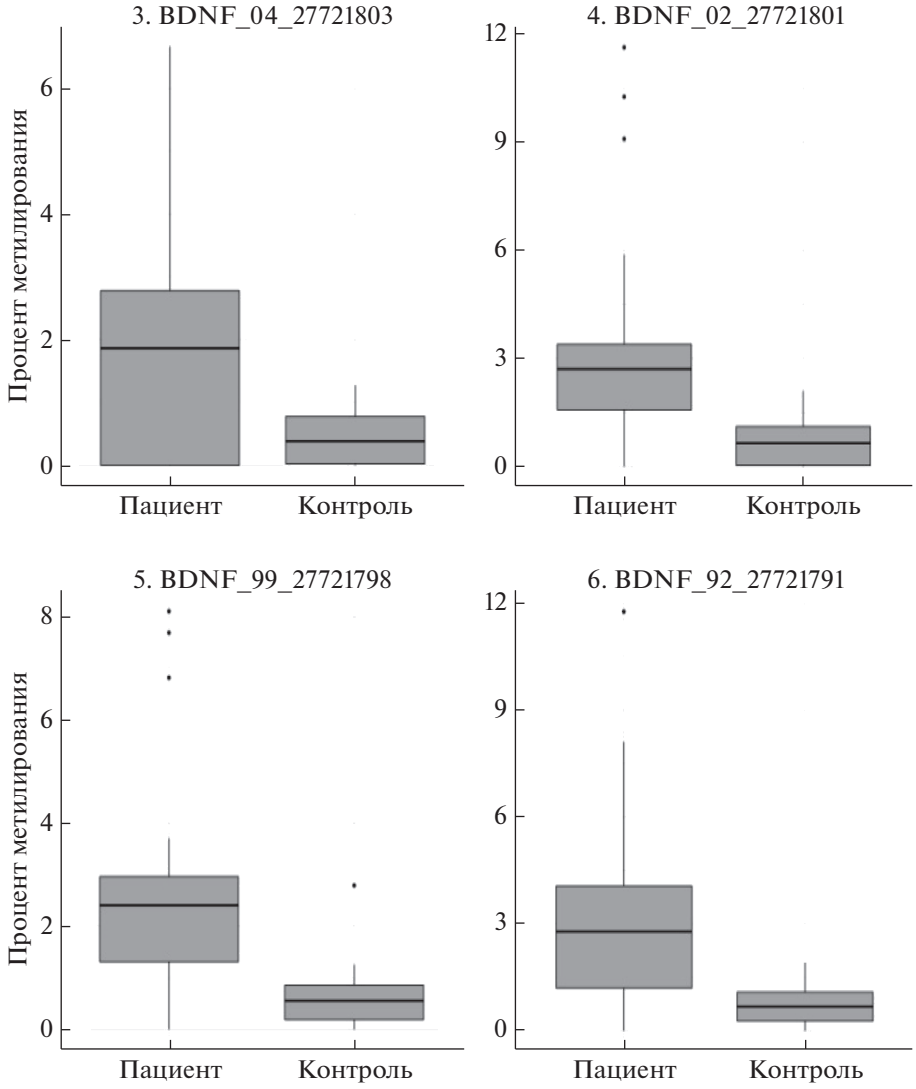

Рис. 1. Распределение уровней метилирования четырех сайтов гена *BDNF* в группах пациентов и контроля.

пликонах. Данные об их метилировании были также получены при проведении NGS. Для сравнительного анализа большого числа дифференциально метилированных сайтов применяли карты интенсивности сигналов (heatmap).

Наиболее значимые результаты получены для 18 сайтов, входящих в состав промотора гена *CACNA1C* (рис. 4). Визуальное сравнение интенсивности сигналов четко указывает на более высокий уровень метилирования у пациентов по

Таблица 1. Сравнение уровней метилирования целевых сайтов генов *BDNF*, *COMT* и *CACNA1C* в группах пациентов и контроля

Анализируемые CpG-сайты	Уровень метилирования		Уровень значимости *p*
	пациенты	контроль	
BDNF_27721791	2.78	0.67	0.00007816
BDNF_27721798	2.41	0.55	0.00003795
BDNF_27721801	2.70	0.65	0.000003296
BDNF_27721803	1.86	0.38	0.004353
COMT_19949901	75.87	63.76	0.000071
CACNA1C_2339515	56.36	46.35	0.00001048

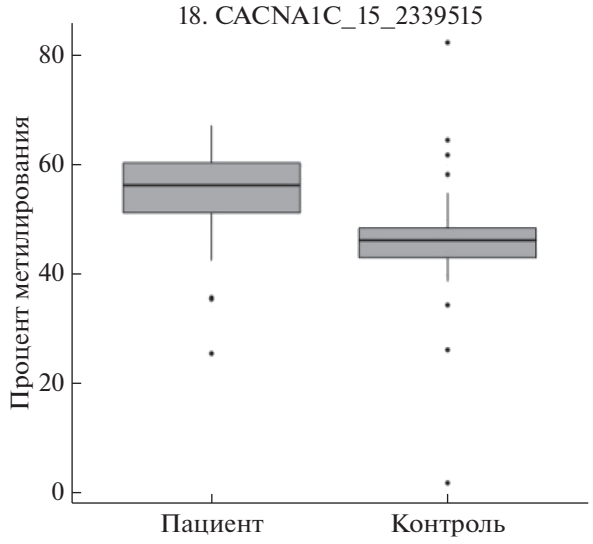

Рис. 2. Распределение уровней метилирования сайта 2339515 гена *CACNA1C* в группах пациентов и контроля.

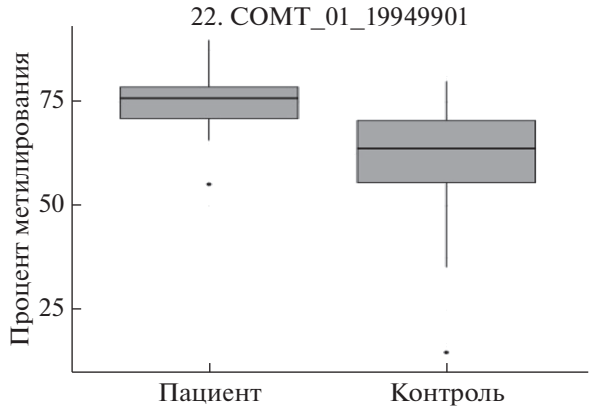

Рис. 3. Распределение уровней метилирования сайта 19949901 гена *COMT* в группах пациентов и контроля.

сравнению с контролем. При иерархической кластеризации по уровню метилирования 18 сайтов гена наблюдается четкое объединение образцов ДНК пациентов в отдельные кластеры. Анализ распределения образцов показал, что построение дендрограммы на основании статуса метилирования позволяет выделить в отдельную группу более 85% пациентов.

Таким образом, выявлены информативные гены, анализ метилирования регуляторных областей которых может рассматриваться как высокоспецифичный тест для определения психоэмоционального статуса человека как в научных целях, так и в практике здравоохранения и криминалистики. Полученные результаты должны быть подтверждены на независимых выборках.

Работа проведена в рамках реализации мероприятия 5 “Разработка методики определения статуса генетических локусов и прижизненной модификации участков ДНК, влияющих на психоэмоциональный статус человека” Научно-технической программы Союзного государства “Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства” (“ДНК-идентификация”).

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.

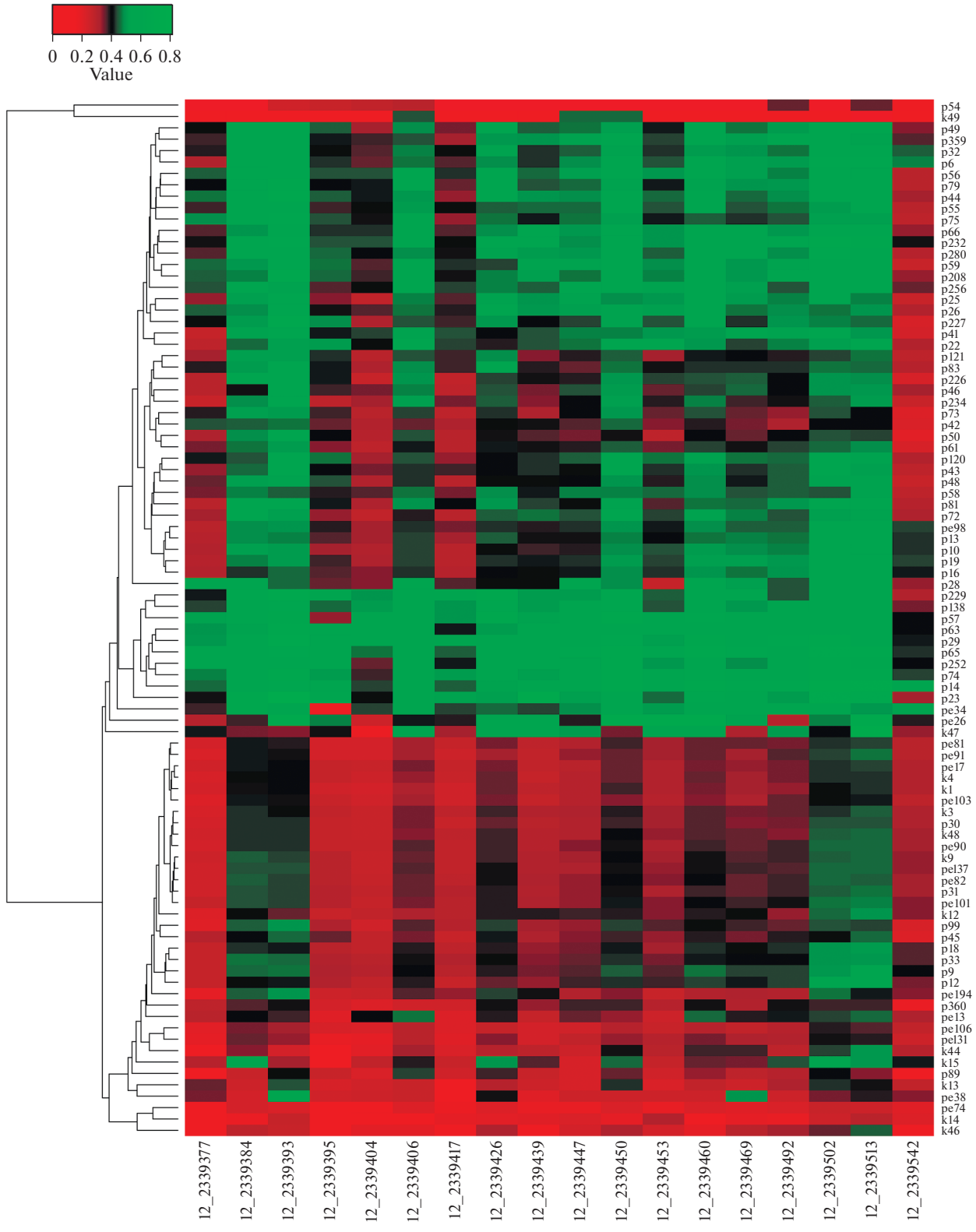

Рис. 4. Карта интенсивности сигналов (heatmap) дифференциально метилированных сайтов в промоторной области гена *CACNA1C*. Справа по вертикали – шифры образцов. По горизонтали – сайты метилирования.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Sasaki A., de Vega W.C., McGowan P.O.* Biological embedding in mental health: An epigenomic perspective // Biochem. and Cell. Biol. 2013. V. 9. № 1. P. 14–21. https://doi.org/10.1139/bcb-2012-0070
2. *Provenkal N., Binder E.* The neurobiological effects of stress as contributors to psychiatric disorders: Focus on epigenetics // Curr. Opinion in Neurobiol. 2015. V. 30. P. 31–37. https://doi.org/10.1016/j.conb.2014.08.007
3. *Rozanov V.A.* Epigenetics: Stress and behavior // Neurophysiology. 2012. V. 44. № 4. P. 332–350.
4. *McEwen B.* Brain on stress: How the social environment gets under the skin // Proc. Natl Acad. Sci. USA. 2012. V. 109. № 2. P. 17180–17185.
5. *Campbell K.A., Colacino J.A., Park S.K., Bakulski K.M.* Cell types in environmental epigenetic studies: Biological and epidemiological frameworks // Curr. Envir. Health Repot. 2020. V. 7. № 3. P. 185–197. https://doi.org/10.1007/s40572-020-00287-0
6. *Hing B., Gardner C., Potash J.B.* Effects of negative stressors on DNA methylation in the brain: Implications for mood and anxiety disorders // Am. J. Med. Genet. Neuropsyshiatric Genet. 2014. V. 165B. № 7. P. 541–554. https://doi.org/10.1002/ajmg.b.32265
7. *Heller D., Grodner D., Tanenhaus M.K.* The role of perspective in identifying domains of reference // Cognition. 2008. V. 108. № 3. P. 831–836. https://doi.org/10.1016/j.cognition.2008.04.008
8. *Бабенко А.С., Моссэ К.А., Седляр Н.Г. и др.* Разработка панели для анализа профиля метилирования целевых эпигенетических локусов, ассоциированных с психоэмоциональным статусом человека // Мол. и прикл. генетика. 2020. Т. 29. С. 37–48.
9. *Shimizu E., Hashimoto K., Okamura N. et al.* Alterations of serum levels of brain-derived neurotrophic factor (BDNF) in depressed patients with or without antidepressants // Biol. Psychiatry. 2003. V. 54. № 1. P. 70–75. https://doi.org/10.1016/s0006-3223(03)00181-1
10. *Bocchio-Chiavetto L., Bagnardi V., Zanardini R. et al.* Serum and plasma BDNF levels in major depression: A replication study and meta-analyses // World J. Biol. Psychiatry. 2010. V. 11. № 6. P. 763–773. https://doi.org/10.3109/15622971003611319
11. *Terracciano A., Lobina M., Grazian Piras M. et al.* Neuroticism, depressive symptoms, and serum BDNF // Psychosom. Med. 2011. V. 73. № 8. P. 638–642. https://doi.org/10.1097/PSY.0b013e3182306a4f
12. *Minelli A., Zanardini R., Bonvicini C. et al.* BDNF serum levels, but not BDNF Val66Met genotype, are correlated with personality traits in healthy subjects // Eur. Arch. Psychiatry Clin. Neurosci. 2011. V. 261. № 5. P. 323–329. https://doi.org/10.1007/s00406-011-0189-3
13. *Dell'Osso B., D'Addario C., Carlotta Palazzo M. et al.* Epigenetic modulation of *BDNF* gene: Differences in DNA methylation between unipolar and bipolar patients // J. Affect Disord. 2014. V. 166. P. 330–333. https://doi.org/10.1016/j.jad.2014.05.020
14. *Strauss J.S., Khare T., De Luca V.* Quantitative leukocyte BDNF promoter methylation analysis in bipolar disorder // Int. J. Bipolar Disord. 2013. V. 12. P. 1–28. https://doi.org/10.1186/2194-7511-1-28
15. *Craddock N., Owen M., O'Donovan M.* The catechol-O-methyl transferase (COMT) gene as a candidate for psychiatric phenotypes: evidence and lessons // Mol. Psychiatry. 2006. V. 11. P. 446–458. https://doi.org/10.1038/sj.mp.4001808
16. *Chen J., Lipska B.K., Halim N. et al.* Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): Effects on mRNA, protein, and enzyme activity in postmortem human brain. // Am. J. Hum. Genet. 2004. V. 75. № 5. P. 807–821. https://doi.org/10.1086/425589
17. *Hu X., Oroszi G., Chun J. et al.* An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk // Alcohol Clin. Exp. Res. 2005. V. 29. № 1. P. 8–16. https://doi.org/10.1097/01.alc.0000150008.68473.62
18. *Зуева И.Б., Улитина А.С., Гораб Д.Н. и др.* Роль аллельных вариантов генов ангиотензин-превращающего фермента ACE и серотонинового транспортера SLC6A4 в развитии когнитивного дефицита у лиц с метаболическим синдромом // Артериальная гипертензия. 2012. Т. 18. № 6. С. 531–539.
19. *Norrholm S.D., Jovanovic T., Smith A.K. et al.* Differential genetic and epigenetic regulation of catechol-O-methyltransferase is associated with impaired fear inhibition in posttraumatic stress disorder // Front. Behav. Neurosci. 2013. V. 10. P. 7–30. https://doi.org/10.3389/fnbeh.2013.00030
20. *Bhat S., Dao D.T., Terrillion C.E. et al. CACNA1C* (Cav1.2) in the pathophysiology of psychiatric disease // Prog. Neurobiol. 2012. V. 99. № 1. P. 1–14. https://doi.org/0.1016/j.pneurobio.2012.06.001
21. *Bagot R.C., Labonte B., Pena C.J., Nestler E.J.* Epigenetic signaling in psychiatric disorders: stress and depression // Dialogues Clin. Neurosci. 2014. V. 16. № 3. P. 281–295. https://doi.org/10.31887/DCNS.2014.16.3/rbagot
22. *Starnawska A., Demontis D., Pen A. et al. CACNA1C* hypermethylation is associated with bipolar disorder // Trans. Psychiatry. 2016. V. 6. № 6. P. 1–7. https://doi.org/10.1038/tp.2016.99

Association of Methylation of Neuromediator Brain System Genes with Psychoemotional Human Characteristics

I. B. Mosse[a], *, N. G. Sedlyar[a], A. S. Babenko[b], K. A. Mosse[a], R. S. Shulinsky[a], and A. V. Kilchevsky[a]

[a] *Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, 220072 Republic of Belarus*

[b] *Belarusian State Medical University, Minsk, 220116 Republic of Belarus*

**e-mail: I.Mosse@igc.by*

Epigenetic regulation of gene activity has been implicated in a number of mental disorders, including developmental disorders of the brain, personality disorders, and deviant behavior. A study of the methylation level of 24 selected CpG sites in the groups of patients with the listed disorders and the control group was carried out. It was shown that CpG sites in the regulatory regions of three genes of human neuromediator systems – BDNF, COMT and CACNA1C – are hypermethylated, what can lead to a decrease in the expression of these genes. The results obtained can be the basis for the development of diagnostic test systems and the search for pharmacological targets for the detection and correction of these conditions, and are also of interest for personal identification inforensics.

Keywords: methylation of CpG sites, epigenetic regulation, high-throughput sequencing, psychoemotional status.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1423–1437

УДК 575.174.015.3

ИСПОЛЬЗОВАНИЕ МАСШТАБНОГО ПАРАЛЛЕЛЬНОГО СЕКВЕНИРОВАНИЯ В КРИМИНАЛИСТИКЕ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПЛАТФОРМ ДЛЯ СЕКВЕНИРОВАНИЯ

© 2021 г. Т. В. Тяжелова[1, *], И. Л. Кузнецова[1], Т. В. Андреева[1, 2], С. С. Кунижева[1, 2], Е. И. Рогаев[1, 2, 3]

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

[2]Московский государственный университет им. М.В. Ломоносова, Москва, 119991 Россия

[3]Научно-технологический университет "Сириус", Краснодарский край, Сочи, 354340 Россия

*****e-mail: tatvltyz@gmail.com*

Поступила в редакцию 10.06.2021 г.
После доработки 18.06.2021 г.
Принята к публикации 20.07.2021 г.

Последнее десятилетие характеризуется значительным прорывом в развитии новых технологий и инструментов для анализа ДНК. Новая технология масштабного параллельного секвенирования (МПС), иначе называемая "секвенирование следующего поколения", к настоящему времени уже широко используется во многих направлениях биологических и биомедицинских исследований. Внедрение новых технологий параллельного секвенирования в область судебной экспертизы происходит медленнее по сравнению с научными и клиническими исследованиями, в основном в связи со сложностями совмещения с уже имеющимися базами данных ДНК-профилей, а также необходимостью проведения строгих валидационных экспериментов для аккредитации приборов и наборов реагентов. В обзоре представлены существующие в мире решения по использованию масштабного параллельного секвенирования в криминалистике, технические аспекты экспериментов, разработанные коммерческие решения и перспективы применения, а также рассмотрены преимущества и недостатки новых технологий секвенирования по сравнению с методами ДНК-анализа, традиционно используемыми в судебной экспертизе.

Ключевые слова: масштабное параллельное секвенирование (МПС), секвенирование следующего поколения (NGS), короткий тандемный повтор (STR), однонуклеотидный полиморфизм (SNP), криминалистика.

DOI: 10.31857/S0016675821120122

Перспектива одновременного анализа большого количества полиморфных маркеров делает технологию масштабного параллельного секвенирования (**МПС**) очень мощным и относительно легко применимым инструментом в лабораториях судебной экспертизы. В настоящее время наиболее активно МПС используется в диагностических лабораториях для анализа индивидуальных геномов человека или целевого (таргетного) секвенирования выборочных участков ДНК для выявления генетических вариантов, связанных со здоровьем или с развитием заболеваний. В криминалистике с использованием МПС станет возможным одновременный анализ и стандартных аутосомных ДНК-маркеров (STR и SNP), и митохондриальной ДНК, и маркеров половых хромосом [1, 2]. Прежде всего технологии МПС будут иметь решающее значение для анализа ДНК человека в таких случаях, как массовые бедствия или другие события, когда криминалистические образцы ДНК подвергаются негативному воздействию и их качество ухудшается.

О возможностях и перспективах использования МПС в криминалистической работе было впервые объявлено в 2015 г. в годовом отчете Нидерландского института судебной экспертизы (NFI) [3]. Первым институтом, аккредитованным на использование технологии МПС в судебном анализе, стала Лаборатория судебно-медицинской экспертизы ДНК Университета Лейдена. Первыми заявленными результатами были уменьшение ложноположительных совпадений при сравнении разных образцов ДНК и успешное разделение различных профилей ДНК в сложной смеси [4]. В 2015 г. использование технологий NGS в криминалистике ограничивалось лишь отдельными пилотными исследованиями. Однако уже к 2017 г. опубликовано большое количество работ по тестирова-

нию различных приборов для МПС и различных панелей криминалистических маркеров. Проведенный в 2017 г. опрос 33 европейских лабораторий, работающих в области судебно-медицинской экспертизы, показал, что 52% из них уже приобрели один или более приборов для МПС [5].

МЕТОДОЛОГИЯ МАСШТАБНОГО ПАРАЛЛЕЛЬНОГО СЕКВЕНИРОВАНИЯ

В основе масштабного параллельного секвенирования лежит несколько различных методологических подходов: секвенирование путем синтеза после амплификации ДНК [6], секвенирование отдельной молекулы в реальном времени на ячейке [7] или с использованием нанопор [8]. Большинство платформ секвенирования второго поколения используют метод секвенирования путем синтеза. Эта технология основана на клональной амплификации ДНК ферментом полимеразой. При мономолекулярном секвенировании этап предварительной амплификации отсутствует.

К основным существенным характеристикам различных методов МПС относятся длина прочтений, производительность и частота ошибок секвенирования. Обзор данных характеристик для трех поколений методов секвенирования приведен в табл. 1. Длина прочтения, производительность и время рабочего процесса каждой из платформ в настоящее время изменяются быстрыми темпами в связи с постоянным совершенствованием методологий, поэтому данные, представленные в табл. 1, показывают только общее представление о каждом из поколений методов и платформах для секвенирования. Несмотря на то что уровень ошибок, получаемых при использовании всех существующих платформ секвенирования нового поколения, выше по сравнению с секвенированием методом Сэнгера, этот недостаток преодолевается в технологиях МПС путем многократного прочтения каждого локуса, т.е. высоким покрытием [2].

В настоящее время в сфере судебной генетики наибольшее применение нашли настольные варианты секвенаторов – это прибор MiSeq (и его модификация ForenSeq), разработанный компанией Illumina,а также секвенаторы от компании ThermoFisher Scientific – Ion Torrent PGM и Ion S5 [5].

Платформа Illumina реализует технологию циклического секвенирования в сочетании со стратегией обратимой терминации флуоресцентно меченных модифицированных дизоксирибонуклеотидтрифосфатов, концептуально аналогичной секвенированию методом Сэнгера. Замены оснований – это наиболее частая ошибка, возникающая во время секвенирования на машинах Illumina.

Платформы Ion Torrent (PGM и S5) представляют собой полупроводниковый секвенатор, который измеряет изменения pH, как следствие выделения ионов водорода при синтезе ДНК [9]. Для платформы Ion Torrent наиболее частыми ошибками являются инсерции и делеции гомополимерных участков длиной более 6 пн, поскольку корреляция между включенными нуклеотидами и изменением детектируемого напряжения не пропорциональна размеру гомополимера [8, 10]. Количество ошибок, генерируемых при секвенировании на платформе Ion Torrent (≥1%), выше, чем на платформе Illumina [11]. Однако отсутствие оптического сканирования и циклического секвенирования на платформе Ion Torrent значительного снижает время секвенирования ДНК по сравнению с платформой Illumina [12]. Протоколы ручной подготовки библиотеки ДНК для платформ Ion Torrent PGM и Ion S5 более трудоемки и громоздки по сравнению с рабочим процессом Illumina, но наличие автоматизированной станции подготовки библиотек Ion Torrent (Ion Chef System) значительно упрощает и стандартизирует рабочий процесс. Недавно аналогичные решения по автоматизации пробоподготовки были предложены и для платформы Illumina [13].

МПС КАК ИНСТРУМЕНТ ДЛЯ ГЕНОТИПИРОВАНИЯ STR- И SNP-МАРКЕРОВ

Первыми маркерами, которые стали использоваться для идентификации личности, стали маркеры ДНК, на основе семейств так называемых минисателлитных или гипервариабельных повторов ДНК. Такой тип повторов ДНК в геноме характеризуется высокой аллельной вариабельностью числа тандемных повторов (VNTR, variable number of tandem repeats). На первых этапах внедрения методов ДНК-идентификации в судебную практику для получения индивидуального профиля ДНК человека (ДНК-фингерпринт) использовали достаточно трудоемкую технологию, которая включала расщепление геномной ДНК специальными ферментами на короткие фрагменты, которые разделяли электрофоретически по размеру и гибридизовали с радиоактивно мечеными пробами, соответствующими отдельным VNTR-маркерам [14, 15].

Благодаря быстрому развитию методов молекулярной биологии, в частности появлению метода полимеразной цепной реакции (ПЦР) анализа ДНК, а также разработкам платформ Applied Biosystems для капиллярного эдектрофореза, используемых для прямого определения последовательности ДНК (секвенирования ДНК) или фрагментного анализа ПЦР-продуктов, ДНК-типиро-

Таблица 1. Основные характеристики для трех поколений методов секвенирования

Метод	Год	Принцип действия	Длина чтения, число нуклеотидов (нт)	Производительность и время работы	Ошибка секвенирования, %
I поколение секвенирования					
Сэнгера	1977	Клонирование/обрыв цепи	25–1200	96.84×10^3, 2 ч	
II поколение секвенирования					
454	2005	эм-ПЦР/SBS/пиросеквенирование	100–1000	10^5 нт, 0.7 Гб, 24 ч	0.003
Solexa/HiSeqR_/ MiSeqR_	2006	Мостиковая ПЦР/SBS/обратная терминация	2 × 100 (Hiseq); 2 × 300 (Miseq)	6×10^9 нт, 1.8 Тб, несколько дней (в зависимости от длины чтения)	<0.01
SOLiD	2007	эм-ПЦР/лигирование/ зонды	35–75	6×10^9 нт, 320 Гб, 1–2 недели	0.001
Ion Torrent	2010	эм-ПЦР/ионно-чувствительный SBS/изменение pH	200–400	$60–80 \times 10^6$ нт, 50 Гб, 2 ч	<1
III поколение секвенирования					
PacBio R	2010	Одномолекулярное секвениирование в реальном времени на ячейке (SMRT/ZMW wells)	8000–20000	35×10^4 нт, 7 Гб, 0.5–6 ч	<0.001
Oxford Nanopore	2014	Измерение электрического тока при прохождении молекулы нуклеиновой кислоты через нанопору	500–200000	10^5 нт, 2–4 Тб, до 48 ч	до 15

Примечание. эм-ПЦР – эмульсионная ПЦР, Гб – гигабайт, Тб – терабайт.

вание стало широко использоваться в практике судебных экспертиз. Одним из первых стандартизированных наборов, получившим международное применение в области криминалистики, стал набор базовых STR (short tandem repeats)-локусов CODIS (Combined DNA Index System). Этот набор был разработан по заказу ФБР при совместном усилии корпорации Promega и Applied Bio systems для создания стандартов ДНК-идентификации для национальной базы ДНК США, с целью увеличения эффективности идентификации прежде всего военных лиц, преступников, а также лиц, пропавших без вести. В 1997 г. был утвержден набор CODIS, состоящий из 13 базовых STR-локусов: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11 и амелогенинового маркера пола. Также был создан набор локусов STR-маркеров для территории Евросоюза (ESS): FGA, TH01, D1S1656, D2S441, D3S1358, D8S1179, D10S1248, D12S391, D18S51, D21S11, D22S1045. Локусы, входящие в данные наборы, стали мировым стандартом и были рекомендованы для использования во всех криминалистических лабораториях мира [16].

Традиционный метод STR-анализа при помощи гель- или капиллярного электрофореза (КЭ) основан на выявлении количества повторяющихся “единиц” STR-маркера по размеру ПЦР-ампликона. Анализ таких маркеров с использованием платформ МПС позволяет получить больше информации – определить не только длину повторяющегося участка ДНК, но и полную последовательность как самой области повторов, так и фланкирующих ее регионов, которые часто содержат дополнительные полиморфные варианты. Многочисленные исследования продемонстрировали, что анализ STR-маркеров на основе полной последовательности локуса приводит к значительному увеличению дискриминационного потенциала маркеров по сравнению с анализом, основанным только на длине этих маркеров [17–21]. Также было отмечено, что выявляемые при этом дополнительные полиморфные варианты могут быть популяционно-специфическими. Например, в ис-

следовании четырех основных популяционных групп США было показано, что все локусы стандартных STR-маркеров, за исключением маркера TPOX, содержат дополнительные варианты последовательности [20]. Для каталогизации новых данных вариабельных последовательностей STR-маркеров создан проект STRSeq, который суммирует все наблюдаемые "дополнительные" аллели последовательностей основных локусов STR-маркеров, используемых в судебной экспертизе [21].

Возможность выявить дополнительные вариации в последовательности ДНК при анализе STR-маркеров с использованием методологии масштабного параллельного секвенирования имеет и ряд других преимуществ. Во-первых, увеличивается точность предсказания при использовании этих маркеров для прямого сопоставления или вычислений родственных взаимосвязей. Во-вторых, более удачно решается проблема так называемых "статтер-аллелей" (или "статтер-пиков") – наиболее распространенного артефакта, наблюдаемого при анализе STR-маркеров и вызванного проскальзыванием ДНК во время амплификации. Образование статтер-аллеля может встречаться в 6–10% продуктов амплификации. Если в копируемой нити наблюдается "проскальзывание", то некоторые из амплифицированных цепей получаются короче, чем должен быть продукт ПЦР. С помощью масштабного параллельного секвенирования было показано, что статтер-аллели маркера зависят от нуклеотидной последовательности основного аллеля и, таким образом, уровень статтер-аллелей может быть различным для аллелей одинаковой длины, имеющих различные нуклеотидные составы. Это дает дополнительное преимущество МПС по сравнению с методом фрагментного анализа STR-маркеров, в частности позволяет дифференцировать аллели в смешанных образцах [22]. В-третьих, при работе со смесью образцов ДНК выявление дополнительных аллелей, которые ранее были бы замаскированы из-за одинаковой длины различных аллелей STR-маркера, снижает вероятность случайного совпадения профилей ДНК разных индивидов, присутствующих в смешанном образце. Использование дополнительных вариаций в локусе STR-маркера было успешно продемонстрировано для анализа смеси образцов в работе Phillips и др. [23]. Кроме того, было показано, что число прочтений, полученных при секвенировании, позволяет точно определить пропорции, в которых смешаны образцы [19].

Разделение фрагментов ДНК по размеру при проведении исследования методом МПС не является необходимым условием анализа STR-маркеров, как при использовании фрагментного анализа. Все маркеры могут быть амплифицированы с получением фрагментов наименьшей возможной длины, что важно для анализа деградированной ДНК, а количество одновременно исследуемых маркеров не ограничивается количеством флуоресцентных красителей, которые могут быть детектированы электрофоретическими системами капиллярного электрофореза. Это открывает возможность одновременного использования большего числа STR-маркеров, что, например, было продемонстрировано при использовании большой панели Y-хромосомных и аутосомных STR, применение которых значительно повышает эффективность анализа смешанных образцов мужской и женской ДНК в образцах, полученных после сексуального насилия [24].

Современная судебно-медицинская экспертиза анализа образцов ДНК не ограничивается STR-маркерами. Однонуклеотидные полиморфные варианты (SNP или SNV) имеют ряд преимуществ при анализе деградированных криминалистических образцов ДНК по сравнению с STR-маркерами, потому что для их определения можно использовать более короткие фрагменты ДНК [25]. Например, была показана возможность определения полного профиля из 20 однонуклеотидных маркеров на ДНК из пятнен крови после длительного хранения образцов (243 дня), тогда как только 9% STR-маркеров было возможно типировать на этих же образцах. В случае образцов ДНК, выделенной из слюны после длительного хранения (147 дней), были определены 16 однонуклеотидных маркеров (81%) и 18% STR-маркеров [26].

Также на основании анализа криминалистических образцов различных тканей разной степени деградации была продемонстрирована возможность получения качественного профиля на основании SNP-панелей, тогда как генотипирование микросателлитных маркеров почти не дало положительных результатов [27]. Из 36 различных криминалистических образцов, для которых не смогли получить STR-профили или они были низкого качества, для 16 образцов (45%) были получены полные SNP-профили [28].

За последние несколько лет был опубликован ряд работ, посвященных разработке высокоинформативных SNP-маркеров, эффективных для ДНК-идентификации индивида [29, 30]. В них была показана возможность предсказаний вероятного цвета глаз, волос, кожи, а также этнического происхождения на основании анализа сочетаний гаплотипов SNP-маркеров. Наборы ДНК-маркеров для определения внешности HirisPlex [29], этногеографического происхождения AISNP [30] и идентификации личности [31] в настоящее время лежат в основе большинства коммерческих панелей для ДНК-идентификации методами масштабного параллельного секвенирования (табл. 2).

Таблица 2. Характеристики наиболее распространенных коммерческих наборов реагентов для ДНК-идентификации с помощью МПС

Наименование панели	Состав маркеров (по состоянию на 01.05.2021)	Рекомендации по секвенированию
Фирма Verogene (США)		
ForenSeq DNA Signature Prep Kit	A: 27 маркеров аутосомных STR, 24 Y-STR, 7 X-STR, 94 SNP-маркера для идентификации личности; B: 27 маркеров аутосомных STR, 24 Y-STR, 7 X-STR, 172 SNP-маркера (94 для идентификации личности, 22 для определения фенотипа, 56 для определения биогеографического происхождения)	Количество ДНК 1 нг, число библиотек за прогон 8–96, ~28 ч MiSeq FGx Reagent Kit, ~22 ч MiSeq FGx Reagent Micro Kit
ForenSeq MainstAY Kit	27 маркеров аутосомных STR, 25 Y-STRs-маркеров (минимальный набор для определения Y-гаплотипа по базе YHRD*), средний размер ампликонов 235 пн (максимальный 481 пн)	Количество ДНК 1 нг, число библиотек за прогон 8–96, ~28 ч MiSeq FGx Reagent Kit, ~22 ч MiSeq FGx Reagent Micro Kit
ForenSeq mtDNA Whole Genome Kit	245 ампликонов, восстанавливающих полную последовательность мтДНК генома (16569 пн)	Количество геномной ДНК 100 пг, число библиотек за прогон 16, ~28 ч MiSeq FGx Reagent Kit
ForenSeq mtDNA Control Region Kit	18 ампликонов, восстанавливающих последовательность контрольного региона мтДНК (1200 пн)	Количество геномной ДНК 100 пг, число библиотек за прогон 3–48, ~28 ч MiSeq FGx Reagent Kit, ~22 ч MiSeq FGx Reagent Micro Kit
ForenSeq Kintelligence Kit	10230 SNP-маркеров для судебно-медицинской экспертизы (94 для идентификации личности, 22 для определения фенотипа, 56 для определения биогеографического происхождения, 9867 для определения родства, 106 X-SNP и 85 Y-SNP, средний размер ампликонов 150 пн	Количество ДНК 1 нг, число библиотек за прогон 3, ~28 ч MiSeq FGx Reagent Kit
Фирма Promega (США)		
PowerSeq 46GY System	22 аутосомных STR-маркера, 23 Y-STR-маркера, амелогенин, размер ампликонов 140–300 пн	Количество ДНК 1 нг, число библиотек за прогон 96, ~28 ч MiSeq FGx Reagent Kit
PowerSeq CRM Nested System	HVI, HVII и HVIII контрольного региона MT, размер ампликонов 144–237 пн	Количество ДНК 1 нг, число библиотек за прогон до 16, ~28 ч MiSeq FGx Reagent Kit
Фирма NimaGen (Нидерланды)		
EasySeqTM Human ID and Sample Tracking Kit	34 SNP-маркера для идентификации личности и амелогенин	Количество ДНК 1 нг, число библиотек за прогон до 16, ~28 ч MiSeq FGx Reagent Kit
Фирма QIAGENE		
QIAseq Investigator Missing Persons SNP Panel	1448 ампликонов (1200 аутосомных SNP-маркеров для идентификации личности и родственных связей, 33 X-SNP-маркера, 55 биогеографического происхождения). Панель включает 46 микрогаплотипов	Количество ДНК 10–40 нг, число библиотек за прогон до 60. В случае панели QIAseq Investigator Missing Persons, SNP panel – до 24, ~28 ч MiSeq FGx Reagent Kit возможно использовать платформы и Ion Torrent
QIAseq Investigator ID SNP Panel	139 SNP-маркеров для идентификации личности	

Таблица 2. Окончание

Наименование панели	Состав маркеров (по состоянию на 01.05.2021)	Рекомендации по секвенированию
QIAseq Investigator Ancestry SNP Panel	102 SNP-маркера для определения биогеографического происхождения (Global ancestry), 75 SNP-маркеров для определения биогеографического происхождения (Middle East)	То же
QIAseq Human Mitochondria Panel	111 ампликонов, восстанавливающих полную последовательность МТ-генома (16411 пн)	
QIAseq Investigator Human Mitochondria Control Region Panel	13 ампликонов, восстанавливающих последовательность контрольного региона мтДНК (1200 пн)	
Фирма BGI (Китай)		
MGIEasy Signature Identification Library Prep Kit	52 аутосомных STR-маркера, 27 STR-маркеров X-хромосомы, 48 STR-маркеров Y-хромосомы, 227 SNP-маркеров (145 для идентификации личности, 29 для определения фенотипа, 53 для определения биогеографического происхождения), HVR 1–HVR 3 МТ, амелогенин	Количество ДНК 10–40 нг, число библиотек за прогон до 60, 28 ч MGISEQ-2000
Фирма Thermo Fisher Scientific		
Precision ID GlobalFiler NGS STR Panel	30 аутосомных STR-маркеров, 1 Y-STR-маркер, 3 маркера определения пола (амелогенин, SRY, SNP rs2032678), размер ампликонов 140–300 пн	Количество ДНК 1 нг, число библиотек за прогон 8, 2 ч Ion S5 System
Precision ID Identity Panel	124 SNP-маркера для идентификации личности, средний размер ампликонов 138 пн	
Precision ID Ancestry Panel	165 SNP-маркеров для определения биогеографического происхождения, средний размер ампликонов 127 пн	
Precision ID mtDNA Whole Genome Panel	81 ампликон, восстанавливающий полную последовательность МТ генома (16569 пн), средний размер ампликонов 163 пн	Количество ДНК 100 пг, число библиотек за прогон 8, 2 ч Ion S5 System
Precision ID mtDNA Control Region Panel	7 ампликонов, восстанавливающих последовательность контрольного региона мтДНК, средний размер ампликонов 153 пн	

* YHRD – Y-STR Haplotype Reference Database.

Преимущества анализа SNP-маркеров для судебных целей с помощью МПС, по сравнению с альтернативными технологиями, такие как SNaPshot и системы чипов, заключаются в первую очередь в том, что можно одновременно определять большее количество маркеров на малых количествах ДНК [32–35]. Кроме того, технология МПС позволяет дополнительно к выявлению индивидуальных генетических вариантов определять также микрогаплотипы, т.е. наборы SNP-маркеров, расположенных в непосредственной близости друг от друга (менее 100 пн). Такие микрогаплотипы демонстрируют большой потенциал в криминалистических исследованиях как для целей идентификации неизвестного индивида, так и для определения вероятного этнического происхождения [36–38].

Анализ полной последовательности ампликонов ПЦР при типировании SNP-маркеров с помощью технологии МПС дает также возможность использовать любой наблюдаемый полиморфизм в исследуемом участке ДНК для увеличения дискриминационного потенциала целевого SNP-маркера. Так, при анализе фланкирующих областей локусов, входящих в панель ForenSeq™ DNA Signature Prep Kit, в одной из популяций коренных американцев было обнаружено, что из 94 ампликонов, полученных с помощью набора Forenseq DNA Signature Prep Kit, 22 содержали дополнительные полиморфные варианты, из кото-

Таблица 3. Список маркеров, отмеченных в валидационных исследованиях различными недостатками при генотипировании с помощью МПС

MiSeq FGx	Ion Torrent	BGI	MiniON
STR-маркеры			
DY389II, DY448, DX10148	D3S1358, D7S820, D8S1179	DYS389II, DYS612	–
SNP-маркеры			
Rs459929*, Rs1029047*, Rs2399332*, Rs7251928, Rs7722456*, Rs110488710*	Rs321198, Rs576261, Rs917118, Rs4530059, Rs10318258*	rs1736442, rs1736442, rs914165, rs735480	Rs733164, Rs873196, Rs1029047, Rs1493232, Rs1031825*

* Маркеры удалены из последних версий панелей.

рых 14 оказались информативными, т.е. не были полностью сцеплены с аллелем основного маркера [36].

КОММЕРЧЕСКИЕ НАБОРЫ ДЛЯ КРИМИНАЛИСТИКИ

Одним из первых коммерчески доступных наборов для криминалистического анализа с использованием платформы для масштабного параллельного секвенирования Ilumina MiSeq FGx стал набор реагентов ForenSeq DNA Signature Prep Kit (Verogen, CA), который позволяет одновременно анализировать до 231 маркера. Этот набор реагентов содержит два набора праймеров (набор A и набор B), в состав каждого из которых входят праймеры для анализа используемых в криминалистике 27 аутосомных STR-маркеров, 24 STR-маркера Y-хромосомы, 7 STR-маркеров X-хромосомы, маркеры пола и 94 SNP-маркера для индивидуальной идентификации (набор праймеров A). Набор праймеров B дополнительно включает праймеры для 22 фенотипических маркеров и 56 маркеров этногеографического происхождения [39].

Первая версия этого набора была протестирована Churchill et al. в 2016 г. с помощью секвенатора MiSeq (Illumina) [40]. В работе использовали образцы стандартной ДНК: 2800M (Promega® Corporation, Madison, WI) и AmpFlSTR® Control DNA 9947A (Life Technologies Corporation), а также 10 образцов ДНК индивидов из разных мировых популяций. Чувствительность набора ForenSeq DNA Signature Prep Kit оценивали путем изменения количества ДНК, взятой в анализ: от 1 нг до 50 пг в двух контрольных образцах. При использовании 1 нг были получены полные профили по всем SNP- и STR-маркерам, совпавшие с профилями контрольных образцов ДНК. С уменьшением количества ДНК (500, 250 и 100 пг) наблюдалось “выпадение” некоторых аллелей (0.5, 1.7 и 5.6% от всех аллелей соответственно). При использовании 50 пг ДНК были успешно генотипированы только 69.4% от всех маркеров. Таким образом была показана относительная нечувствительность к колебаниям количества ДНК в диапазоне от 1 нг до 100 пг. При проведении сравнительного анализа 10 образцов с помощью ForenSeq DNA Signature Prep Kit и наборов Identifiler Plus и Yfiler (Thermo Fisher Scientific) было показано полное совпадение аллелей, полученных для каждого из образцов двумя методами, за исключением двух локусов: при анализе методом капиллярного электрофореза наблюдалось выпадение аллелей в локусах DYS392 и DYS635, а при использовании МПС – в локусе DYS392 [40].

При анализе смешанных ДНК образцов изменения в гомо- и гетерозиготности маркеров обнаруживались, если второстепенная (минорная) ДНК присутствовала в образце даже на уровне 5%. Авторы [41] считают, что при числе прочтений менее 85000 на образец возможно увеличение количества неоднозначно определенных генотипов, что является достоверным наблюдением и может использоваться в качестве руководства в судебно-медицинской практике. Независимая проверка набора ForenSeq DNA Signature Prep Kit была проведена в работах Xavier et al. (набор праймеров A) [41] и Silvia et al. (набор праймеров B) [42]. В обоих исследованиях для всех стандартных контрольных образцов, использованных в работах, были получены полные профили STR-маркеров, за исключением двух случаев “выпадения” аллелей маркера Penta E при использовании смеси праймеров B [41]. Также в обеих работах профили всех STR-маркеров совпали с таковыми, полученными методом капиллярного электрофореза, за исключением ранее упомянутого маркера DXS10148 [20]. Заметим, что этот маркер позже был удален из панели.

При анализе чувствительности набора реагентов с использованием смеси праймеров A было выявлено “выпадение” аллеля для маркера DX10103 при использовании 250 пг ДНК. При использовании 50 пг для 93% STR-маркеров генотип образцов совпал с контрольным вариантом [42]. При этом были выявлены случаи отсутствия ампликонов (2.8%), низкого покрытия (0.6%), некорректное

определение генотипа было получено для 0.6% маркеров. Для маркера D22S1045 было отмечено уменьшение среднего числа прочтений по мере увеличения размера аллеля, что наблюдалось и в других исследованиях [21, 43, 44]. Предполагают, что причиной этому мотив повтора – последовательность ATT [45]. Также сообщалось о проблемах с покрытием маркера DYS392, имеющего аналогичный мотив повтора – ATA [43, 45]. Результаты генотипирования аутосомных SNP-маркеров совпадали в трех повторностях с покрытием от 1072 до 21 чтений на локус, с правильным и полным генотипированием при использовании более 100 пг матрицы ДНК [46]. В целом эти исследования показали, что некоторые аутосомные SNP-маркеры в этом наборе, включая rs6955448 [46, 47], rs7041158 [45] и rs2920816 [45], некорректно генотипируются и в дальнейшем они были исключены из набора реагентов.

С помощью набора ForenSeq DNA Signature Prep Kit также исследовали смеси мужской и женской ДНК, выделенной из буккальных соскобов [41]. Были исследованы смеси образцов буккального эпителия мужчины и женщины. SNP-маркеры мужчины были выявлены в количестве 3.3% (смесь мужского и женского образца 1 : 20), 4.4% (смесь 1 : 10) и 34.8% (смесь 1 : 1). Недопредставленность мужских SNP-маркеров в смесях связана с совпадением некоторых аллелей мужчины с женскими аллелями [41]. Silvia и др. проанализировали образцы ДНК, выделенные из пятен крови, слюны или спермы, с выстиранной или ношенной одежды из хлопка, нейлона и джинсовой ткани [42]. Для испытаний набора использовали 1 нг ДНК. При использовании прибора ForenSeq, 13 из 15 образцов дали полные STR-профили, совпадающие с контрольным типированием (AmpFISTR Identifiler PCR Amplification Kit Life Technologies/3130xl Genetic Analyzer). В шести образцах наблюдалось место выпадения маркера Penta E. Надо отметить, что длина ампликона Penta E составляет от 362 до 467 пн, он является одним из самых длинных STR-маркеров. В образце ДНК, полученной с ношенной футболки, также три маркера (CSF1PO, D16S539, D2S1338) не были типированы корректно ни одной системой. В образце ДНК с выстиранной футболки уже девять маркеров не были генотипированы (CSF1PO, D13S317, D16S539, D18S51, D2S1338, TH01, TPOX, D7S820a, vWAa). Вероятная причина этого – деградация ДНК, полученной из образцов исследуемого материала. Для 16 образцов были получены полные профили SNP-маркеров, соответствующие профилю ДНК донора. Один локус (rs1426654) в образце ДНК с ношенных джинс был по-разному генотипирован системой MiSeqFGx (по сравнению с контрольным типированием в фирме 23andMe), вероятно из-за того, что порог гетерозиготности в MiSeq FGx был установлен на уровне 25% [42].

Вторая лидирующая на современном рынке компания, выпускающая приборы МПС для криминалистических лабораторий, – это ThermoFisher Scientific. Эта компания разработала и выпустила целый ряд панелей Precision ID для использования на платформе Ion Torrent: панель идентификационных STR-маркеров, панель идентификационных SNP-маркеров, панель SNP-маркеров для определения этногеографического происхождения, панель SNP-маркеров для определения цвета радужной оболочки глаз и цвета волос, а также для определения последовательности митохондриальной ДНК, включая отдельную панель для анализа гипервариабельных регионов митохондриальной ДНК. За последние годы большое число публикаций было посвящено тестированию этих панелей [48–51]. Например, при работе с панелью маркеров, включающей 165 маркеров для определения происхождения индивида (Precision ID Ancestry Panel), все группы исследователей отмечали, что рекомендованное фирмой-производителем число циклов амплификации (21 цикл) недостаточно для получения полных профилей по включенным в состав набора маркерам в случае использовании малых количеств (менее 1 нг) ДНК – в таких условиях 30% локусов имеют покрытие менее 100 [48–51]. При этом увеличение циклов амплификации до 25 приводит к значительному улучшению результатов, хотя и повышает аллельный дисбаланс за счет предпочтительной амплификации более коротких фрагментов ДНК. Во всех исследованиях было отмечено низкое покрытие маркера rs1296819. Pereira и др. [48] пришли к выводу, что три SNP-маркера (rs7722456, rs459929 и rs7251928) работали настолько плохо, что их следует исключить из панели (см. табл. 3). По результатам описанных выше работ как в дизайн панелей Precision ID, так и в программное обеспечение был внесен ряд изменений [52–54].

В настоящее время основной набор для анализа STR-маркеров (панель Precision ID GlobalFiler NGS STR) включает все локусы расширенной панели CODIS, а также девять дополнительных аутосомных STR-локусов, четыре маркера пола, включая DYS391. Этот набор был протестирован на приборах Ion PGM [55] и Ion S5 [56]. Значительный дисбаланс в покрытии между разными маркерами все равно сохраняется. При этом Müller et al. [55] продемонстрировали, что дисбаланс воспроизводимо изменялся, когда панель использовалась в разных лабораториях. Wang et al. [57] предложили некоторые модификации, чтобы уменьшить разницу в уровнях покрытия маркеров.

При работе с панелями STR-маркеров наблюдаемые коэффициенты статтер-аллелей различались между исследованиями [55, 57]. Wang et al. сообщают о 8%-ном проскальзывании для всех (кроме трех) маркеров. В то время как Müller et al.

сообщают о более высоких коэффициентах – до 15–21% для некоторых маркеров (в обеих работах использовали 1 нг ДНК). Причина такой большой разницы между исследованиями не очевидна, хотя использование разных типов инструментов – одна из возможных причин. При этом в обоих исследованиях набор продемонстрировал хорошую чувствительность, и большинство аллелей детектировались и корректно определялись при количествах ДНК-матрицы ниже 100 пг [55, 57].

В качестве сравнительного анализа использования платформ IonTorrent и MiSeq в криминалистике можно рассматривать исследования панелей маркеров, разработанных различными фирмами. Так, панель идентификационных маркеров фирмы Qiagen, содержащая 140 аутосомных SNP, была протестирована с использованием платформ MiSeq [33] и Ion Torrent [58]. Было обнаружено, что из 140 маркеров один (rs1058083) дает стабильно низкое покрытие на обеих платформах во всех исследованиях, и в целом одни и те же SNP работают плохо во всех лабораториях, что указывает на низкую эффективность амплификации в начальной ПЦР. Покрытие, получаемое для некоторых маркеров, вероятно может зависеть от применяемой для секвенирования платформы. Так, покрытие маркера rs9951171 при секвенировании на платформе Ion Torrent в работе [58] было в 4 раза ниже, чем при использовании платформы MiSeq [33]. Типирование маркера rs1029047 оказалось проблематичным во всех исследованиях как с гомозиготными, так и с гетерозиготными генотипами; для него был выявлен значительный аллельный дисбаланс. Примечательно, что проблемы с генотипированием данного маркера также были описаны при его анализе с помощью панели от ThermoFisher [57]. Это, вероятно, связано с расположением вблизи маркера rs1029047 полиА тракта.

Аллельный дисбаланс также наблюдался во всех работах при анализе маркера rs2399332. Авторы интерпретируют этот результат по-разному: наличием полиморфизма в сайте связывания праймера [33], включением лишнего основания [47, 58], соседством с полиТ трактом, секвенирование которого приводит к ошибочной гетероплазмии в этом участке [57].

Некоторые SNP-маркеры также характеризуются менее выраженным аллельным дисбалансом (преимущественно в случае использования платформы MiSeq) или высокими показателями неправильного включения. К сожалению, на сегодняшний день нет единого стандарта параметров для анализа полученных данных. Один из примеров – порог гомозиготного генотипа, рекомендованный в трех различных исследованиях: 95% в работе [47], 90% [22] или 85% [49].

Помимо фирмы Qiagen, многие коммерческие компании, например такие как Promega (США) и NimaGene (Нидерланды), также выпустили панели для ДНК-идентификации с использованием технологии МПС.

Набор PowerSeq Auto Kit (Promega) включает 23 стандартных аутосомных STR-маркеров из системы CODIS, один Y-STR и амелогениновый маркер пола. При валидационных исследованиях PowerSeq на платформе MiSeq [59, 60], маркер D22S1045 не вызвал замечаний, в отличие от ранее упомянутых проблем в наборе ForenSeq DNA Signature Prep Kit (Verogen). Исследования чувствительности с помощью системы PowerSeq Auto демонстрируют “выпадение” аллелей только при использовании менее 62 пг ДНК [60]. Parson и др. провели эксперименты по оптимизации системы PowerSeq Auto/Y и указали на целый ряд усовершенствований лабораторного процесса путем удаления, модификации или автоматизация различных шагов [61].

Для использования дополнительных информативных полиморфных вариантов в STR-маркерах также требуется коррекция номенклатуры этих маркеров. Были предложены различные системы номенклатуры STR-аллелей, содержащих дополнительные SNP [17, 61, 62]; кроме того, комиссия ДНК Международного общества судебной генетики опубликовала ряд своих рекомендаций [63]. Тем не менее на сегодняшний день общепризнанных рекомендаций для номенклатуры STR еще не выпущено. Одна из причин этой задержки – недостаточный объем популяционных данных для создания единой номенклатуры. Эта проблема частично решается в последние годы – уже представлены результаты крупных популяционных исследований [20, 21, 64, 65] и разработана база STRseq для каталогизации вариантов STR-маркеров [20].

ТЕХНОЛОГИЧЕСКИЕ ПЕРСПЕКТИВЫ СЕКВЕНИРОВАНИЯ НА ДРУГИХ ПЛАТФОРМАХ

Компания MGI Tech Co., Ltd. (Шэньчжэнь, Китай) недавно запустила серию новых машин МПС, основанных на технологии комбинаторного синтеза зонд-якорь (cPAS, combinatorial Probe-Anchor Synthesis) “наношариков” ДНК (DNB, DNA Nanoballs) [66]. При подготовке библиотек для секвенирования ДНК расщепляют до фрагментов длиной около 500 пн, к ней лигируют специальные адаптеры, обеспечивающие последующую циркуляризацию фрагментов ДНК и амплификацию их по типу катящегося кольца с укладкой длинной цепи ДНК в компактные “наношарики”. Использование такого типа клональной амплификации исключает накопление ошибок ДНК-полимеразы в ходе ПЦР, так как матрицей в каждом цикле служит

только исходный фрагмент ДНК. Дальнейшее секвенирование происходит с использованием подхода комбинаторного синтеза зонд-якорь, при этом генерируются считывания парных концов длиной 50–150 нуклеотидов.

На сегодняшний день опубликованы лишь единичные работы, представляющие использование платформы MGI при проведении судебно-медицинской экспертизы. Первая работа была посвящена определению отцовства с использованием SNP-маркеров в исследовании Chang и др. [67].

Совсем недавно компания MGIEasy (MGI Tech, Шэньчжэнь, Китай) разработала криминалистическую панель для использования на платформе MGI [68]. Указанная панель позволяет проводить одновременный анализ 362 генетических маркеров, включая маркеры ядерного и митохондриального геномов. Панель включает маркеры пола, 52 аутосомных STR-маркера, 27 STR-маркеров X-хромосомы, 48 STR-маркеров Y-хромосомы и 227 SNP-маркеров для идентификации личности, определения вероятного этногеографического происхождения и вероятной внешности, а также позволяет анализировать все гипервариабельные регионы мтДНК. В качестве преимущества системы MGIEasy разработчики выделяют малый размер ампликонов для маркеров, входящих в панель (около 150 пн), при этом для 210 SNP-маркеров он составляет меньше 90 пн. В работе Ran Li представлены первые результаты валидации этого набора и самой платформы MGI для криминалистики [68].

Хотя секвенирование с помощью технологии синтеза (SBS) в настоящее время все же доминирует в области МПС, предпринимается множество попыток ввести другие подходы и преодолеть недостатки секвенирования 2-го поколения [2].

В настоящее время наиболее перспективной технологией является нанопоровое секвенирование, основанное на детекции изменения ионного тока при прохождении молекулы ДНК через белковую пору. Эта технология представляет собой секвенирование в реальном времени с возможностью определять последовательность ДНК длиной несколько сотен тысяч нуклеотидов.

Недавно был представлен портативный секвенатор на основе нанопор MinION (Oxford Nanopore Technologies). Cornelis et al. впервые исследовали возможность его использования в криминалистике для анализа SNP-маркеров [69]. Авторы проанализировали контрольный образец 9947 A (Promega, Madison, US) по 52 SNP-маркерам, разработанным консорциумом SNPforID [27], и сопоставили данные с генотипами, полученными с помощью секвенирования на платформе Illumina. Для анализа использовали 2.5 нг ДНК, полученные ампликоны лигировали друг с другом случайным образом для создания более длинных фрагментов ДНК, необходимых для нанопорового секвенирования. Среднее полученное в результате секвенирования покрытие локуса составило 17933, что достаточно для генотипирования SNP-маркеров. Для маркера rs1029047 было получено наименьшее количество чтений, вероятно из-за расположения в участке поли-А тракта. Об аналогичных трудностях секвенирования гомополимерных областей также сообщали в работе Loman et al. [70]. Для двух локусов SNP (rs143232 и rs1031825), один из которых соседствует, а второй находится внутри гомополимерного участка (см. табл. 3), наблюдался значительный аллельный дисбаланс. В результате анализа при помощи пакета программ Metrichor service, разработанного Oxford Nanopore Technologies, генотип для маркера rs1031825 в гомозиготном образце был неправильно определен как гетерозиготный. Очевидно, при разработке панелей маркеров, ориентированных на систему MinION, следует особенно строго избегать гомополимерных участков.

В настоящий момент к основным недостаткам системы MinION относятся высокая частота ошибок при секвенировании (до 15%, по сравнению с 0.5% у Illumina) и несовместимость с существующими криминалистическими базами данных, содержащими в основном STR-профили индивидов [71, 72], что однако относится и к SNP-панелям для других платформ. Тем не менее недавние исследования [72, 73], в которых проверялась применимость этой технологии для идентификации ДНК, демонстрируют хорошие перспективы для использования системы MinION в судебно-медицинских исследованиях. Так, Zaaijer et al. [72] разработали специальный модуль для биоинформационной обработки данных, полученных с помощью MinION, и показали, что этот подход позволяет достичь совпадение профилей ДНК для 91–195 SNP с точностью 99.9%, при этом сам процесс секвенирования занимает всего 3 минуты. Хотя подход, использованный в этом исследовании, пока не совместим с существующими методологиями судебной экспертизы и базами данных, он демонстрирует прекрасный потенциал для оперативной ДНК-идентификации в полевых условиях, когда доступ к стандартным технологиям профилирования ДНК ограничен, а также в критически ограниченных по времени случаях, когда надо повторно идентифицировать неизвестного индивида или сделать анализ близкого родства у жертв катастроф непосредственно на месте происшедшего [73].

ПЕРСПЕКТИВЫ И НАПРАВЛЕНИЯ РАЗВИТИЯ

К преимуществам использования МПС можно отнести возможность анализировать и другие типы нуклеотидных последовательностей. На-

пример, анализ статуса метилирования определенных регионов ДНК для определения возраста индивида [74], использование РНК-маркеров для идентификации жидкостей организма [75]. Интересно, что уже апробируются для криминалистических целей недавно разработанные протоколы МПС, которые позволяют использовать образцы с малым количеством геномного материала для одновременного анализа РНК и ДНК. Использование новых видов маркеров актуально для особенных случаев идентификации, например дифференцирования одного из монозиготных близнецов [76].

В связи с тем, что современные генетические технологии уже широко применяются в криминалистике, возникла необходимость создания на основе результатов генетических экспертиз судебных баз данных. На данный момент уже более 60 стран мира создали национальные базы данных судебно-медицинской экспертизы ДНК на основе STR-маркеров, среди которых, например, общепринятая база данных CODIS и Национальная база данных ДНК Великобритании. Количество и объемы информации, содержащиеся в этих базах данных, продолжают быстро расти. Например, национальная база данных судебной экспертизы Китая насчитывает более 27 миллионов записей об индивидуальных STR-профилях. Совместить уже имеющиеся базы данных и профили ДНК, полученные с использованием МПС, – задача, которую необходимо решить. Разработаны программы и веб-приложения, чтобы помочь судебным экспертам просто и эффективно работать с данными MPS, особенно при типировании STR маркеров [77, 78]. Разрабатываются алгоритмы, которые генерируют имена для секвенированных аллелей STR на основе старой номенклатуры, но с учетом новых SNP, в соответствии с рекомендациями Международного общества судебной генетики (ISFG) [61], например программа STRNaming [79].

Кроме того, уже сейчас возникает необходимость внесения изменений в международные и национальные законодательства с целью хранения и использования в судебной практике генетических данных, определенных с помощью как традиционных методов судебно-генетического анализа, так и новых видов геномных данных, полученных с помощью методов и платформ МПС [80, 81]. В Российской Федерации в настоящее время рассматриваются изменения закона “О государственной геномной регистрации в Российской Федерации”, с целью формирования федеральной базы данных геномной информации (ФБДГИ) для более эффективного использования данных полного генома для раскрытия и предотвращения преступления, в том числе особо тяжких (террористических актов, убийств и изнасилований) [82].

Еще одна немаловажная проблема, стоящая на пути массового внедрения МПС в область судебной медицины и криминалистики, – необходимость подготовки большого количества профильных специалистов высокого уровня.

ЗАКЛЮЧЕНИЕ

На сегодняшний день среди платформ для масштабного параллельного секвенирования приборы Illumina имеют значительные преимущества перед другими платформами благодаря высокой точности анализа, большой пропускной способности, а также широкого диапазона доступных коммерческих панелей. Тем не менее высокая степень кросс-платформенной совместимости позволяет предположить, что в будущем возможна успешная адаптация разработанных криминалистических панелей для использования на любых платформах для масштабного параллельного секвенирования.

Технологии секвенирования ДНК совершенствуются очень быстрыми темпами: модернизируются новые методы МПС (секвенирование 3-го поколения), разрабатываются полностью автоматизированные системы (конвейеры), в которых весь процесс подготовки образцов и секвенирования происходит без вмешательства человека. Очевидно, что уже в ближайшем будущем платформы МПС будут использоваться при проведении рутинных судебно-медицинских исследований ДНК.

Работа поддержана Государственным контрактом № 011-17 от 26.09.2017 г.

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. *Margulies M., Egholm M., Altman W.E. et al.* Genome sequencing in microfabricated high-density picolitre reactors // Nature. 2005. V. 437(7057). P. 376–380. https://doi.org/10.1038/nature03959
2. *Goodwin S., McPherson J.D., McCombie W.R.* Coming of age: Ten years of next-generation sequencing technologies // Nat. Rev. Genet. 2016. V. 17. № 6. P. 333–351. https://doi.org/10.1038/nrg.2016.49
3. https://dnadatabank.forensischinstituut.nl/binaries/dna-jaarverslag-2015_tcm37-87649.pdf
4. *Iozzi S., Carboni I., Contini E. et al.* Forensic genetics in NGS era: New frontiers for massively parallel typing // Forensic Sci. Int. Genet. Suppl. Ser. 2015. V. 5. P. e418–e419. https://doi.org/10.1016/j.fsigss.2015.09.166

5. *Alonso A., Müller P., Roewer L. et al.* European survey on forensic applications of massively parallel sequencing // Forensic Sci. Int. Genet. 2017. V. 29. P. 23–25. https://doi.org/10.1016/j.fsigen.2017.04.017
6. *Fuller C.W., Middendorf L.R., Benner S.A. et al.* The challenges of sequencing by synthesis // Nat. Biotechnol. 2009. V. 27(11). P. 1013–1023. https://doi.org/10.1038/nbt.1585
7. *Rhoads A., Au K.F.* PacBio sequencing and its applications // Genomics Proteomics Bioinformatics. 2015. V. 13(5). P. 278–289. https://doi.org/10.1016/j.gpb.2015.08.002
8. *Jain M., Fiddes I.T., Miga K.H. et al.* Improved data analysis for the MinIONnanopore sequencer // Nat. Methods. 2015. V. 12(4). P. 351–356. https://doi.org/10.1038/nmeth.3290
9. *Rothberg J.M., Hinz W., Rearick T.M. et al.* An integrated semiconductor device enabling non-optical genome sequencing // Nature. 2011. V. 475(7356). P. 34852. https://doi.org/10.1038/nature10242
10. *Bentley D.R., Balasubramanian S., Swerdlow H.P. et al.* Accurate whole human genome sequencing using reversible terminator chemistry // Nature. 2008. V. 456. P. 53–59. https://doi.org/10.1038/nature07517
11. *Zhao X., Li H., Wang Z. et al.* Massively parallel sequencing of 10 autosomal STRs in Chinese using the ion torrent personal genome machine (PGM) // Forensic Sci. Int. Genet. 2016. V. 25. P. 3438. https://doi.org/10.1016/j.fsigen.2016.07.014
12. *Quail M.A., Smith M., Couplnad P. et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers // BMC Genomics. 2012. V. 13. P. 341. https://doi.org/10.1186/1471-2164-13-341
13. *Hollard C., Ausset L., Chantrel Y. et al.* Automation and developmental validation of the ForenSeq™ DNA Signature Preparation kit for high-throughput analysis in forensic laboratories // Forensic Sci. Int. Genet. 2019. V. 40. P. 37–45. https://doi.org/10.1016/j.fsigen.2019.01.010
14. *Jeffreys A.J., Wilson V., Thein S.L.* Hypervariable "minisatellite" regions in human DNA // Nature. 1985. V. 314. P. 67–73. https://doi.org/10.1038/314067a0
15. *Rogaev E.I.* Two novel human DNA tandem repeat families from the hypervariable DNA probe homologous to human apolipoprotein CII-gene intron and *D. virilis* satellite // Nucl. Acids. Res. 1989. V. 17. P. 1246. https://doi.org/10.1093/nar/17.3.1246
16. *Перепечина И.О.* Обмен информацией о результатах криминалистического исследования ДНК: международный опыт // Сб. матер. 54-х криминалистических чтений "Судебная 88 экспертиза в парадигме российской науки (к 85-летию Ю.Г. Корухова)": В 2-х ч. М.: Акад. управления МВД России, 2013. P. 140–146.
17. *Gelardi C., Rockenbauer E., Dalsgaard S. et al.* Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles // Forensic Sci. Int. Genet. 2014. V. 12. P. 38–41. https://doi.org/10.1016/j.fsigen.2014.04.016
18. *Gettings K.B., Aponte R.A., Vallone P.M. et al.* STR allele sequence variation: Current knowledge and future issues // Forensic Sci. Int. Genet. 2015. V. 18. P. 118–130. https://doi.org/10.1016/j.fsigen.2015.06.005
19. *Devesse L., Ballard D., Davenport L. et al.* Concordance of the ForenSeq system and characterisation of sequence-specific autosomal STR alleles across two major population groups // Forensic Sci. Int. Genet. 2017. V. 34. P. 57–61. https://doi.org/10.1016/j.fsigen.2017.10.012
20. *Phillips C., Gettings K.B., King J.L. et al.* "The devil's in the detail": release of an expanded, enhanced and dynamically revised forensic STR sequence guide // Forensic Sci. Int. Genet. 2018. V. 34. P. 162–169. https://doi.org/10.1016/j.fsigen.2018.02.017
21. *Gettings K.B., Borsuk L.A., Ballard D. et al.* STRSeq: A catalog of sequence diversity at human identification short tandem repeat loci // Forensic Sci. Int. Genet. 2017. V. 31. P. 111–117. https://doi.org/10.1016/j.fsigen.2017.08.017
22. *van der Gaag K.J., de Leeuw R.H., Hoogenboom J. et al.* Massively parallel sequencing of short tandem repeats – population data and mixture analysis results for the PowerSeq system // Forensic Sci. Int. Genet. 2016. V. 24. P. 86–96. https://doi.org/10.1016/j.fsigen.2016.05.016
23. *Phillips C., Fang R., Ballard D. et al.* Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel // Forensic Sci. Int. Genet. 2007. V. 1(2). P. 180–185. https://doi.org/10.1016/j.fsigen.2007.02.007
24. *Tozzo P., Gabbin A., Politi C. et al.* Combined statistical analyses of forensic evidence in sexual assault: A case report and brief review of the literature // J. Forensic Sci. 2020. V. 65. P. 1767–1773. https://doi.org/10.1111/1556-4029.14487
25. *Amorim A., Pereira L.* Pros and cons in the use of SNPs in forensic kinship investigation: A comparative analysis with STRs // Forensic Sci. Int. 2005. V. 150. № 1. P. 17–21. https://doi.org/10.1016/j.forsciint.2004.06.018
26. *Dixon L.A., Murray C.M., Archer E.A. et al.* Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes // Forensic Sci. Int. 2005. V. 154. № 1. P. 62–77. https://doi.org/10.1016/j.forsciint.2004.12.011
27. *Sanchez 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
28. *Dario P., Oliveira A.R., Ribeiro T. et al.* SNPforID 52-plex in casework samples: "Cracking" bones and other difficult samples // Forensic Sci. Int. Genet. Suppl. Ser. 2015. V. 5. P. e118–e120. https://doi.org/10.1016/j.fsigss.2015.09.048
29. *Kidd K.K., Speed W.C., Pakstis A.J. et al.* Progress toward an efficient panel of SNPs for ancestry inference // Fo-

rensic Sci. Int. Genet. 2014. V. 10. P. 23–32. https://doi.org/10.1016/j.fsigen.2014.01.002

30. *Walsh S., Liu F., Wollstein A. et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA // Forensic Sci. Int. Genet. 2013. V. 7. № 1. P. 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005
31. *Kidd K.K., Pakstis A.J., Speed W.C. et al.* Developing a SNP panel for forensic identification of individuals // Forensic Sci. Int. 2006. V. 164. № 1. P. 20–32. https://doi.org/10.1016/j.forsciint.2005.11.017
32. *Daniel R., Santos C., Phillips C. et al.* A SNaPshot of next generation sequencing for forensic SNP analysis // Forensic Sci. Int. Genet. 2015. V. 14. P. 50–60. https://doi.org/10.1016/j.fsigen.2014.08.013
33. *Grandell I., Samara R., Tillmar A.O.* A SNP panel for identity and kinship testing using massive parallel sequencing // Int. J. Legal. Med. 2016. V. 130. № 4. P. 90514. https://doi.org/10.1007/s00414-016-1341-4
34. *Kidd K.K., Pakstis A.J., Speed W.C. et al.* Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics // Forensic Sci. Int. Genet. 2014. V. 12. P. 215–224. https://doi.org/10.1016/j.fsigen.2014.06.014
35. *Pakstis A.J., Fang R., Furtado M.R. et al.* Mini-haplotypes as lineage informative SNPs and ancestry inference SNPs // Europ. J. Hum. Genet. 2012. V. 20. № 11. P. 1148–1154. https://doi.org/10.1038/ejhg.2012.69
36. *Wendt F.R., King J.L., Novroski N.M. et al.* Flanking region variation of ForenSeq DNA signature Prep Kit STR and SNP loci in Yavapai native Americans // Forensic Sci. Int. Genet. 2017. V. 28. P. 146–154. https://doi.org/10.1016/j.fsigen.2017.02.014
37. *Pang J.B., Rao M., Chen Q.F. et al.* A 124-plex microhaplotype panel based on next-generation sequencing developed for forensic applications // Sci. Reports. 2020. V. 10. 1945. https://doi.org/10.1038/s41598-020-58980-x
38. *Bulbul O., Pakstis J., Soundararajan U. et al.* Ancestry inference of 96 population samples using microhaplotypes // Int. J. Legal Med. 2018. V. 132. № 3. P. 703–711. https://doi.org/10.1007/s00414-017-1748-6
39. *Jager A.C., Alvarez M.L., Davis C.P. et al.* Developmental validation of the MiSeqFGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories // Forensic Sci. Int. Genet. 2017. V. 28. P. 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011
40. *Churchill J.D., Schmedes S.E., King J.L., Budowle B.* Evaluation of the Illumina(®) Beta version ForenSeq DNA Signature Prep Kit for use in genetic profiling // Forensic Sci. Int. Genet. 2016. V. 20. P. 20–29. https://doi.org/10.1016/j.fsigen.2015.09.009
41. *Xavier C., Parson W.* Evaluation of the Illumina ForenSeq DNA Signature Prep kit – MPS forensic application for the MiSeqFGx benchtop sequencer // Forensic Sci. Int. Genet. 2017. V. 28. P. 188–194. https://doi.org/10.1016/j.fsigen.2017.02.018
42. *Silvia A.L., Shugarts N., Smith J.* A preliminary assessment of the ForenSeqFGx system: Next generation sequencing of an STR and SNP multiplex // Int. J. Legal Med. 2017. V. 131. № 1. P. 73–86. https://doi.org/10.1007/s00414-016-1457-6
43. *Eduardoff M., Gross T.E., Santos C. et al.* Inter-laboratory evaluation of the EUROFORGEN Global ancestry-informative SNP panel by massively parallel sequencing using the Ion PGM // Forensic Sci. Int. Genet. 2016. V. 23. P. 178–189. https://doi.org/10.1016/j.fsigen.2016.04.008
44. *Parson W., Strobl C., Huber G. et al.* Evaluation of next generation mtGenome sequencing using the ion torrent personal genome machine (PGM) // Forensic Sci. Int. Genet. 2013. V. 7. № 5. P. 543–549. https://doi.org/10.1016/j.fsigen.2013.06.003
45. *Almalki N., Chow H.Y., Sharma V. et al.* Systematic assessment of the performance of Illumina's MiSeqFGx forensic genomics system // Electrophoresis. 2017. V. 38. № 6. P. 846–854. https://doi.org/10.1002/elps.201600511
46. *Churchill J.D., Novroski N.M.M., King J.L. et al.* Population and performance analyses of four major populations with Illumina's FGx forensic genomics system // Forensic Sci. Int. Genet. 2017. V. 30. P. 81–92. https://doi.org/10.1016/j.fsigen.2017.06.004
47. *Just R.S., Moreno L.I., Smerick J.B., Irwin J.A.* Performance and concordance of the ForenSeq system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens // Forensic Sci. Int. Genet. 2017. V. 28. P. 1–9. https://doi.org/10.1016/j.fsigen.2017.01.001
48. *Pereira V., Mogensen H.S., Borsting C., Morling N.* Evaluation of the precision ID ancestry panel for crime case work: A SNP typing assay developed for typing of 165 ancestral informative markers // Forensic Sci. Int. Genet. 2017. V. 28. P. 138–145. https://doi.org/10.1016/j.fsigen.2017.02.013
49. *Al-Asfi M., McNevin D., Mehta B. et al.* Assessment of the precision ID ancestry panel // Int. J. Legal Med. 2018. V. 132(6). P. 1581–1594. https://doi.org/10.1007/s00414-018-1785-9
50. *Jin S., Chase M., Henry M. et al.* Implementing a biogeographic ancestry inference service for forensic casework // Electrophoresis. 2018. V. 39. № 21. P. 2757–2765. https://doi.org/10.1002/elps.201800171
51. *Fordyce S.L., Mogensen H.S., Borsting C. et al.* Second-generation sequencing of forensic STRs using the Ion Torrent HID STR 10-plex and the Ion PGM // Forensic Sci. Int. Genet. 2015. V. 14. P. 132–140. https://doi.org/10.1016/j.fsigen.2014.09.020
52. *Guo F., Zhou Y., Liu F. et al.* Evaluation of the early access STR kit v1 on the Ion Torrent PGM platform // Forensic Sci. Int. Genet. 2016. V. 23. P. 111–120. https://doi.org/10.1016/j.fsigen.2016.04.004
53. *Kulstein G., Pably P., Furst A. et al.* "The acid test" – validation of the ParaDNA® Body Fluid ID Test forroutine forensic casework // Int. J. Legal Med. 2019. V. 133. № 3. P. 751–757. https://doi.org/10.1016/j.fsigen.2014.09.020

54. *Novroski N.M.M., King J.L., Churchill J.D. et al.* Characterization of genetic sequence variation of 58 STR loci in four major population groups // Forensic Sci. Int. Genet. 2016. V. 25. P. 214–226. https://doi.org/10.1016/j.fsigen.2016.09.007

55. *Müller P., Alonso A., Barrio P.A. et al.* Systematic evaluation of the early access applied biosystems precision ID Globalfiler mixture ID and Globalfiler NGS STR panels for the ion S5 system // Forensic Sci. Int. Genet. 2018. V. 36. P. 95–103. https://doi.org/10.1016/j.fsigen.2018.06.016

56. *Gettings K.B., Kiesler K.M., Faith S.A. et al.* Sequence variation of 22 autosomal STR loci detected by next generation sequencing // Forensic Sci. Int. Genet. 2016. V. 21. P. 15–20. https://doi.org/10.1016/j.fsigen.2015.11.005

57. *Wang Z., Zhou D., Wang H. et al.* Massively parallel sequencing of 32 forensic markers using the precision ID GlobalFiler NGS STR panel and the ion PGM system // Forensic Sci. Int. Genet. 2017. V. 31. P. 126–134. https://doi.org/10.1016/j.fsigen.2017.09.004

58. *de la Puente M., Phillips C., Santos C. et al.* Evaluation of the Qiagen 140-SNP forensic identification multiplex for massively parallel sequencing // Forensic Sci. Int. Genet. 2017. V. 28. P. 35–43. https://doi.org/10.1016/j.fsigen.2017.01.012

59. *Zeng X., King J., Hermanson S. et al.* An evaluation of the PowerSeq Auto System: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing // Forensic Sci. Int. Genet. 2015. V. 19. P. 172–179. https://doi.org/10.1016/j.fsigen.2015.07.015

60. *Montano E.A., Bush J.M., Garver A.M. et al.* Optimization of the PromegaPowerSeq Auto/Y system for efficient integration within a forensic DNA laboratory // Forensic Sci. Int. Genet. 2018. V. 32. P. 26–32. https://doi.org/10.1016/j.fsigen.2017.10.002

61. *Parson W., Ballard D., Budowle B. et al.* Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements // Forensic Sci. Int. Genet. 2016. V. 22. P. 54–63. https://doi.org/10.1016/j.fsigen.2016.01.009

62. *Huszar T.I., Jobling M.A., Wetton J.H.* A phylogenetic framework facilitates Y-STR variant discovery and classification via massively parallel sequencing // Forensic Sci. Int. Genet. 2018. V. 35. P. 97–106. https://doi.org/10.1016/j.fsigen.2018.03.012

63. *Gettings K., Ballard D., Bodner M. et al.* Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting // Forensic Sci. Int. Genet. 2019. V. 43. P. 10216. https://doi.org/10.1016/j.fsigen.2019.102165

64. *Gettings K.B., Borsuk L.A., Steffen C.R. et al.* Sequence-based U.S. population data for 27 autosomal STR loci // Forensic Sci. Int. Genet. 2018. V. 37. P. 106–115. https://doi.org/10.1016/j.fsigen.2018.07.013

65. *Gunnarsdottir E.D., Li M., Bauchet M.F. et al.* High-throughput sequencing of complete human mtDNA genomes from the Philippines // Genome Res. 2011. V. 21. № 1. P. 1–11. https://doi.org/10.1101/gr.107615.110

66. *Drmanac R., Sparks A.B., Callow M.J. et al.* Human genome sequencing using unchained base reads on Self-Assembling DNA nanoarrays // Science. 2010. V. 327. iss. 5961. P. 78–81. https://doi.org/10.1126/science.1181498

67. *Chang L., Yu H., Miao X. et al.* Development and comprehensive evaluation of a noninvasive prenatal paternity testing method through a scaled trial // Forensic Sci. Int. Genet. 2019. V. 43. P. 102158. https://doi.org/10.1016/j.fsigen.2019.102158

68. *Ran Li., Xuefeng Shen, Hui Chen et al.* Developmental validation of the MGIEasy Signature Identification Library Prep Kit, An all-in-one multiplex system for forensic applications // Int. J. Legal Med. 2021. V. 135. P. 739–753. https://doi.org/10.1007/s00414-021-02507-0

69. *Cornelis S., Gansemans Y., Deleye L. et al.* Forensic SNP genotyping using nanopore MinION sequencing // Sci. Reports. 2017. V. 7. P. 41759. https://doi.org/10.1038/srep41759

70. *Loman N.J., Quick J., Simpson J.T.* A complete bacterial genome assembled de novo using only nanopore sequencing data // Nat. Methods. 2015. V. 12. № 8. P. 733–735. https://doi.org/10.1038/nmeth.3444

71. *Kchouk M., Gibrat J.-F., Elloumi M.* Generations of sequencing technologies: From first to next generation // Biol. Med. 2017. V. 9. P. 395. https://doi.org/10.4172/0974-8369.1000395

72. *Zaaijer S., Gordon A., Speyer D. et al.* Rapid re-identification of human samples using portable DNA sequencing // eLife. 2017. V. 6. P. e27798. https://doi.org/10.7554/eLife.27798

73. *Zubakov D., Kokmeijer I., Ralf A. et al.* Towards simultaneous individual and tissue identification: A proof-of-principle study on parallel sequencing of STRs, amelogenin, and mRNAs with the Ion Torrent PGM // Forensic Sci. Int. Genet. 2015. V. 17. P. 122–128. https://doi.org/10.1016/j.fsigen.2015.04.002

74. *Vidaki A., Daniel B., Syndercombe Court D.* Forensic DNA methylation profiling – potential opportunities and challenges // Forensic Sci. Int. Genet.2013. V. 7. P. 499–507. https://doi.org/10.1016/j.fsigen.2013.05.004

75. *Albani P.P., Fleming R.* Novel messenger RNAs for body fluid identification // Sci. Justice. 2018. V. 58. P. 145–152. https://doi.org/10.1016/j.scijus.2017.09.002

76. *Turrina S., Bortoletto E., Giannini G., De Leo D.* Monozygotic twins: Identical or distinguishable for science and law? // Med Sci Law. 2021. V. 61. № 1. P. 62–66. https://doi.org/10.1177/0025802420922335

77. *Ganschow S., Silvery J., Kalinowski J., Tiemann C.* toaSTR: A web application for forensic STR genotyping by massively parallel sequencing // Forensic Sci. Int. Genet. 2018. V. 37. P. 21–28. https://doi.org/10.1016/j.fsigen.2018.07.006

78. *Hoogenboom J., van der Gaag K., de Leeuw R. et al.* FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing

noise // Forensic Sci. Int. Genet. 2017. V. 27. P. 27–40. https://doi.org/10.1016/j.fsigen.2016.11.007

79. *Hoogenboom J., TitiaSijen T., van der Gaag K.* STRNaming: Generating simple, informative names for sequenced STR alleles in a standardised and automated manner // Forensic Sci. Int. Genet. 2021. V. 52. P. 102473. https://doi.org/10.1016/j.fsigen.2021.102473
80. *Чемерис А., Анисимов В., Аминев Ф.* Актуальные вопросы правового регулирования, создания и функционирования баз данных генома человека // Правовое государство: теория и практика. 2020. Т. 16. № 82. С. 134–144. https://doi.org/10.33184/pravgos-2020.2.12
81. *Vladimirova D.* Analysis of judicial practice in the field of genomic and genetic research including legal positions of the ECHR and practice of Russian courts // Ross. Pravo. Obraz. Prakt. Nauk. 2020. № 5. P. 16–25. https://doi.org/10.34076/2410-2709-2020-5-16-25
82. https://sozd.duma.gov.ru/bill/1048800-7

Using NGS in Forensics: A Comparative Analysis of Sequencing Platforms

T. V. Tyazhelova[a, *], I. L. Kuznetsova[a], T. V. Andreeva[a, b], S. S. Kunizheva[a, b], and E. I. Rogaev[a, b, c]

[a] *Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, 119991 Russia*

[b] *Lomonosov Moscow State University, Moscow, 119991 Russia*

[c] *"Sirius" University of Science and Technology, Krasnodarskii kray, Sochi, 354340 Russia*

**e-mail: tatvltyz@gmail.com*

The last decade has been characterized by a significant breakthrough in the development of new technologies and tools for DNA analysis. A new technology of massive parallel sequencing, otherwise known as "next-generation sequencing" (NGS) is by now already widely used in many areas of biological and biomedical research. The introduction of new parallel sequencing technologies into the forensic field is slower compared to scientific and clinical research, mainly due to the difficulties of compatibility with already existing DNA profile databases, as well as the need for rigorous validation experiments for accreditation of instruments and assays sets of reagents. This review presents the worldwide existing solutions for the use of massively parallel sequencing in forensics, technical aspects of experiments, developed commercial solutions and application prospects, as well as the advantages and disadvantages of new sequencing technologies compared to the methods of DNA analysis traditionally used in forensic science.

Keywords: massive parallel sequencing (MPS), next generation sequencing (NGS), short tandem repeat (STR), single nucleotide polymorphism (SNP), forensic science.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1438–1449

УДК 575.17:599.9

ГЕНЕТИКО-ДЕМОГРАФИЧЕСКИЕ ПАРАМЕТРЫ НАСЕЛЕНИЯ МОСКВЫ. МИГРАЦИОННЫЕ ПРОЦЕССЫ

© 2021 г. О. Л. Курбатова[1], *, А. С. Грачева[1], **, Е. Ю. Победоносцева[1], И. Г. Удина[1]

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

**e-mail: okurbat@list.ru*

***e-mail: palesa@yandex.ru*

Поступила в редакцию 19.05.2021 г.
После доработки 07.06.2021 г.
Принята к публикации 08.06.2021 г.

Определение основных генетико-демографических параметров миграционных процессов, значимых для изучения структуры генофонда населения Москвы, базируется на анализе материалов демографической статистики, Всероссийских переписей населения и данных выборочного анкетирования двух групп жителей Москвы, различающихся по возрастным характеристикам: “старшая” группа в среднем 1948 г. рождения, “младшая” – в среднем 1987 г. рождения. Показано значительное увеличение интенсивности и дальности миграции в Москву за последние 40 лет. Коэффициент миграции, рассчитанный по данным анкетирования “старшей” группы москвичей, равен 0.413 для мужчин и 0.517 для женщин, средний радиус миграции – 1008 км; по данным анкетирования “младшей” группы – коэффициент миграции 0.636 для мужчин и 0.657 для женщин, средний радиус миграции – 1782 км. Выявлены гендерные особенности миграционных процессов в Москве, которые обусловливают значительные различия половозрастного состава городских жителей различных национальностей (численное преобладание мужчин среди представителей “молодых” этнических диаспор). Данные об источниках формирования генофонда населения Москвы на современном этапе показывают, что доля родившихся в России в “старшей” возрастной группе (89%) намного больше, чем в “младшей” (76%). В первом случае генофонд формируется на 3/4 за счет населения Центрального федерального округа (ЦФО), заметен вклад Приволжского федерального округа и Украины; во втором случае в ЦФО родилась лишь половина анкетируемых, много уроженцев Приволжского, Северо-Кавказского, Южного и Сибирского округов, республик Средней Азии, Украины, Армении. Данные анкетирования подтверждаются анализом данных миграционной статистики Росстата о составе мигрантов в Москву по странам выхода. Различия параметров миграции в двух возрастных группах указывают на возможность динамики этнорегионального состава населения Москвы в последующих поколениях, что непременно вызовет изменение частот многих генетических маркеров, в том числе частотных профилей маркеров ДНК-идентификации. Выявленные в настоящем исследовании особенности миграционных процессов в Москве указывают на необходимость своевременного обновления и актуализации генетических баз данных для целей ДНК-идентификации в мегаполисе.

Ключевые слова: демографическая генетика, мегаполисы, Москва, анкетирование, миграционная статистика, коэффициент миграции, миграционное расстояние, этнический состав, динамика генофонда, базы данных для целей ДНК-идентификации.

DOI: 10.31857/S0016675821120080

Демографической генетике Москвы посвящены многочисленные исследования, проводимые в ИОГен РАН [1–14]. Приток мигрантов всегда был основным, а по большей части, единственным источником роста численности населения Москвы. Население Москвы росло особенно быстрыми темпами, начиная со второй половины XIX в. (рис. 1), благодаря резкому увеличению притока мигрантов после реформы 1861 г., освободившей крестьян от крепостной зависимости. Перепись 1897 г. зафиксировала, что численность населения Москвы превысила 1 млн чел. (1039 тыс.) – с этой даты Москву можно считать мегаполисом [15].

Основные генетически значимые параметры миграционных процессов в Москве изучены нами в многолетней динамике – начиная с середины XIX в. Коэффициенты миграции, рассчитанные по данным переписей и церковно-приходских книг, на рубеже XIX и XX вв. были чрезвычайно высокими: $0.7 < m < 0.8$ [9] и достигли максимума в годы Первой Мировой войны. В середине XX в. коэффициент миграции был почти такой же

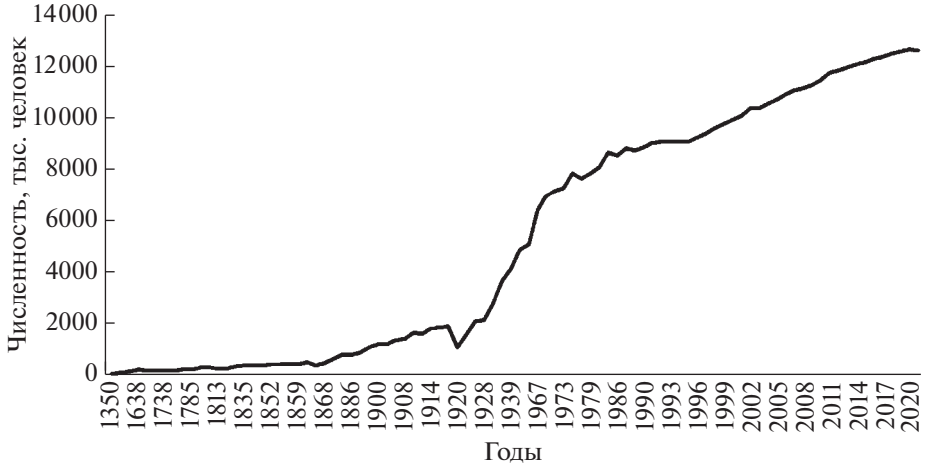

Рис. 1. Динамика численности населения Москвы.

(0.76), как и на рубеже XIX и XX вв., и до конца XX столетия колебался вблизи значений $m = 0.4–0.5$ [1, 7].

За столетний период заметно выросла дальность перемещения мигрантов – на рубеже XIX и XX вв. средний радиус миграции $\bar{R}$ составлял 230 км (наибольший вклад в генофонд москвичей вносили уроженцы самой Москвы, Московской губернии и близлежащих губерний Нечерноземья). В начале XXI в. данный показатель превысил 1000 км (увеличились потоки мигрантов из южных регионов РФ и стран “ближнего зарубежья”) [8, 13, 14]. Параллельно с расширением области миграционного притяжения Москвы происходит и увеличение разнообразия этнического состава населения.

Проведенный анализ свидетельствует о том, что интенсивность миграции и этнорегиональный состав мигрантов подвержены заметным и не всегда предсказуемым изменениям во времени. В этой связи решено было продолжить изучение динамики основных генетически значимых параметров миграции на современном этапе.

В настоящее время Москва – самый крупный мегаполис и столица Российской Федерации с численностью населения 12636.3 тыс. чел. (на 1 января 2021 г.) [16]. По “среднему варианту” прогноза Росстата до 2036 г. естественный прирост населения Москвы будет иметь отрицательное значение (естественная убыль), а миграционный прирост будет колебаться на уровне 100 тыс. чел. в год [16].

В связи с тем, что многие, ранее использованные нами источники демографической информации стали теперь недоступными, настоящее исследование основано в основном на данных анкетирования москвичей с привлечением материалов переписей населения, которые остаются единственными источниками данных о национальности.

Цель настоящего исследования – оценка основных генетико-демографических параметров миграции населения Москвы в первые десятилетия XXI в., важных для грамотного формирования популяционных выборок при создании генетических баз данных для целей медицинской генетики и криминалистики.

МАТЕРИАЛЫ И МЕТОДЫ

Материалом для исследования послужили материалы демографической статистики Территориального органа Федеральной службы государственной статистики по г. Москве [16] (далее – **Мосстат**), материалы Всероссийских переписей населения 2002 и 2010 гг. (далее **ВПН-2002** и **ВПН-2010**) [17, 18]. Также использовали данные анкетирования двух групп жителей Москвы: 1) “старшая” группа – практически здоровые лица, проходившие диспансеризацию на базе 15 городской клинической больницы г. Москвы в 2001 г.; 2) “младшая” группа – матери новорожденных, родившихся в 2017 г. в роддоме при Городской клинической больнице им. Мухина г. Москвы (матери сообщили информацию и об отцах детей). Возрастные распределения и их параметры для двух выборок москвичей содержатся в разделе “Результаты”.

Анкеты содержат подробные генеалогические и демографические данные (национальность, место и дата рождения самого анкетируемого и его предков в трех предшествующих поколениях). От всех анкетированных было получено информированное согласие на использование данных в анонимном генетико-демографическом исследовании.

На основе данных анкетирования, а также материалов ВПН-2002, ВПН-2010 и данных текущей демографической статистики [16–19] были рассчитаны следующие генетико-демографические

Таблица 1. Половозрастной состав населения Москвы на 1 января 2020 г. по данным Мосстата (тыс. чел.) [19]

Возрастные группы, лет	Мужчины и женщины	Мужчины	Женщины
0–4	681.0	351.6	329.4
5–9	640.4	330.4	310.0
10–14	537.7	274.5	263.2
15–19	516.9	259.4	257.5
20–24	497.8	243.5	254.3
25–29	729.1	362.1	367.0
30–34	1132.3	540.0	592.3
35–39	1155.4	555.0	600.4
40–44	1003.3	482.2	521.1
45–49	950.5	464.7	485.8
50–54	819.1	388.3	430.7
55–59	947.5	431.6	515.9
60–64	857.8	367.0	490.8
65–69	696.5	279.4	417.2
70 и старше	1512.8	528.2	984.6
Все население	12678.1	5857.9	6820.2

Примечание. Серым фоном выделена репродуктивная часть популяции.

параметры населения Москвы: N_e – эффективный объем популяции (количество человек); m – коэффициент миграции (доля мигрантов в общей структуре населения мегаполиса); R – миграционное расстояние (в км); этнический состав населения мегаполиса. Подробное описание методов расчета указанных параметров приведено в статье [20].

Этническая принадлежность анкетируемых лиц приводится на основе их самоидентификации.

РЕЗУЛЬТАТЫ

Половозрастной состав и оценка N_e

Анализ половозрастного состава населения Москвы (табл. 1) демонстрирует несовпадение с традиционной моделью половозрастной пирамиды: максимальная численность приходится не на детские возраста, а на возрастные когорты 30–34 года, 35–39 лет, 40–44 года (годы рождения 1986–1990, 1981–1985, 1976–1980 соответственно). Соотношение полов (отношение численности мужчин к численности женщин) во всей популяции составляет 0.8589. Мужчины числено преобладают в младших возрастных когортах до 20-летнего возраста, а дальше нарастает численное преобладание женщин. Однако в репродуктивной части популяции (выделена в табл. 1) мужчин больше, чем женщин, в связи с тем, что репродуктивный период длиннее для мужского пола. Вся репродуктивная часть –6805.2 тыс. чел., мужчины – 3726.8 тыс. чел., женщины – 3078.4 тыс. чел. Величина N_e, рассчитанная с учетом неравного соотношения полов в репродуктивной части московской популяции, составляет N_e = 6743.4 тыс. чел.

Средний возраст населения Москвы на 1 января 2019 г. составил 42.30 лет; 40.08 – у мужчин и 44.21 – у женщин [16].

Гендерные особенности миграционных потоков обусловливают значительные различия половозрастного состава городских жителей различных национальностей. Так, по данным ВПН-2010, на 100 женщин у русских приходится 83 мужчины, у татар – 84, у украинцев – 96, у белорусов – 97, у евреев – 109, у армян – 120, у грузин – 130, у азербайджанцев – 155, у узбеков – 186. Самые “молодые” этнические группы в Москве – узбеки (средний возраст 34.2 года, 33.8 у мужчин и 34.9 у женщин), азербайджанцы (сред. возраст 34.3 лет, 35.4 у мужчин и 32.5 у женщин) и армяне (37.8 лет, 37.8 и 37.8 соответственно). Самая “старая” этническая группа города – евреи (сред. возраст 54.7 лет, 53.7 у мужчин и 55.8 у женщин). Промежуточные значения – у белорусов (49.6, 47.5 и 51.6 лет соответственно) и украинцев (49.2, 48.2 и 50.1). Моложе средний возраст у татар (42.1, 39.6 и 44.3) и у грузин (40.4, 40.9 и 39.7). Средний возраст русских в Москве составляет 40.8 лет (38.2 у мужчин и 42.8 у женщин) [18].

Анализ распределения анкетируемых москвичей по годам рождения показывает, что в “старшей” группе, как мужчины, так и женщины в среднем 1940–1950-х годов рождения с большим размахом изменчивости, почти в 70 лет (табл. 2,

Таблица 2. Параметры возрастных распределений анкетированных лиц и коэффициенты миграции в двух выборках

Выборки анкетируемых	*N*	Параметры возрастных распределений				*m*
		$\bar{x} \pm s.e.$	min	max	σ	
“Старшая” группа						
Мужчины	93	1949 ± 1.95	1914	1984	18.79	0.413
Женщины	204	1948 ± 1.07	1922	1985	15.25	0.517
“Младшая” группа						
Отцы	165	1985 ± 0.54	1959	1998	7.05	0.636
Матери	242	1988 ± 0.27	1974	2001	5.10	0.657

Примечание. *N* – число известных дат рождения; $\bar{x} \pm s.e.$ – средний год рождения и его статистическая ошибка; σ – стандартное отклонение; *m* – коэффициент миграции.

рис. 2); в “младшей” группе родители новорожденных в среднем 1980–1990-х годов рождения (табл. 2, рис. 2). Средние возрастные характеристики анкетируемых лиц “старшей” группы соответствуют когортам 70–80-летних в табл. 1, однако на момент проведения анкетирования возраст опрашиваемых был меньше. Возрастные характеристики “младшей” возрастной группы соответствуют когорте 30–40 летних в табл. 1. Таким образом, две группы анкетируемых разделяет интервал времени, соответствующий 1.5 поколения. Важность учета возрастных характеристик выборок в популяционных исследованиях была отмечена ранее [21].

Генетические параметры миграции по данным анкетирования и переписей населения

По данным анкетирования “старшей” возрастной группы коэффициент миграции $m = 0.485$, причем у мужчин ниже, чем у женщин (0.413 и 0.517 соответственно). Для “младшей” группы значения коэффициента миграции намного больше – 0.636 для мужчин и 0.657 для женщин (табл. 2).

Коэффициент миграции, рассчитанный на основе данных ВПН-2002, равен 0.4318; на основе данных ВПН-2010 – 0.4180, причем в обоих случаях значение этого параметра у мужчин незначительно меньше, чем у женщин (табл. 3). Значениям *m*, полученным по данным переписей, в

Рис. 2. Возрастные распределения анкетированных жителей Москвы.

Таблица 3. Расчет коэффициента миграции для Москвы по данным ВПН-2002 и ВПН-2010 [17, 18]

Москва	ВПН-2002			ВПН-2010		
	всего	мужчины	женщины	всего	мужчины	женщины
Всего указавших место рождения	9680950	4568304	5112646	10898165	5026509	5871656
Живут в Москве с рождения	5500829	2629847	2870982	6342983	2936231	3406752
Доля живущих с рождения	0.5682	0.5757	0.5615	0.5820	0.5841	0.5802
Коэффициент миграции	0.4318	0.4243	0.4385	0.4180	0.4159	0.4198

Примечание. Данные переписей представлены по табл. 1 из: “Продолжительность проживания населения частных домохозяйств в месте постоянного жительства по полу и возрастным группам по субъектам Российской Федерации”. Том 10. “Продолжительность проживания населения в месте постоянного жительства”.

Таблица 4. Распределение жителей Москвы по месту рождения по данным анкетирования и по данным ВПН-2002 и ВПН-2010 [17, 18]

Место рождения	“Старшая” группа		“Младшая” группа		ВПН-2002		ВПН-2010	
	N	%	*N*	%	*N*	%	*N*	%
Россия, в т.ч.:	261	88.78	310	76.17	8796686	88.63	10154733	91.42
ЦФО, в т.ч.:	221	74.92	218	53.56	7503908	75.61	8777764	79.02
Москва	152	51.53	143	35.14	5497687	55.39	6888041	62.01
МО	17	5.76	28	6.88	698921	7.04	634748	5.71
СЗФО	3	1.02	1	0.25	199267	2.01	199633	1.80
ЮФО	4	1.36	10	2.46	245436	2.47	175259	1.58
ПФО	19	6.44	36	8.85	454494	4.58	488391	4.40
УрФО	2	0.68	4	0.98	92033	0.93	106391	0.96
СКФО**	3	1.02	23	5.65	—	—	108997	0.98
СФО	5	1.69	12	2.95	188728	1.90	190061	1.71
ДВФО	2	0.68	6	1.47	112820	1.14	108237	0.97
Азербайджан	3	1.02	5	1.23	117718	1.19	78249	0.70
Армения	2	0.68	9	2.21	71547	0.72	58295	0.52
Беларусь	4	1.36	3	0.74	89283	0.90	75652	0.68
Грузия	7	2.37			82331	0.83	55177	0.50
Казахстан	5	1.69	5	1.23	66389	0.67	74934	0.67
Киргизия			19	4.67	14686	0.15	33243	0.30
Молдавия			8	1.97	55405	0.56	40150	0.36
Таджикистан			11	2.70	44617	0.45	40183	0.36
Туркмения			3	0.74	12693	0.13	13041	0.12
Узбекистан	2	0.68	12	2.95	58797	0.59	82115	0.74
Украина	12	4.07	15	3.69	392323	3.95	303262	2.73
Другие страны***	1	0.34	7	1.72	122246	1.23	98923	0.89
Всего указали место рождения	294	100	407	100	9924721	100	11107957	100

Примечание. Данные переписей представлены по табл. 3 из: “Население по месту рождения и месту проживания на территории Российской Федерации”. Том 10. “Продолжительность проживания населения в месте постоянного жительства”. СЗФО – Северо-Западный федеральный округ, ЦФО – Центральный федеральный округ, ЮФО – Южный федеральный округ, ПФО – Приволжский федеральный округ, УрФО – Уральский федеральный округ, СКФО – Северо-Кавказский федеральный округ, СФО – Сибирский федеральный округ, ДВФО – Дальневосточный федеральный округ. ** – Северо-Кавказский федеральный округ с центром в городе Пятигорске был выделен из состава Южного федерального округа Указом Президента России от 19 января 2010 г. В составе ЮФО остались: Адыгея, Калмыкия, Краснодарский край и Астраханская, Волгоградская и Ростовская области [17, 18]. В материалах ВПН-2002 данные по ЮФО включают и данные по территориям, впоследствии вошедшим во вновь образованный СКФО. *** – данные по странам “дальнего зарубежья” вошли в категорию “другие страны”. В этой категории у анкетируемых “старшей” группы – Германия (1 чел.), в “младшей” группе – Абхазия (2 чел.), Германия (2 чел.), Египет (2 чел.), Турция (1 чел.).

наибольшей степени соответствует значение этого параметра в “старшей” возрастной группе.

Подавляющее большинство (почти 90%) анкетируемых лиц в “старшей” группе родилось в России – примерно столько же уроженцев России среди москвичей отмечено в ВПН-2002 и ВПН-2010 (табл. 4). В “младшей” возрастной группе уроженцев России гораздо меньше (76%) В “старшей” группе 3/4 москвичей родилось в ЦФО, в том числе чуть больше половины в самой столице; заметен также вклад населения ПФО, а из стран “ближнего зарубежья” – Украины и Гру-

Таблица 5. Средний радиус миграции для жителей Москвы, рассчитанный по данным анкетирования и по данным ВПН-2002 и ВПН-2010 [17, 18]

Источник данных	N	$\bar{R} \pm s.e.$ (км)	Min	Max	σ
Анкетирование: “старшая” группа	143	1008.55 ± 106.76	100	8415	1276.67
мужчины	38	862.74 ± 156.88	100	3651	967.09
женщины	105	1061.32 ± 133.88	100	8415	1371.89
Анкетирование: “младшая” группа	264	1781.59 ± 105.99	20	10211	1722.11
мужчины	105	1711.92 ± 165.46	20	9172	1695.42
женщины	159	1827.60 ± 138.25	20	10211	1743.30
Перепись 2002 г.	4304788	1277.25 ± 0.81	100	9000	1677.16
Перепись 2010 г.	4149531	1337.48 ± 0.83	100	9000	1694.02

Примечание. N – число лиц, родившихся не в Москве, с указанным местом рождения. $\bar{R}$ – средний радиус миграции; *s.e.* – статистическая ошибка; σ – стандартное отклонение. Из расчета радиуса миграции по данным переписей исключены жители стран “дальнего зарубежья”.

зии. В “младшей” группе распределение по местам рождения совсем иное – в ЦФО родилось чуть больше половины анкетируемых, в самой Москве – одна треть, много уроженцев ПФО, СКФО, ЮФО, СФО, а из стран “ближнего зарубежья” – Киргизии, Украины, Узбекистана, Таджикистана, Армении (страны перечислены в порядке убывания числа мигрантов из них в Москву).

Среднее миграционное расстояние, рассчитанное по данным анкет “старшей” группы, составляет 1008 км (863 км для мужчин, 1061 км для женщин), а для “младшей” группы гораздо больше – 1782 км (1712 км для мужчин, 1828 км для женщин) (табл. 5). Оценка среднего радиуса миграции в Москву на основе данных ВПН-2002 $\bar{R} = 1277$ км, а ВПН-2010 – $\bar{R} = 1337$ км (табл. 5).

Генетические параметры миграции по данным миграционной статистики Росстата

Данные Росстата, представленные в табл. 6, свидетельствуют о том, что в миграционных потоках в Москву мигранты из России составляют подавляющее большинство – примерно 90%. На втором и третьем местах с большим отрывом – Украина и Молдова.

В последние десять лет число мигрантов, прибывших в столицу, постепенно росло (от 125 тыс. в 2011 г. до 354 тыс. в 2018 г.); однако в 2019 г. заметно сократилось (до 300 тыс.), особенно сократилась международная миграция (до 26 тыс., очевидно вследствие ограничений, связанных с коронавирусом). Если рассматривать абсолютные значения числа мигрантов, то максимальное их число из Украины, Молдовы, Узбекистана, Армении, Азербайджана, Таджикистана, Грузии, Киргизии, Казахстана, Беларуси, Туркмении, Латвии прибыло в 2018 г. Максимальный приток из США наблюдался 2016 г., из Германии – в 2011 г., из Литвы и “других стран” – в 2013 г., из Эстонии и Израиля – в 2015 г.

В 2019 г. доля прибывших в Москву из России достигла максимальной величины – 91.42%, а доля международных мигрантов сократилась до 8.58%; на втором месте по-прежнему Украина (2.54%); на третьем – Молдова (1.12%), на четвертом – Таджикистан (0.66%); на пятом – Узбекистан (0.59%); на шестом – Азербайджан (0.56%) (табл. 6). Как видно, республики Средней Азии и Закавказья, согласно статистике Росстата, не набирают и 1%.

Динамика состава мигрантов в Москву, представленная на рис. 3, свидетельствует об относительной стабильности вклада мигрантов из отдельных стран по годам. Исключением является Украина, доля мигрантов из которой снижалась до 2013 г., а в 2014 г. (в силу известных политических событий), возросла, в 2019 г. вновь упала. Доля мигрантов из Молдовы, находившаяся в интервале от 1.3 до 1.8%, к 2019 г. падает наиболее заметно – до 1.12%. Доля мигрантов из Таджикистана, Грузии, Армении, Казахстана, Азербайджана, Узбекистана, Белоруссии упала не так значитель-

Таблица 6. Распределение мигрантов, прибывших в Москву в 2009–2019 гг., по странам выбытия (рассчитано по данным Росстата) [16]

Страны выбытия	2010		2011		2012		2013		2014		2015		2016		2018		2019	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Россия	110827	88.04	106833	85.56	185471	90.25	202488	90.00	199890	89.11	266229	90.26	212414	89.17	314868	89.02	274402	91.42
Украина	4147	3.29	4218	3.38	4761	2.32	5148	2.29	6962	3.10	8253	2.80	7337	3.08	11331	3.20	7627	2.54
Молдова	1649	1.31	2075	1.66	2722	1.32	3109	1.38	3591	1.60	4816	1.63	4254	1.79	6209	1.76	3353	1.12
Узбекистан	1111	0.88	1321	1.06	1602	0.78	1576	0.70	1791	0.80	2165	0.73	1973	0.83	2814	0.80	1768	0.59
Армения	1276	1.01	1342	1.07	1703	0.83	1930	0.86	2043	0.91	2094	0.71	1681	0.71	2345	0.66	1367	0.46
Азербайджан	1080	0.86	1200	0.96	1169	0.57	1280	0.57	1401	0.62	1772	0.60	1631	0.68	2573	0.73	1690	0.56
Таджикистан	513	0.41	554	0.44	682	0.33	802	0.36	880	0.39	1236	0.42	1244	0.52	2527	0.71	1996	0.66
Грузия	1082	0.86	1198	0.96	1191	0.58	1208	0.54	1144	0.51	1284	0.44	1105	0.46	1360	0.38	879	0.29
Киргизия	538	0.43	690	0.55	813	0.40	767	0.34	868	0.39	1082	0.37	1045	0.44	1633	0.46	1330	0.44
Казахстан	714	0.57	925	0.74	856	0.42	949	0.42	951	0.42	1116	0.38	884	0.37	1496	0.42	1098	0.37
Беларусь	649	0.52	826	0.66	686	0.33	900	0.40	998	0.44	1015	0.34	875	0.37	1598	0.45	1176	0.39
США	194	0.15	312	0.25	319	0.16	301	0.13	325	0.14	375	0.13	416	0.17	295	0.08	168	0.06
Туркмения	246	0.20	314	0.25	383	0.19	333	0.15	303	0.14	425	0.14	301	0.13	527	0.15	317	0.11
Германия	247	0.20	381	0.31	308	0.15	288	0.13	270	0.12	288	0.10	232	0.10	283	0.08	179	0.06
Латвия	141	0.11	187	0.15	213	0.10	193	0.09	206	0.09	242	0.08	185	0.08	249	0.07	163	0.05
Израиль	162	0.13	204	0.16	176	0.09	191	0.08	212	0.09	216	0.07	156	0.07	129	0.04	79	0.03
Литва	67	0.05	97	0.08	106	0.05	108	0.05	83	0.04	90	0.03	86	0.04	97	0.03	61	0.02
Эстония	95	0.08	102	0.08	113	0.05	91	0.04	94	0.04	119	0.04	79	0.03	106	0.03	60	0.02
Другие	1140	0.91	2080	1.67	2236	1.09	3331	1.48	2311	1.03	2134	0.72	2304	0.97	3252	0.92	2458	0.82
Всего	125878	100	124859	100	205510	100	224993	100	224323	100	294951	100	238202	100	353692	100	300171	100

Примечание. *N* – число человек, прибывших в Москву из каждой страны; в процентах дана доля мигрантов из каждой страны в общем потоке мигрантов.

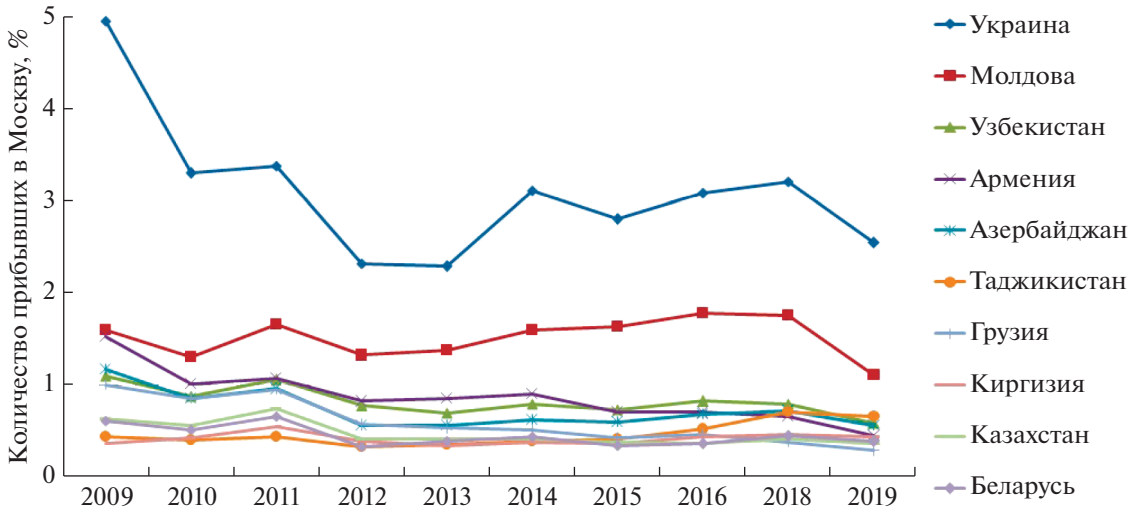

Рис. 3. Динамика состава международных мигрантов в Москве по странам выбытия в 2009–2019 гг. (рассчитано по данным Росстата). На графике представлены только страны, поставляющие наибольшее число мигрантов.

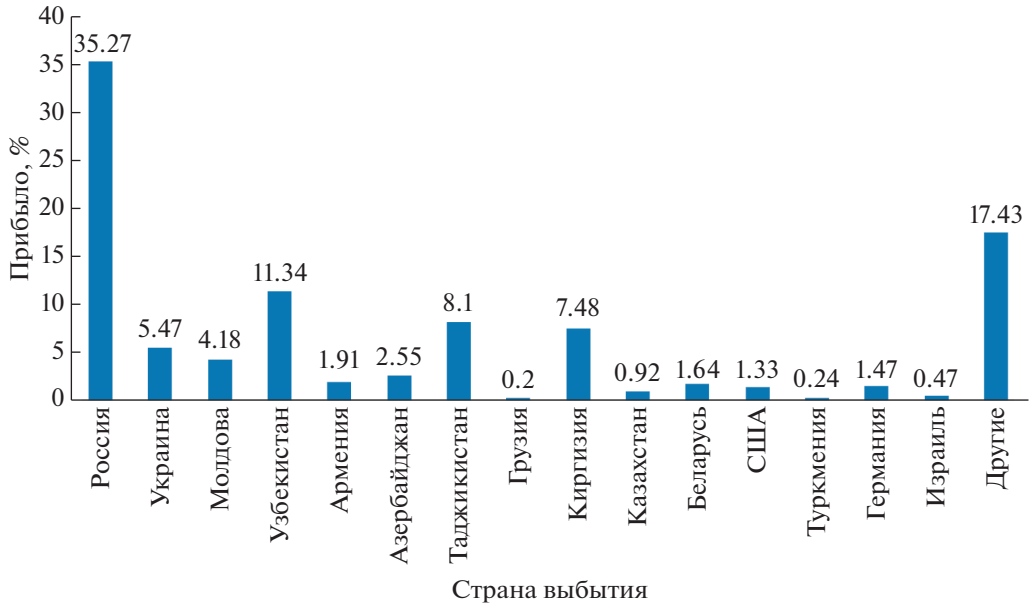

Рис. 4. Распределение мигрантов, прибывших в Москву в 2012 г., по странам выбытия (рассчитано по данным УФМС).

но, а вклад Киргизии остался примерно на том же уровне.

Коэффициент миграции, рассчитанный как число прибывших за 2019 г., умноженное на 25 (условно – длина поколения в годах) и поделенное на среднегодовую численность населения мегаполиса, $m = 0.59$ – значение, находящееся приблизительно посередине между значениями m в “младшей” и “старшей” возрастных группах.

Совершенно иную картину рисуют данные УФМС по г. Москве (рис. 4). По данным этой организации (ныне упраздненной) в 2012 г. в Москву прибыло 2.37 млн мигрантов, причем доля прибывших из России составила только 35%; второе, третье и четвертое места в этом списке занимают Узбекистан, Таджикистан и Киргизия, суммарный вклад которых в поток мигрантов в Москву составил 27%; Украина и Молдова нахо-

дятся на пятом и шестом местах соответственно. Примечательно, что из Китая прибыло примерно столько же, сколько из Азербайджана (2.5%).

Попытка рассчитать коэффициент миграции по данным УФМС заводит в тупик – при пересчете на поколение получается астрономическая сумма – почти 58 млн мигрантов, сумма почти в пять раз превышающая все население Москвы.

Расхождения в данных о странах исхода мигрантов связаны с особенностями миграционного учета граждан, свойственных двум разным учреждениям – Росстату и ФМС.

Этнический состав населения

Удельный вес лиц русской национальности в Москве по данным ВПН-2010 составляет 86%, а по данным ВПН-2002 – 85% (от числа лиц, принявших участие в переписи) (табл. 7). Обращает на себя внимание большая доля лиц, не указавших национальность при прохождении переписи, особенно в 2010 г., что затрудняет анализ этнического состава населения мегаполиса. В 2010 г. в десятку наиболее многочисленных национальностей входят, помимо русских, по убыванию численности – украинцы, татары, армяне, азербайджанцы, евреи, белорусы, грузины, узбеки, таджики. По данным анкетирования жителей Москвы наиболее разнообразен этнический состав "младшей" группы – в ней представлено 27 национальностей (в "старшей" группе – 14). Помимо русских, в "младшую" группу входят (по убыванию численности): киргизы, дагестанцы, армяне, таджики, татары, украинцы, узбеки, чеченцы, азербайджанцы, евреи, белорусы, чуваши, осетины, молдаване, корейцы, казахи, ингуши (табл. 7); категория "другие" представлена следующими этносами: гагаузы, кабардинцы, калмыки, якуты, монголы, сирийцы, турки, туркмены, египтяне.

ОБСУЖДЕНИЕ

Сравнение генетико-демографических параметров двух разновозрастных групп жителей Москвы – "старшей" (в среднем 1940–1950-х годов рождения) и "младшей" (в среднем 1980–1990-х годов рождения) свидетельствует об увеличении интенсивности миграционных потоков в Москву и о расширении области миграционного притяжения мегаполиса. Если в "старшей" группе 3/4 москвичей родились в ЦФО и лишь 11% – за пределами РФ, то в "младшей" группе в ЦФО родилось чуть больше половины, а за рубежом – почти 1/4. "Младшая" группа отличается от "старшей" более высокими значениями коэффициента миграции и среднего миграционного расстояния, а также более разнообразным этническим составом.

Сравнение генетико-демографических параметров миграции в двух крупнейших мегаполисах РФ – Москве и Санкт-Петербурге показывает, что источники формирования генофондов населения этих мегаполисов, расположенных на расстоянии 650 км друг от друга, различны – генофонд населения Санкт-Петербурга на 3/4 формируется за счет населения СЗФО [21]. Это говорит о том, что целесообразно создавать отдельные базы генетических данных для целей профилактической медицины и криминалистики для каждого мегаполиса.

Гендерные особенности миграционных потоков в Москве проявляются в различиях коэффициента миграции и радиуса миграции, а также в неравном соотношении полов в разных этнических группах (численное преобладание женщин у русских, украинцев, белорусов, татар и, напротив, – численное преобладание мужчин среди представителей "молодых" этнических диаспор – армян, азербайджанцев, узбеков). Эти особенности необходимо учитывать при формировании баз данных по однородительским маркерам.

Результаты исследования указывают на необходимость постоянного обновления и актуализации генетических баз данных, в том числе для целей ДНК-идентификации. При формировании референсных групп целесообразно учитывать возрастные характеристики. Выявленные в настоящем исследовании различия разновозрастных выборок по этнотерриториальным источникам их формирования указывают на вектор динамики генофонда населения мегаполиса в поколениях – можно прогнозировать увеличение в Москве частот тех генетических маркеров, распространенность которых выше в южных и юго-восточных регионах РФ и постсоветского пространства.

Авторы благодарят москвичей, принявших участие в анкетировании, и руководство медицинских учреждений Москвы, на базе которых было проведено анкетирования, за ценное сотрудничество.

Исследование проведено в рамках реализации Мероприятия 10 Научно-технической программы Союзного государства "Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства" ("ДНК-идентификация").

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хель-

Таблица 7. Национальный состав жителей Москвы по данным анкетирования и по данным ВПН-2002, ВПН-2010 [17, 18]

Национальность	"Старшая" группа		"Младшая" группа		ВПН-2002		ВПН-2010	
	N	%*	*N*	%*	*N*	%*	*N*	%*
Русские	286	85.37	381	72.71	8808009	84.83	9930410	86.33
Азербайджанцы	4	1.19	6	1.15	95563	0.92	57123	0.50
Армяне	5	1.49	17	3.24	124425	1.20	106466	0.93
Башкиры	1	0.30	—	—	5941	0.06	6609	0.06
Белорусы	5	1.49	3	0.57	59353	0.57	39225	0.34
Грузины	6	1.79	—	—	54387	0.52	38934	0.34
Народы Дагестана	1	0.30	15	2.86	17738	0.17	21156	0.18
Евреи	7	2.09	4	0.76	79359	0.76	53145	0.46
Ингуши	—	—	1	0.19	4050	0.04	4354	0.04
Казахи	—	—	1	0.19	7997	0.08	9393	0.08
Киргизы	—	—	24	4.58	4102	0.04	18736	0.16
Корейцы	1	0.30	1	0.19	8630	0.08	9783	0.09
Молдаване	—	—	1	0.19	36570	0.35	21699	0.19
Мордва	2	0.60	—	—	23387	0.23	17095	0.15
Немцы	1	0.30	—	—	5271	0.05	4817	0.04
Осетины	—	—	1	0.19	10561	0.10	11311	0.10
Таджики	—	—	13	2.48	35385	0.34	27280	0.24
Татары	6	1.79	13	2.48	166083	1.60	149043	1.30
Узбеки	—	—	10	1.91	24312	0.23	35595	0.31
Украинцы	8	2.39	13	2.48	253644	2.44	154104	1.34
Чеченцы			6	1.15	14465	0.14	14524	0.13
Чуваши	2	0.60	2	0.38	16011	0.15	14313	0.12
Другие**	—	—	13	2.48	110385	1.06	89977	0.78
Указавшие национальность	335		524		9965628		10835092	
Национальность не указана	—	—	—	—	417126	4.02	668409	5.81
Всего	380	100	524	100	10382754	100	11503501	100

Примечание. * От всех лиц. ** Группа "другие" в "младшей" группе: гагаузы, кабардинцы, калмыки, якуты, монголы, сирийцы, турки, туркмены, египтяне.

синкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

Исследование одобрено Этическим комитетом ИОГен РАН. От каждого из включенных в исследование участников было получено информированное добровольное согласие. Демографические данные, полученные в результате анкетирования, и результаты исследования представлены и хранятся в анонимной (обезличенной) форме.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Курбатова О.Л., Победоносцева Е.Ю., Имашева А.Г.* Роль миграционных процессов в формировании брачной структуры московской популяции. Сообщение 1. Возраст, место рождения и национальность вступающих в брак // Генетика. 1984. Т. 20. № 3. С. 501–511.
2. *Курбатова О.Л., Победоносцева Е.Ю.* Роль миграционных процессов в формировании брачной структуры московской популяции. Сообщение 2. Брачная ассортативность в отношении возраста, мест рождения и национальности супругов // Генетика. 1988. Т. 24. № 9. С. 1679–1688.
3. *Курбатова О.Л., Победоносцева Е.Ю.* Роль миграционных процессов в формировании брачной структуры московской популяции. Сообщение 3. Брачная ассортативность как фактор, противодействующий аутбридингу // Генетика. 1988. Т. 24. № 9. С. 1689–1695.
4. *Kurbatova O.L., Pobedonostseva E.Yu.* Genetic demography of the Moscow population: Migration, outbreeding and assortative mating // Ann. of Hum. Biol. 1991. V. 18. № 1. P. 31–46.
5. *Kurbatova O.L., Pobedonostseva E.Yu*. Assortative mating among minority ethnic groups in Moscow and other large cities of the CIS // Minority Populations. Genetics, Demography and Health / Eds Bittles A.N., Roberts D.F. London: The Macmillan Press Ltd., 1992. P. 241–256.
6. *Курбатова О.Л., Победоносцева Е.Ю.* Генетико-демографические процессы в многонациональной популяции // Успехи совр. генетики. 1996. Вып. 20. С. 38–61.
7. *Курбатова О.Л., Победоносцева Е.Ю., Свежинский Е.А.* Генетико-демографические процессы в московской популяции в середине 1990-х годов. Миграция и эмиграция как факторы изменения генетического разнообразия популяции // Генетика. 1997. Т. 33. № 12. С. 1688–1696.
8. *Победоносцева Е.Ю., Свежинский Е.А., Курбатова О.Л.* Генетико-демографические процессы в московской популяции в середине 1990-х годов. Анализ этногеографических параметров миграции: изоляция расстоянием // Генетика. 1998. Т. 34. № 3. С. 423–430.
9. *Свежинский Е.А., Курбатова О.Л.* Опыт исторической реконструкции генетико-демографической структуры московской популяции на рубеже XIX и XX вв. // Генетика. 1999. Т. 35. № 8. С. 1149–1159.
10. *Курбатова О.Л., Победоносцева Е.Ю., Свежинский Е.А.* Влияние этноконфессиональных факторов на динамику генофонда населения Москвы // Мусульмане изменяющейся России. М.: Росс. политич. энциклоп. (РОССПЭН), 2002. С. 142–172.
11. *Курбатова О.Л., Победоносцева Е.Ю.* Городские популяции: возможности генетической демографии (миграция, подразделенность, аутбридинг) // Вестник ВОГиС. 2006. Т. 10. № 1. С. 155–188.
12. *Курбатова О.Л., Победоносцева Е.Ю., Веремейчик В.М. и др.* Особенности генетико-демографических процессов в населении трех мегаполисов в связи с проблемой создания генетических баз данных // Генетика. 2013. Т. 49. № 4. С. 513–522.
13. *Курбатова О.Л., Янковский Н.К.* Миграция – основной фактор популяционной динамики городского населения России // Генетика. 2016. Т. 52. № 7. С. 831–851.
14. *Курбатова О.Л.* Демографическая генетика городского населения. Автореф. дис. ... докт. биол. наук. М.: ИОГен РАН, 2014. 49 с.
15. Население Москвы [Электронный ресурс]. Материал из Википедии – свободной энциклопедии. Режим доступа: https://ru.wikipedia.org/wiki/Население_Москвы, свободный. Загл. с экрана. (28.04.2021).
16. Управление Федеральной службы государственной статистики по г. Москве и Московской области [Электронный ресурс]. Режим доступа: https://mosstat.gks.ru, свободный. Загл. с экрана. (27.04.2021).
17. Всероссийская перепись населения 2002 года [Электронный ресурс] / Официальный сайт Всероссийской переписи населения. Режим доступа: http://www.perepis2002.ru/index.html?id=11, свободный. Загл. с экрана. (14.04.2021).
18. Всероссийская перепись населения 2010 года [Электронный ресурс]. Режим доступа: https://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm, свободный. Загл. с экрана. (27.04.2021).
19. Московский статистический ежегодник. 2020. Статистический сборник. М.: Мосстат, 2020. 200 с.
20. *Курбатова О.Л., Удина И.Г., Грачева А.С. и др.* Генетико-демографические параметры населения г. Санкт-Петербурга. Миграционные процессы // Генетика. 2019. Т. 55. № 9. С. 1071–1082.
21. *Курбатова О.Л., Удина И.Г., Грачева А.С. и др.* Генетико-демографические параметры населения г. Новосибирска // Генетика. 2018. Т. 54. Приложение. S74–S84. https://doi.org/10.1134/S0016675818130106

Genetic Demography of the Population of Moscow. Migration Processes

O. L. Kurbatova[a], *, A. S. Gracheva[a], **, E. Yu. Pobedonostseva[a], and I. G. Udina[a]

[a] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia*

**e-mail: okurbat@list.ru*

***e-mail: palesa@yandex.ru*

Determination of the principal genetic-demographic parameters of migration processes that are significant for studying the structure of the Moscow population gene pool is based on an analysis of demographic statistics, All-Russia population censuses and data from a sample survey of two groups of Moscow residents differing in age characteristics: the "older" group, mean year of birth 1948; "younger" – born in 1987 on average. A significant increase in the migration intensity and migration distance in Moscow over the past 40 years has been shown. The migration coefficient, calculated according to the survey data of the "older" group of Muscovites, is equal to 0.413 for men and 0.517 for women, the average migration radius is 1008 km; according to the survey of the "younger" group – the migration coefficient is 0.636 for men and 0.657 for women, the average migration radius is 1782 km. Gender differences in migration flows are manifested in different age structure and in unequal sex ratio within particular ethnic groups of Muscovites (e.g., predominance of men in the "young" ethnic diasporas). Data on the sources of the formation of the Moscow population gene pool at the present stage show that the share of those born in Russia in the "older" age group (89%) is much higher than in the "younger" age group (76%); in the first case, the gene pool is formed by 3/4 at the expense of the population of the Central Federal District (CFD), the contribution of the Volga Federal District and the Ukraine is noticeable; in the second case, only half of the respondents were born in the CFD, many are natives of the Volga, North Caucasian, Southern and Siberian districts, the republics of Central Asia, Ukraine, and Armenia. The survey data are confirmed by the analysis of Rosstat Agency migration statistics data on the composition of migrants to Moscow by country of exit. Differences in migration parameters in two age groups indicate the possibility of dynamics of the ethno-regional composition of the Moscow population in subsequent generations, which will certainly cause a change in the frequencies of many genetic markers, including the frequency profiles of DNA-based identification markers. The features of migration processes in Moscow revealed in this study indicate the need for timely updating of genetic databases for the purposes of DNA-based identification in a megalopolis.

Keywords: genetic demography, megalopolises, Moscow, questionnaires, migration statistics, migration rate, migration distance, ethnic composition, gene pool dynamics, databases for DNA-based identification purposes.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1450–1457

УДК 575.162:575.167:57.024

ПОЛНОГЕНОМНОЕ ИССЛЕДОВАНИЕ СВЯЗИ МЕТИЛИРОВАНИЯ CpG-САЙТОВ С АГРЕССИВНЫМ ПОВЕДЕНИЕМ

© 2021 г. С. А. Боринская[1], *, А. В. Рубанович[1], **, А. К. Ларин[2], А. В. Казанцева[3], Ю. Д. Давыдова[3], Э. В. Генерозов[2], Э. К. Хуснутдинова[3], Н. К. Янковский[1]

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

[2]Федеральный научно-клинический центр физико-химической медицины Федерального медико-биологического агентства, Москва, 119435 Россия

[3]Институт биохимии и генетики Уфимского научного центра Российской академии наук, Уфа, 450054 Россия

**e-mail: borinskaya@vigg.ru*

***e-mail: rubanovich@vigg.ru*

Поступила в редакцию 02.08.2021 г.
После доработки 09.08.2021 г.
Принята к публикации 10.08.2021 г.

Профиль метилирования ДНК может отражать как генетические особенности, так и внешние воздействия, которые влияют на ряд признаков человека, включая характеристики поведения, в том числе антисоциальное поведение. Для исследования связи метилирования CpG-сайтов с агрессивным поведением был определен уровень метилирования 800K-сайтов в ДНК, выделенной из периферической крови у 122 мужчин, совершивших убийства, и у 163 мужчин контрольной группы. Для 25 топовых по эффекту CpG-сайтов ($p < 1E-22$) ассоциация с агрессивным поведением воспроизводились в тестовой выборке и оставалась значимой на уровне эпигенома после поправки на сопутствующие факторы (возраст, курение, национальность). Среди факторов, ассоциированных с генами, с которыми связаны выявленные CpG-сайты, представлены: неблагоприятные условия в детстве, курение, злоупотребление алкоголем и астма, что соответствует ранее опубликованным данным по факторам, связанным с физической агрессией. Полученные результаты вносят вклад в понимание биологических основ агрессии и выявление биомаркеров агрессивного поведения.

Ключевые слова: EWAS, метилирование ДНК, физическая агрессия, биомаркеры агрессивного поведения, антисоциальное поведение.

DOI: 10.31857/S0016675821120043

Агрессивное поведение представляет одну из важных социальных проблем. По данным Всемирной организации здравоохранения, убийства и умышленное причинение тяжкого вреда здоровью входит в ведущие причины смертности во всем мире для людей в возрасте 15–44 лет [1]. В России наиболее высокий уровень убийств наблюдался в 1990-е–2000-е годы, достигая в отдельные периоды 28–30 на 100000 населения, но к 2018–2020 гг. снизился до 5.4–5.8 на 100000 человек [2, 3].

Исследования показывают, что агрессивное поведение, в том числе экстремально агрессивное, ведущее к совершению серийных убийств, имеет наследственную составляющую [4–9], при этом влияние наследственности, оцененное в близнецовых исследованиях и исследованиях приемных детей, составляет около 50% [6, 10].

Средовые воздействия, такие как ряд заболеваний или стресс, или курение матери во время беременности, неблагоприятные условия воспитания в детстве и другие могут приводить к повышению риска формирования асоциального, в том числе агрессивного поведения. Особенно значимы воздействия, перенесенные в пренатальный и постнатальный периоды, которые, как показано для человека и модельных животных, могут вызывать долговременные эпигенетические модификации активности генов, связанных с регуляцией поведения, в том числе за счет изменения уровня метилирования специфичных CpG-сайтов [11–17]. Показано, для некоторых признаков предикция на основе уровня метилирования ассоциированных с признаком эпигенетических маркеров более эффективна, чем на основе полигенного показателя риска [18, 19]. Изменения уровня метилирования CpG-сайтов, связанные с агрессивным поведением, показаны как в иссле-

дованиях отдельных генов, так и в полногеномом анализе ассоциаций эпигенетических маркеров (**EWAS**, epigenome-wide association studies) (см. обзор и ссылки в [17]). Уровень агрессии в этих исследованиях определяется по преступным деяниям, либо опросникам, оценивающим проявления агрессивного поведения. Однако биомаркеры агрессивного поведения, в том числе экстремальной физической агрессии, проявляющейся как умышленное убийство, которые воспроизводились бы от исследования к исследованию, до настоящего времени не выявлены [17, 20], поэтому поиск их представляется актуальным.

В данной работе мы провели полногеномное исследование ассоциации уровня метилирования CpG-сайтов в ДНК периферической крови с агрессивным поведением у мужчин, совершивших убийство, в сравнении смужчинами контрольной популяционной выборки.

МАТЕРИАЛЫ И МЕТОДЫ

В исследовании были использованы образцы венозной крови лиц, совершивших убийства и госпитализированных для психиатрического обследования в 2016–2018 гг. Все образцы крови были собраны у неродственных индивидов под контролем биоэтического комитета ИБГ УФИЦ РАН с соблюдением процедуры “информированного согласия” и сбором данных о месте рождения, национальной принадлежности, курении, потреблении алкоголя и социоэкономических характеристиках. В группу лиц с агрессивным поведением (совершивших убийства) вошло 134 мужчины в возрасте 40.9 ± 13.4 лет – русских, татар, башкир, удмуртов и других национальностей. Контрольная группа представлена образцами: ДНК из ранее собранных популяционных коллекций лаборатории анализа генома ИОГен РАН и коллекций биологических материалов человека ИБГ УНЦ РАН и включала 163 мужчин в возрасте 46.1 ± 9.4 – лет русских, татар, башкир, удмуртов и др. ДНК выделена из лимфоцитов периферической крови методом фенольно-хлороформной экстракции.

Анализ паттернов метилирования проводился согласно методике производителя с помощью чипа InfiniumMethylationEPICBeadChipKit (Illumina, США) (EPIC 850K), который обеспечивает определение уровня метилирования 863904 CpG-сайтов, покрывающих весь геном [21, 22]. Вслед за экспериментальным определением уровня метилирования проводили нормализацию данных с контролем качества определения метилирования по каждому сайту, контролем качества образца (исключаются образцы с большим количеством сайтов с некачественно определенным уровнем метилирования), нормализацией значений для коррекции технических ошибок, нормализация значений для коррекции групповых эффектов (чип/чип корреляция) согласно процедурам, описанным в [22]. Из общего количества сайтов, для которых были получены оценки уровней метилирования (beta) с использованием описанного выше метода нормализации данных, после контроля качества метилирования по сайтам и образцам осталось 823843 CpG-сайтов, а четыре образца не прошли контроль качества.

Исследуемые образцы были разделены на две группы – обучающая выборка (52 мужчин с агрессивным поведением и 126 в контрольной группе) и подтверждающая выборка (82 и 37 образцов соответственно). Обучающая и подтверждающая выборки были независимыми и анализировались в различные периоды исследования. Значимость различий уровней метилирования CpG-сайтов у убийц по сравнению с контролем оценивалась по непараметрическому тесту Манна–Уитни. Анализ возможных эффектов сопутствующих факторов (возраст, курение, национальность) проводился по методу ANOVA, реализованному в программе SPSS v. 12.

Фенотипы, сопряженные с изменениями паттернов метилирования генов, к которым относятся выявленные CpG-сайты, приведены согласно открытой базе данных Атлас EWAS [23], который с одержит описания 288974 CpG в 35594 генах.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При анализе метилома с помощью чипа EPIC 850 K у мужчин с агрессивным поведением в сравнении с мужчинами контрольной выборки в обучающей выборке было обнаружено 19876 CpG-сайтов, уровень метилирования которых отличался со статистической значимостью выше порогового согласно поправке на множественное тестирование (поправка Бонферрони (0.05/823843 = 6.1E–9)). Из них 25 топовых по эффекту CpG-сайтов ($p < 1E–22$) уверенно воспроизводились в тестовой выборке (табл. 1, рис. 1). Влияние сопутствующих факторов (возраст, курение, национальность) оказалось незначимым (табл. 2).

Согласно описанию CpG-сайтов чипа EPIC 850K Infinium Methylation EPIC v1.0 B5 Manifest File (https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html) 22 из выявленных нами сайтов расположены в теле, шельфе или в промотрном участке 20 структурных генов и генов микроРНК, тогда как для трех сайтов связь с генами неизвестна. Согласно информации, представленной в атласе EWAS [23],

Таблица 1. Топовые по эффекту CpG-сайты, уровень метилирования которых ассоциирован с агрессивным поведением

Название CpG-сайта	Ген	Локализация	Фенотипы	Значимость различий	
				обучающая выборка	тестовая выборка
cg20450977	*MTRNR2L8*	Тело гена, вне островка	Неблагоприятные условия в детстве, ИМТ, бесплодие	1.9E-44	1.0E-47
cg24159721	*RP11-206L10.17*	То же	Fibroblasts	8.3E-44	5.5E-48
cg27559706	*DHFR*	Промотер, шельф	Курение, синдром Дауна	1.5E-33	1.5E-36
cg15890734	*DHFR*	Тело гена, шельф	То же	4.0E-32	2.9E-46
cg11628021	–	–	–	4.9E-32	1.5E-03
cg05001044	*RP11-206L10.17*	Тело гена, вне островка	Возраст, болезнь Крейтцфельдта–Якоба, синдром Дауна	6.7E-32	7.7E-34
cg11894324	*RP5-857K21.5*	Тело гена, шельф	–	2.3E-30	2.2E-39
cg16581032	–	–	–	2.2E-29	6.9E-48
cg17759706	*ANKRD11*	Тело гена, шельф	Неблагоприятные условия в детстве, возраст, алкогольная зависимость, курение, астма, ИМТ	4.8E-28	1.6E-05
cg16340030	*STX3*	Тело гена, вне островка	Астма, синдром Дауна	7.3E-28	1.3E-10
cg16267121	*MTRNR2L3*	Промотер, шельф	Астма, колоректальный рак	1.5E-27	2.7E-46
cg23106865	*MTRNR2L6*	Промотер, вне островка	То же	1.7E-27	5.8E-08
cg22400896	*DENND5A*	Тело гена, вне островка	Возраст, курение, лейкемия	3.2E-27	1.2E-05
cg11330075	*hsa-mir-1977, RP5-857K21.4*	Тело гена, шельф	–	8.8E-27	1.2E-45
cg05812299	*MTRNR2L5*	Промотор, шельф	Астма, колоректальный рак	2.9E-25	4.7E-23
cg25191519	*SPG7*	Тело гена, шельф	Неблагоприятные условия в детстве, возраст, алкогольная зависимость, курение, астма, ИМТ, синдром Дауна	3.3E-25	1.3E-05
cg27573888	*C1orf38*	Тело гена, вне островка	–	1.6E-23	2.0E-05
cg27237039	*SNRNP48*	Тело гена, шельф	Возраст, синдром Дауна	3.2E-23	2.6E-03
cg03228760	*SS18L1*	То же	Возраст, курение, алкогольная зависимость	4.1E-23	1.1E-09

Таблица 1. Окончание

Название CpG-сайта	Ген	Локализация	Фенотипы	Значимость различий	
				обучающая выборка	тестовая выборка
cg14215390	*PHF20*	Промотер, берег	Возраст, курение	5.1E-23	1.9E-03
cg12376350	*hsa-mir-1977, RP5-857K21.4*	Тело гена, шельф	–	1.1E-22	4.9E-38
cg22914762	*MTRNR2L13*	Промотор, вне островка	Болезнь Бехчета, рецептор фактора некроза опухоли	1.1E-22	4.9E-15
cg23127585	–	Шельф	Синдром Шегрена	3.6E-22	6.1E-06
cg26634885	*PLCD3*	Промотор, островок	Шизофрения, возраст, курение, астма, синдром Дауна	5.7E-21	7.4E-07
cg03838322	*ZNF326*	Тело гена, шельф	Высокое потребление насыщенных жирных кислот, инсулиновая резистентность	5.8E-21	1.1E-04

Примечание. Прочерки означают, что нет данных.

метилирование других CpG-сайтов в генах, с которыми связаны выявленные нами CpG-сайтов, ассоциировано с курением и возрастом (по восьми случаям среди наших 25 сайтов), с неблагоприятными условиями в детстве и злоупотреблением алкоголя (по трем генам), с астмой (сайты в семи генах), столько же с синдромом Дауна и один сайт в гене, ассоциированном с шизофренией. Для того, чтобы оценить обогащенность выявленной группы генов генами с какими-либо определенными функциями, мы сравнили частоту встречаемости ассоциированных с генами фенотипов в выявленной нами группе и в целом в атласе EWAS (табл. 3). В выявленной группе генов в два раза чаще, чем в среднем по атласу EWAS представлены гены, ассоциированные с неблагоприятными условиями в детстве, курением, потреблением алкоголя и астмой. Изменение уровня метилирования аналогичных групп генов (ассоциированных с неблагоприятными условия в детстве, курением и потреблением алкоголя и иммунными функциями) выявлено в опубликованных исследованиях связи метилирования с агрессивным поведением [17, 24]. В настоящем исследовании указанное превышение частоты было статистически значимым лишь для астмы ($p = 0.019$); для неблагоприятных условий в детстве, курения и потребления алкоголя превышения частот оказались на уровне тенденции ($0.05 < p < 0.07$). Для остальных выявленных фенотипов частота встречаемости близка к средней частоте встречаемости в атласе EWAS.

В исследованиях эпигенетических маркеров, ассоциированных со средовыми воздействиями, такими как курение [25] или неблагоприятными условиями в детстве [26], наблюдается воспроизводимость большинства связанных с этими признаками эпигенетических маркеров от исследования к исследованию. Однако для агрессивного поведения такие устойчиво воспроизводимые маркеры пока не выявлены [17, 20, 24]. Выявленные в данном исследовании маркеры, ассоциированные с экстремально агрессивным поведением (убийством), необходимо подтвердить в дальнейших исследованиях на независимых выборках. Поиск биомаркеров агрессивного поведения, как указывающих на причинные изменения, лежащих в основе связанных с агрессией нейроэндокринных, иммунных и метаболических нарушений, так и маркеров, отражающих сопряженные влияния окружающей среды, соответствует общей стратегии развития мер предотвращения асоциального поведения, включающей как мониторинг семейных и социоэкономических факторов, так и медико-биологических [20, 27–30].

Работа выполнена в рамках мероприятия 5 “Разработка методики определения статуса гене-

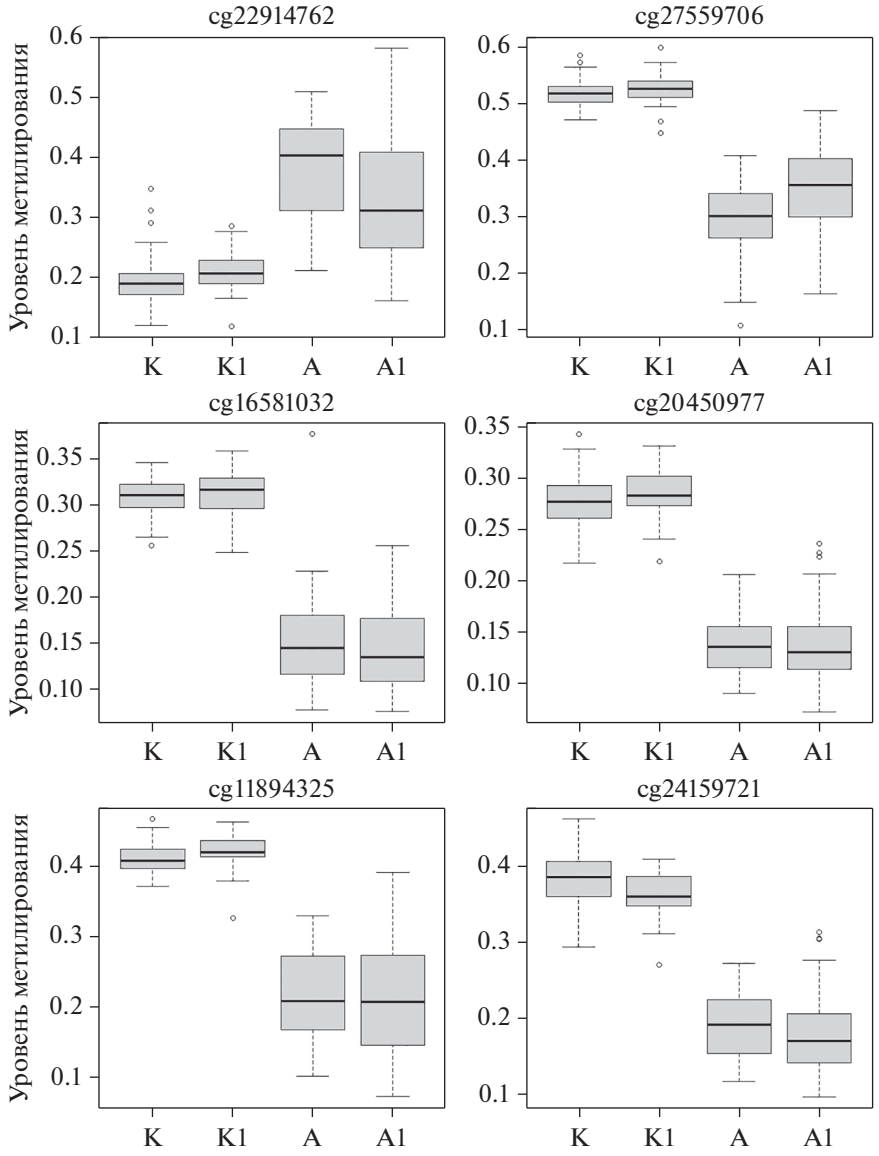

Рис. 1. Примеры гипер- и гипометилирования CpG-сайтов у убийц по сравнению с контролем. А – агрессивное поведение, К – популяционный контроль, индекс "1" маркирует тестовую выборку, оси абсцисс – исследованные выборки.

тических локусов и прижизненной модификации участков ДНК, влияющих на психоэмоциональный статус человека" научно-технической программы Союзного государства "Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства" ("ДНК-идентификация") при финансовой поддержке Министерства науки и образования РФ (Государственный контракт № 011-17 от 26.09.2017).

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.

Таблица 2. Значимость влияния ковариат (возраст, курение, национальность) на различия уровней метилирования между группами с агрессивным поведением и контролем

	Значимость действия фактора (*p*-значения)							
Факторы	статус		возраст		курение		национальность	
Сравнения:	A-K	A1-K1	A-K	A1-K1	A-K	A1-K1	A-K	A1-K1
cg20450977	3.4E-56*	7.8E-41*	0.901	0.432	0.079	0.743	0.478	0.113
cg24159721	9.5E-54*	3.3E-39*	0.576	0.056	0.007*	0.563	0.859	0.069
cg27559706	6.9E-64*	1.7E-23*	0.006*	0.410	0.185	0.016*	0.627	0.425
cg15890734	3.5E-54*	9.9E-36*	0.386	0.141	0.964	0.514	0.129	0.123
cg11628021	4.3E-27*	4.0E-03*	0.358	0.562	0.464	0.748	0.694	0.273
cg05001044	3.0E-62*	6.1E-22*	0.294	0.240	0.765	0.959	0.938	0.770
cg11894324	2.5E-58*	1.9E-26*	0.088	0.212	0.239	0.606	0.312	0.684
cg16581032	2.0E-51*	5.8E-36*	0.428	0.277	0.299	0.459	0.637	0.383
cg17759706	1.8E-18*	6.9E-06*	0.418	0.779	0.492	0.683	0.709	0.632
cg16340030	2.5E-12*	8.5E-06*	0.306	0.423	0.292	0.263	0.612	0.789
cg16267121	3.2E-45*	4.6E-33*	0.035	0.235	0.220	0.644	0.167	0.346
cg23106865	1.0E-18*	1.1E-04*	0.767	0.959	0.273	0.778	0.210	0.029
cg22400896	2.9E-12*	5.5E-05*	0.736	0.120	0.116	0.156	0.446	0.294
cg11330075	1.4E-46*	3.8E-32*	0.736	0.254	0.571	0.747	0.779	0.519
cg05812299	4.9E-38*	1.4E-15*	0.487	0.050	0.435	1.0E-04*	0.196	0.008*
cg25191519	4.2E-17*	1.5E-04*	0.548	0.859	0.392	0.100	0.501	0.073
cg27573888	2.5E-15*	3.2E-05*	0.517	0.307	0.546	0.075	0.229	0.708
cg27237039	3.2E-11*	2.0E-03*	0.501	0.145	0.828	0.405	0.687	0.129
cg03228760	2.3E-18*	3.4E-07*	0.050	0.935	0.964	0.907	0.168	0.990
cg14215390	2.2E-14*	9.0E-03*	0.590	0.218	0.379	0.114	0.574	0.550
cg12376350	1.5E-27*	5.3E-27*	0.760	0.191	0.026*	0.328	0.275	0.068
cg22914762	8.5E-33*	7.4E-10*	0.719	0.084	0.769	0.001*	0.054	0.002*
cg23127585	2.2E-11*	1.3E-04*	0.322	0.425	0.229	0.751	0.636	0.316
cg26634885	3.5E-09*	7.5E-06*	0.226	0.146	0.251	0.047*	0.737	0.169
cg03838322	3.9E-14*	2.0E-03*	0.005	0.999	0.345	0.905	0.560	0.943

Примечание. Значимые различия отмечены звездочкой.

Таблица 3. Частота встречаемости фенотипов, ассоциированных с генами, с которыми связаны выявленные CpG-сайты

Фенотип	Данное исследование		Атлас EWAS		*p*-value
	количество ассоциированных генов	частота	количество ассоциированных генов	частота	
Неблагоприятные условия в детстве	3	0.188	3348	0.094	0.057
Курение	8	0.500	12603	0.354	0.072
Потребление алкоголя	3	0.188	3273	0.092	0.053
Возраст	8	0.500	14296	0.402	0.145
Астма	7	0.438	8352	0.235	0.019
Синдром Дауна	7	0.438	12062	0.339	0.137
Шизофрения	1	0.063	2846	0.080	0.369

СПИСОК ЛИТЕРАТУРЫ

1. World Health Organization. Preventing suicide: A global imperative. World Health Organization. 2014. https://apps.who.int/iris/handle/10665/131056
2. *Богданов С.В., Репецкая А.Л.* Убийства в России и США: сравнительный анализ криминальной статистики // Криминол. журн. Байкальского гос. ун-та эконом. права. 2009. № 4. С. 13–22.
3. ЕМИСС [Электронный ресурс]. Режим доступа: https://www.fedstat.ru. (дата обращения: 05.07.2021).
4. *Craig I.W., Halton K.E.* Genetics of human aggressive behavior // Hum. Genet. 2009. V. 126. № 1. P. 101–113. https://doi.org/10.1007/s00439-009-0695-9
5. *Pavlov K.A., Chistiakov D.A., Chekhonin V.P.* Genetic determinants of aggression and impulsivity in humans // J. Appl. Genet. 2012. V. 53. № 1. P. 61–82. https://doi.org/10.1007/s13353-011-0069-6
6. *Plomin R., DeFries J.C., Knopik V.S., Neiderhiser J.* Behavior Genetics (6th ed.). N.Y.: Worth Publ., 2013. 560 p.
7. *Tiihonen J., Rautiainen M., Ollila H.M. et al.* Genetic background of extreme violent behavior // Mol. Psychiatry. 2015. V. 20. № 6. P. 786–792. https://doi.org/10.1038/mp.2014.130
8. *Лазебный О.Е., Куликов А.М., Бутовская П.Р. и др.* Анализ агрессивного поведения с помощью 250 SNP-маркеров у молодых русских мужчин // Генетика. 2020. Т. 56. № 9. С. 1075–1086. https://doi.org/10.31857/S0016675820080093
9. *Драгович А.Ю., Боринская С.А.* Генетическая и геномная основа агрессивного поведения человека // Генетика. 2019. Т. 55. № 12. С. 1381–1396.
10. *Waltes R., Chiocchetti A.G., Freitag C.M.* The neurobiological basis of human aggression: A review on genetic and epigenetic mechanisms // Am. J. Med. Genet. Part B. Neuropsychiatric Genetics. 2015. V. 171. № 5. P. 650–675. https://doi.org/10.1002/ajmg.b.32388
11. *Martin E.M., Fry R.C.* Environmental influences on the epigenome: Exposure-associated DNA methylation in human populations // Annu Rev. Public Health. 2018. V. 39. P. 309–333. https://doi.org/10.1146/annurevpublhealth-040617-014629
12. *Tielbeek J.J., Johansson A., Polderman T.J.C. et al.* Genome-wide association studies of a broad spectrum of antisocial behavior // JAMA Psychiatry. 2017. V. 74. № 112. P. 1242. https://doi.org/10.1001/jamapsychiatry.2017.3069
13. *Chistiakov D.A., Chekhonin V.P.* Early-life adversity-induced long-term epigenetic programming associated with early onset of chronic physical aggression: Studies in humans and animals // World J. Biol. Psychiatry. 2019. V. 20. № 4. P. 258–277. https://doi.org/10.1080/15622975.2017.1322714
14. *Dintsova V.V., Roetman P.J., Ip H.F. et al.* Genomics of human aggression: Current state of genome-wide studies and an automated systematic review tool // Psychiatr. Genet. 2019. V. 29. № 5. P. 170–190. https://doi.org/10.1097/YPG.0000000000000239
15. *Ip H.F., van der Laan C.M., Brikell I. et al.* Genetic Association Study of Childhood Aggression across raters, instruments and age // Transl. Psychiatry. 2021. V. 11. № 1. P. 413. https://doi.org/10.1038/s41398-021-01480-x
16. *van Dongen J., Hagenbeek F.A., Suderman M. et al.* DNA methylation signatures of aggression and closely related constructs: A meta-analysis of epigenome-wide studies across the lifespan // Mol. Psychiatry. 2021. V. 26. P. 1–15. https://doi.org/10.1038/s41380-020-00987-x
17. *Shah S., Bonder M.J., Marioni R.E. et al.* Improving phenotypic prediction by combining genetic and epigenetic associations // Am. J. Hum. Genet. 2015. V. 97. № 1. P. 75–85. https://doi.org/10.1016/j.ajhg.2015.05.014
18. *McCartney D.L., Hillary R.F., Stevenson A.J. et al.* Epigenetic prediction of complex traits and death // Genome Biol. 2018. V. 19. № 1. P. 136. https://doi.org/10.1186/s13059-018-1514-1
19. *Мустафин Р.Н., Казанцева А.В., Еникеева Р.Ф. и др.* Эпигенетика агрессивного поведения // Генетика. 2019. Т. 55. № 9. С. 987–997.
20. *Mitjans M., Seidel J., Begemann M. et al.* Violent aggression predicted by multiple pre-adult environmental hits // Mol. Psychiatry. 2019. V. 24. P. 1549–1564. https://doi.org/10.1038/s41380-018-0043-3
21. *Pidsley R., Zotenko E., Peters T.J. et al.* Critical evaluation of the Illumina Methylation EPIC BeadChip microarray for whole-genome DNA methylation profiling // Genome Biol. 2016. V. 17. № 1. P. 208. https://doi.org/10.1186/s13059-016-1066-1
22. *Wu M.C., Kuan P.F.* A guide to Illumina bead chip data analysis // DNA Methylation Protocols. Methods in Molecular Biology / Ed. Tost J. N.Y.: Humana Press, 2018. V. 1708. P. 303–329.
23. EWAS Atlas [Электронный ресурс]. Режим доступа: https://ngdc.cncb.ac.cn/ewas/atlas/index (дата обращения: 05.07.2021).
24. *Cecil C.A.M., Walton E., Pingault J.B. et al.* DRD4 methylation as a potential biomarker for physical aggression: An epigenome-wide, cross-tissue investigation // Am. J. Med. Genet. B. Neuropsychiatr Genet. 2018. V. 177(8). P. 746–764. https://doi.org/10.1002/ajmg.b.32689
25. *Silva C.P., Kamens H.M.* Cigarette smoke-induced alterations in blood: A review of research on DNA methylation and gene expression // Exp. Clin. Psychopharmacol. 2021. V. 29. № 1. P. 116–135. https://doi.org/10.1037/pha0000382
26. *Cecil C.A.M., Zhang Y., Nolte T.* Neurosci childhood maltreatment and DNA methylation: A systematic review // Biobehav. Rev. 2020. V. 112. P. 392–409. https://doi.org/10.1016/j.neubiorev.2020.02.019
27. *Raine A.* Biosocial studies of antisocial and violent behavior in children and adults: A review // J. Abnorm. Child Psychol. 2002. V. 30. № 4. P. 311–326. https://doi.org/10.1023/a:1015754122318
28. *Fazel S., Gulati G., Linsell L. et al.* Schizophrenia and violence: Systematic review and meta-analysis // PLoS Med. 2009. V. 6. № 8. P. e1000120. https://doi.org/10.1371/journal.pmed.1000120
29. *Tuvblad C., Baker L.A.* Human aggression across the lifespan: Genetic propensities and environmental moderators // Adv. in Genetics. 2011. V. 75. P. 171–214. https://doi.org/10.1016/B978-0-12-380858-5.00007-1

30. *Ehrenreich H.* The impact of environment on abnormal behavior and mental disease // EMBO Rep. 2017. V. 18. № 5. P. 661–665. https://doi.org/10.15252/embr.201744197

Epigenome Wide Study of Association of CpG Sites Methylation and Aggressive Behavior

S. A. Borinskaya[a], *, A. V. Rubanovich[a], **, A. K. Larin[b], A. V. Kazantseva[c], Yu. V. Davydova[c], E. V. Generozov[b], E. K. Khusnutdinova[c], and N. K. Yankovsky[a]

[a]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia

[b]Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, 119435 Russia

[c]Institute of Biochemistry and Genetics UFRC Russian Academy of Sciences, Ufa, 450054 Russia

**e-mail: borinskaya@vigg.ru*

***e-mail: rubanovich@vigg.ru*

The DNA methylation profile may reflect both genetic features and exposure to external factors that affect certain human traits, including behavioral characteristics. To study an association of CpG methylation with aggressive behavior, the methylation level of 800K sites was determined in peripheral blood DNA samples from 122 men who have committed murder and 163 men of a control group. For 25 top CpG sites ($p < 1E-22$), the association with aggressive behavior was reproduced in the test sample and remained significant at the epigenome level after correction for concurrent factors (age, smoking status, and ethnicity). Phenotypes associated witht hegenes that correspond to the identified CpG sites include child abuse, smoking, alcohol abuse, and asthma, in agreement with published data on physical aggression-associated factors. There sults contribute to under standing the biological basis of aggression and identifying biomarkers of aggressive behavior.

Keywords: EWAS, DNA methylation, physical aggression, biomarkers of aggressive behavior, antisocial behavior.

ГЕНЕТИКА, 2021, том 57, № 12, с. 1458–1464

УДК 575.224.22

АДАПТИВНЫЕ ИЗМЕНЕНИЯ ГЕНОВ ДЕСАТУРАЦИИ ЖИРНЫХ КИСЛОТ У КОРЕННОГО НАСЕЛЕНИЯ СЕВЕРО-ВОСТОКА СИБИРИ

© 2021 г. Б. А. Малярчук[1], *, М. В. Деренко[1], Г. А. Денисова[1]

[1]Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, 685000 Россия

**e-mail: malyarchuk@ibpn.ru*

Поступила в редакцию 05.03.2021 г.
После доработки 12.05.2021 г.
Принята к публикации 27.05.2021 г.

В традиционной “арктической” диете коренных жителей крайнего Северо-Востока Сибири (эскимосов, чукчей, коряков) издревле преобладали мясо и жир ластоногих и китов, богатые омега-3 полиненасыщенными жирными кислотами. Это отразилось на особенностях липидного обмена северных аборигенов. Например, среди них преимущественно распространены варианты генов десатураз жирных кислот (*FADS1* и *FADS2*), кодирующие ферменты с пониженной активностью. В настоящей работе исследована распространенность 22 пн-инсерции гена *FADS2* (rs66698963), которая влияет на экспрессию гена *FADS1*, в популяциях северо-восточной (коряки, эвены) и южной (буряты) частей Сибири. У коряков обнаружена минимальная частота 22 пн-инсерции (5.6%) – почти на порядок реже, чем у бурят (45.3%). Кроме этого, у коряков зарегистрирована высокая частота (10.9%) нонсенс-мутации в гене *CYB5R2* (позиция 7694023 хромосомы 11), приводящей к преждевременной терминации синтеза NADH-цитохром b5 редуктазы, необходимой для функционирования десатураз, кодируемых *FADS*-генами. Выявленные генетические особенности коренного населения Крайнего Севера объясняются долговременной адаптацией к традиционной диете, очень богатой липидами, в связи с чем отпадает необходимость дополнительного синтеза полиненасыщенных жирных кислот с помощью десатураз жирных кислот.

Ключевые слова: гены десатураз жирных кислот, *FADS1*, *FADS2*, *CYB5R2*, популяции человека, коренное население Сибири.

DOI: 10.31857/S0016675821120109

Десатуразы жирных кислот, кодируемые генами *FADS1* и *FADS2*, катализируют синтез двойных связей в ацильных цепях жирных кислот. Известно, что *FADS*-гены возникли в результате генных дупликаций и образовали кластер генов в участке длиной примерно 100 тыс. пар нуклеотидов (пн) на хромосоме 11 человека (11q12-13.1) [1]. В функциональном отношении десатуразы жирных кислот имеют большое значение, поскольку они задействованы в процессе синтеза полиненасыщенных жирных кислот (**ПНЖК**), а последние участвуют в формировании оптимального физического состояния биологических мембран как в норме, так и в условиях температурного стресса. Предполагается, что при низкотемпературном стрессе концентрации ПНЖК в липидах мембран повышаются за счет усиления экспрессии генов десатураз ЖК и образования дополнительных количеств ПНЖК [2]. В связи с адаптацией к холоду большой интерес представляют исследования коренного населения Крайнего Севера (эскимосов, чукчей и коряков), издревле проживающего в крайне экстремальных условиях природной среды. Одним из проявлений экстремальности жизни на Крайнем Севере в прошлом был дефицит растительных углеводов и избыток липидов животного происхождения. В традиционной диете аборигенов Крайнего Севера преобладали мясо и жир ластоногих и китов, богатые омега-3 ПНЖК – главным образом эйкозапентаеновой и докозагексаеновой кислотами. Это отразилось на особенностях углеводного и липидного обмена коренных жителей крайнего Северо-Востока Сибири. Так, часть генетической вариабельности у них представлена вариантами полиморфизма, обусловившими недостаточность карнитин-пальмитоилтрансферазы 1А, панкреатической амилазы, сахаразы-изомальтазы, трегалазы [3].

Установлено, что в области расположения *FADS*-генов находятся два гаплоблока сцепления: гаплоблок 1, определяемый вариантами полиморфизма в гене *FADS1* и первой части гена *FADS2*, и гаплоблок 2, определяемый вариантами полиморфизма оставшейся последовательности

гена *FADS2* (от экзона 6 до 5'-нетранслируемой области) [4]. В пределах гаплоблока 1 выявлены два основных гаплотипа – A и D, варианты полиморфизма которых сцеплены на расстоянии 38.9 тыс. пн. Эти гаплотипы значительно различаются по уровню экспрессии и соответственно по способности кодируемых десатураз синтезировать длинноцепочечные ПНЖК. Показано, что гаплотипы A и D кодируют ферменты соответственно с пониженной и повышенной десатуразной активностью [4–6]. Распространение гаплотипа D, по мнению Ameur et al. [4], обусловлено необходимостью повышения эффективности синтеза жирных кислот в условиях их дефицита в пище (например, в Африке), а также может быть связано с выгодами в развитии головного мозга в условиях пищевого недостатка ПНЖК. В современных условиях, однако, носительство гаплотипа D может быть неблагоприятным в отношении развития атеросклероза [4]. В современных европейских популяциях преобладает гаплотип D, однако среди донеолитических охотников и собирателей Европы был распространен в основном гаплотип A. Установлено, что в последующие эпохи под действием отбора у европейцев стал более частым гаплотип D [7–9].

Популяционно-генетические исследования показали, что гаплотип A значительно реже встречается в африканских и европейских популяциях, но намного чаще, например, среди индейцев Америки [4, 8, 10, 11]. В популяциях коренного населения Сибири частота гаплотипа A существенно превалирует над частотой гаплотипа D в северной части Сибири [12]. Наиболее высокие частоты гаплотипа A зарегистрированы и среди эскимосов Чукотки, Аляски и Гренландии [6, 8, 13], а также американских индейцев [11, 19]. Обнаружено также, что у гренландских эскимосов и американских индейцев гены *FADS1* и *FADS2* находятся под действием положительного отбора, который способствовал почти полной фиксации гаплотипа A в этих популяциях [11, 13, 14].

Между тем в ряде других исследований арктических популяций Сибири (эскимосов, чукчей, коряков, нганасан, якутов) и Северной Америки (эскимосы) действие отбора на гены *FADS1* и *FADS2* не выявлено, несмотря на высокие частоты гаплотипа A в этих популяциях [15–18]. Тем не менее Amorim et al. [11] считают, что изначально сигнал положительного отбора в генах *FADS1* и *FADS2* возник у древнего населения Берингии, предкового по отношению к американским индейцам, в связи с необходимостью адаптироваться к холоду и ограниченным пищевым ресурсам, предоставленным Арктикой. Предполагается, что возникшие таким образом адаптивные генетические варианты сохранились в генофондах америндов и эскимосов [11]. Между тем за действие отбора у аборигенов Арктики/Америки и у европейцев ответственны разные комбинации вариантов полиморфизма в *FADS*-генах, поскольку у европейцев под действием отбора оказались локусы, расположенные в гаплоблоке 1, а у эскимосов – в основном локусы, расположенные в гаплоблоке 2 [7, 8, 11, 13, 19]. Это может быть обусловлено различными причинами возникновения отбора вариантов полиморфизма *FADS*-генов в популяциях: у европейцев – это сдвиг в сторону растительной диеты в после-неолитическое время, вследствие чего более благоприятными оказались D-гаплотипы с повышенной десатуразной активностью, а у предков эскимосов – это возникновение специфической “арктической” диеты с очень высоким содержанием ПНЖК, вследствие чего более предпочтительным оказалось поддержание максимальной частоты вариантов гаплотипа A с возможно еще более пониженной десатуразной активностью (в сравнении с европейскими A-вариантами) [19].

Еще одним геномным участком, задействованным в адаптации к растительной диете, является локус rs66698963, расположенный в интроне 1 гена *FADS2* рядом с регулируемым стеролами функциональным элементом SRE. Установлено, что инсерция длиной 22 пн в этом участке гена *FADS2* влияет на экспрессию гена *FADS1* [5]. Более того, предполагается, что благодаря именно этой инсерции гаплотип D кодирует фермент с повышенной десатуразной активностью [20]. Сигналы отбора для локуса rs66698963 зарегистрированы в различных популяциях Африки, Южной и Восточной Азии [5, 8, 20]. Так, в некоторых индийских популяциях, длительно практикующих вегетарианство, частота инсерции в локусе rs66698963 достигает 80–90%, в то время как в европейских популяциях ее частота составляет менее 50% [20]. Это объясняется тем, что 22 пн-инсерция помогает вегетарианцам обходиться без животной пищи, поскольку ее носители имеют больше метаболических возможностей для синтеза длинноцепочечных ПНЖК из биохимических предшественников [5]. Таким образом, следуя этой логике, в популяциях коренного населения Северо-Востока Сибири, диета которого длительное время была основана на потреблении продуктов морского зверобойного промысла, можно ожидать низкую частоту 22 пн-инсерции гена *FADS2*. Действительно, предварительные исследования показали, что ее частота у сибирских эскимосов составляет примерно 25% [8].

В последние годы накопились данные о том, что особенности “арктической” диеты (избыток липидов и белков при недостатке углеводов) существенно повлияли на генетические характеристики коренного населения Крайнего Севера. У эскимосов, чукчей и коряков обнаружен дефицит ферментов как углеводного обмена (гены *AMY2A, SI* и *TREH*), так и липидного обмена (гены *CPT1A*

и *CRAT*) [3]. Кроме этого, у гренландских эскимосов были выявлены связанные с гомеостазом глюкозы мутации в генах *TBC1D4* и *ADCY3*, которые повышают риск ожирения и диабета 2-го типа [21, 22]. Интересно, что в ряде случаев генетические изменения были обусловлены нонсенс-мутациями, приводящими к стоп-кодонам или к нарушению сплайсинга (для генов *SI*, *TBC1D4* и *ADCY3*), а также делецией целого гена (как в случае гена *AMY2A*). Появление и распространение такого рода мутаций в популяциях может быть связано с ослаблением отбора, возникшим из-за дефицита углеводных субстратов в традиционной диете северных аборигенов.

Известно, что избыток липидов в пище также привел к понижению активности некоторых ферментов (например, карнитин-ацилтрансфераз CPT1A и CrAT) в популяциях коренного населения Крайнего Севера [23, 24]. Исследования геномного полиморфизма показали также наличие и нонсенс-мутаций в генах липидного обмена у коренного населения Сибири [25, 26]. Так, у коряков и эскимосов была выявлена очень редкая нонсенс-мутация в гене *CYB5R2* (позиция 7694023 хромосомы 11), кодирующем NADH-цитохром b5 редуктазу [26]. Этот фермент относится к числу ключевых элементов в системе десатурации жирных кислот, поскольку именно NADH-редуктаза обеспечивает передачу электронов к десатуразам, кодируемым *FADS*-генами. Повреждение гена *CYB5R2*, по всей видимости, может способствовать снижению активности десатураз жирных кислот.

Целью настоящей работы является исследование распространенности 22 пн-инсерции (rs66698963) гена *FADS2* в популяциях Северо-Восточной Сибири – у коряков и эвенов северо-западного побережья Охотского моря, издавна занимавшихся прибрежным рыболовством, добычей тюленей и охотой, а также у бурят Южной Сибири – кочевников-скотоводов, в традиционной диете которых основное место занимают продукты животного и животно-растительного происхождения. Кроме этого, в настоящем исследовании проведен популяционный скрининг частоты нонсенс-мутации 11: 7694023 гена *CYB5R2* у коряков и охотских эвенов.

МАТЕРИАЛЫ И МЕТОДЫ

Изучены выборки коренного населения Северо-Восточной Сибири (коряки и эвены Северо-Эвенского района Магаданской области) и Южной Сибири (буряты из различных районов Республики Бурятия). Материалом для исследования служила геномная ДНК, выделенная из цельной крови с помощью стандартного метода, который включает в себя лизис клеток 1%-ным раствором додецилсульфата натрия и протеиназой К (Sigma, США) и последующую депротеинизацию фенолом и хлороформом. Генотипирование локуса rs66698963 гена *FADS2* проводили как описано ранее [5]. Для этого участок гена *FADS2*, включающий анализируемый локус, амплифицировали с помощью полимеразной цепной реакции (ПЦР), а затем продукты ПЦР фракционировали электрофоретически в 5%-ных полиакриламидных гелях. Для детекции ДНК использовали окраску гелей бромистым этидием с последующей визуализацией ДНК в УФ-свете.

ПЦР-фрагмент гена *FADS2* секвенировали также по Сэнгеру с помощью тех же олигонуклеотидных праймеров, которые использовали для проведения ПЦР, и наборов для циклического секвенирования ДНК Big Dye Terminator (Applied Biosystems v. 3.1) на генетическом анализаторе ABI Prism 3500xL (Applied Biosystems, США).

Нуклеотидную последовательность гена *CYB5R2*, включающую в свой состав локус 11: 7694023, амплифицировали с помощью пары олигонуклеотидных праймеров cybL (5'-GCCT-GAAACATCCTTCCTGT-3') и cybR (5'-GGCAG-GTTCTGTCAGCTTTC-3'). Нумерация нуклеотидов приводится согласно референтной последовательности генома человека GRCh37.p13 (hg19). Праймеры подобраны с помощью анализа нуклеотидной последовательности гена *CYB5R2* (под номером NC_000011.10 в базе данных GenBank) с использованием программы Primer3 [27]. Участки ДНК амплифицировали в течение 35 циклов в температурном режиме: 94°C – 30 с, 52°C – 60 с и 72°C – 60 с. Продукты ПЦР-амплификации секвенировали как описано выше.

Для выравнивания и анализа нуклеотидных последовательностей ДНК использовали программы пакета MEGA5 [28]. Частоту аллелей анализируемых локусов, гетерозиготность и соответствие распределения генотипов равновесию Харди–Вайнберга рассчитывали с помощью пакета программ Arlequin 3.01 [29].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Полиморфизм локуса rs66698963 гена FADS2

Секвенирование ПЦР-фрагмента гена *FADS2* у четырех коряков показало наличие делеционно-инсерционного полиморфизма в этом участке. В одном случае зарегистрирована гетерозигота с нормальным (*n*) и инсерционным (*i*) вариантами в локусе rs66698963 и в трех случаях – гомозиготы *nn*. В соответствии с полученными ранее данными [5] инсерция в локусе rs66698963 имеет длину 22 пн (рис. 1). Эта инсерция наблюдается в сочетании с дополнительной инсерцией трех нуклеотидов (CCA) в локусе rs138766446, расположенном на расстоянии 53 пн от локуса rs66698963, что также ранее отмечалось Reardon et al. [5] (рис. 1).

n TGCCCCC---CCCCAGCCCTCCTCCCTGCCTGCCCCCAGGGCCTCCTCCCTGCCTCCCCAG

i TGCCCCCCCCACCCCAGCCCTCCTCCCTGCCTGCCCCCAGGGCCTCCTCCCTGCCTCCCCAG

n GG----------ACTTCT

i GGACTTCTCCCTGCCTCCCCAGGGACTTCT

Рис. 1. Нуклеотидные последовательности гена *FADS2* между позициями 61 834 980 и 61 835 070 хромосомы 11 для вариантов *n* (норма) и *i* (инсерция) локуса rs66698963.

Электрофоретический анализ ПЦР-продуктов гена *FADS2* в 5%-ных полиакриламидных гелях позволил определить распространенность аллелей и генотипов по локусу rs66698963 в трех выборках коренного населения Сибири (табл. 1). Электрофоретический анализ не позволяет надежно зарегистрировать дополнительную 3 пн-инсерцию в локусе rs138766446 и поэтому полученные нами результаты (как и в других работах [5, 20]) описываются далее лишь в отношении 22 пн-инсерции в локусе rs66698963.

Как видно из табл. 1, частота 22 пн-инсерции увеличивается в ряду популяций коряков (5.6%), эвенов (13.6%) и бурят (45.3%). Геномные исследования, проведенные ранее в рамках проекта “1000 геномов” (http://www.1000genomes.org), показали, что частота 22 пн-инсерции в локусе rs66698963 варьирует от очень высоких значений у населения Южной Азии (84%) и Африки (74%) до более умеренных в Восточной Азии (56%), Европе (46%) и Америке (44%). Согласно опубликованным ранее результатам анализа геномных данных, самые низкие частоты этой инсерции наблюдались у сибирских эскимосов (~25%) [8]. Между тем полученные нами результаты анализа полиморфизма участка гена *FADS2*, включающего в себя локус rs66698963, показывают, что на Северо-Востоке Сибири частота 22 пн-инсерции еще ниже (5.6%). На юге Сибири у бурят эта инсерция отмечается с частотой, лишь немного меньшей, чем в региональной восточноазиатской выборке проекта “1000 геномов”.

Низкую частоту 22 пн-инсерции гена *FADS2* у коряков можно объяснить долговременной адаптацией к традиционной “арктической” диете, очень богатой липидами, в связи с чем отпадает необходимость дополнительного синтеза ПНЖК. Предполагается, что предки коряков Северного Приохотья усвоили основы культуры морского зверобойного промысла от предков эскимосов еще несколько тысячелетий назад и, по всей видимости, даже ассимилировали какие-то группы эскимосов [30].

У охотских эвенов также обнаружена не столь высокая, как у бурят, частота 22 пн-инсерции, что может быть обусловлено длительным соседством и брачными контактами между коряками и эвенами Северного Приохотья на протяжении нескольких веков. Известно, что эвены – кочевники-оленеводы, появились на побережье Охотского моря, занятом коряками, в середине XVII в. [31]. В процессе межэтнических взаимодействий часть коряков вошла в состав эвенских родов, что естественно повлияло на генетические характеристики эвенов и сблизило их с коряками.

Частота нонсенс-мутации 11: 7694023 гена CYB5R2

Исследование полногеномной изменчивости у 102 представителей коренного населения Сибири позволило выявить новый вариант полиморфизма – нонсенс-мутацию G → A гена *CYB5R2* в позиции 7694023 хромосомы 11, приводящую к стоп-кодону Gln12Stop в NADH-цитохром b5 редуктазе [25, 26]. Этот вариант полиморфизма был обнаружен только в выборках коряков ($N = 16$) и эскимосов ($N = 4$) с частотой 25%. Анализ опуб-

Таблица 1. Распределение частот генотипов и аллелей локуса rs66698963 в некоторых популяциях Сибири

Популяция	Генотипы			Аллели		H_o	H_e	HWE, p
	nn	*ni*	*ii*	*n*	*i*			
Коряки ($N = 71$)	0.887	0.113	0	0.944	0.056	0.112	0.107	1.0
Эвены ($N = 70$)	0.743	0.243	0.014	0.864	0.136	0.243	0.236	1.0
Буряты ($N = 74$)	0.27	0.554	0.176	0.547	0.453	0.554	0.499	0.36

Примечание. *n* – нормальный вариант; *i* – инсерционный вариант локуса rs66698963. H_o и H_e – наблюдаемая и ожидаемая гетерозиготность; HWE, p – статистическая значимость отклонения от равновесия Харди–Вайнберга (достоверно при $p < 0.05$).

Таблица 2. Распределение частот генотипов и аллелей локуса 11: 7694023 гена *CYB5R2* у коряков и эвенов

Популяция	Генотипы			Аллели		H_o	H_e	HWE, p
	GG	*GA*	*AA*	*G*	*A*			
Коряки ($N = 78$)	0.807	0.167	0.026	0.891	0.109	0.167	0.195	0.211
Эвены ($N = 68$)	0.971	0.029	0 (0)	0.985	0.015	0.029	0.029	1.0

Примечание. H_o и H_e – наблюдаемая и ожидаемая гетерозиготность; HWE, p – статистическая значимость отклонения от равновесия Харди–Вайнберга (достоверно при $p < 0.05$).

ликованных данных показывает отсутствие мутации 11: 7694023 у более чем 125000 представителей различных этнорасовых групп (база данных gnomAD v2.1.1 (https://gnomad.broadinstitute.org)). Однократно эта мутация была зарегистрирована только у корейских пациентов с раком легких [32]; она зарегистрирована под номером COSM1188159 в каталоге соматических мутаций при раке COSMIC (https://cancer.sanger.ac.uk/cosmic).

Для оценки распространенности мутации G → A в позиции 7694023 хромосомы 11 на Северо-Востоке Сибири нами проведен скрининг ее частоты у коряков и охотских эвенов (табл. 2). В результате установлено, что частота А-варианта у коряков составляет 10.9%, а у эвенов всего лишь 1.5%. Среди коряков обнаружены два индивидуума, гомозиготных по этой мутации (частота 2.6%).

Мутация 11: 7694023 интересна тем, что она имеет непосредственное отношение к процессу десатурации жирных кислот. Эта мутация произошла в гене *CYB5R2*, кодирующем NADH-цитохром b5 редуктазу. Данный фермент выполняет в организме целый ряд функций, связанных с десатурацией и элонгацией жирных кислот, биосинтезом холестерина, клеточным дыханием, биотрасформацией ксенобиотиков. NADH-цитохром b5 редуктаза является центральным элементом в системе биохимических реакций десатурации жирных кислот, поскольку именно NADH-редуктаза обеспечивает передачу электронов к десатуразам, кодируемым *FADS*-генами. Таким образом, обнаруженная в популяциях коренного населения Северо-Востока Сибири замена Gln12Stop в NADH-цитохром b5 редуктазе, по всей видимости, терминирует синтез этого белка. В таком случае недостаток NADH-цитохром b5 редуктазы может привести к снижению активности десатураз жирных кислот.

Результаты проведенного исследования позволяют предположить, что, по всей видимости, избыток ПНЖК в традиционной диете северных аборигенов привел к ненужности дополнительного синтеза жирных кислот, что проявилось в отборе ферментов со сниженной активностью – как десатураз, кодируемых *FADS*-генами, так и других участников этого биохимического процесса – например, NADH-цитохром b5 редуктазы. Важно отметить, что если в прошлом ферменты метаболизма ПНЖК со сниженной активностью могли иметь адаптивное значение для коренных северных народов, то в настоящее время при переходе от традиционной диеты к “европейскому” типу питания их присутствие может негативно повлиять на здоровье и стать причиной развития атеросклероза и воспалительных заболеваний в популяциях коренного населения Крайнего Севера. Это обусловлено тем, что для диеты “европейского” типа характерно снижение в рационе относительной доли омега-3 жирных кислот и отношения омега-3/омега-6 кислот.

В исследованиях эскимосов Гренландии было обнаружено, что аллели гаплотипа А, находящиеся под воздействием отбора, ассоциируются с антропометрическими фенотипами (ростом и весом) и метаболическими характеристиками (уровни инсулина и холестерина в крови), а также повлияли на жирнокислотный спектр в мембранах эритроцитов [13]. Таким образом, ожидается, что гаплотипы со сниженной активностью десатураз могли играть протективную роль в отношении атеросклероза и воспалительных процессов [4, 13, 33]. Однако результаты подобного рода клинико-генетических исследований пока довольно противоречивы [34–36], в связи с чем необходимы дальнейшие исследования биохимических и физиологических последствий генетически детерминированного снижения активности десатураз жирных кислот.

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

1. *Marquardt A., Stöhr H., White K., Weber B.H.* cDNA cloning, genomic structure, and chromosomal localization of three members of the human fatty acid desat-

urase family // Genomics. 2000. V. 66. P. 175–183. https://doi.org/10.1006/geno.2000.6196

2. *Лось Д.А.* Структура, регуляция экспрессии и функционирование десатураз жирных кислот // Успехи биол. химии. 2001. Т. 41. С. 163–198.
3. *Малярчук Б.А.* Долговременные ген-средовые взаимодействия и генетика нарушений метаболизма в популяциях коренного населения Северо-Востока Азии // Экол. генетика. 2018. Т. 16. № 2. С. 30–35. https://doi.org/10.17816/ecogen16230-35
4. *Ameur A., Enroth S., Johansson A. et al.* Genetic adaptation of fatty-acid metabolism: A human specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids // Am. J. Hum. Genet. 2012. V. 90. P. 809–820. https://doi.org/10.1016/j.ajhg.2012.03.014
5. *Reardon H.T., Zhang J., Kothapalli K.S. et al.* Insertion-deletions in a FADS2 intron 1 conserved regulatory locus control expression of fatty acid desaturases 1 and 2 and modulate response to simvastatin // Prostaglandins Leukot. Essent. Fatty Acids. 2012. V. 87. P. 25–33. https://doi.org/10.1016/j.plefa.2012.04.011
6. *Voruganti V.S., Higgins P.B., Ebbesson S.O. et al.* Variants in CPT1A, FADS1, and FADS2 are associated with higher levels of estimated plasma and erythrocyte Delta-5 desaturases in Alaskan Eskimos // Front. Genet. 2012. V. 3. 86. eCollection 2012. https://doi.org/10.3389/fgene.2012.00086
7. *Mathieson I., Lazaridis I., Rohland N. et al.* Genome-wide patterns of selection in 230 ancient Eurasians // Nature. 2015. V. 528. P. 499–503. https://doi.org/10.1038/nature16152
8. *Ye K., Gao F., Wang D. et al.* Dietary adaptation of FADS genes in Europe varied across time and geography // Nat. Ecol. Evol. 2017. V. 1. 0167. https://doi.org/10.1038/s41559-017-0167
9. *Mathieson S., Mathieson I.* FADS1 and the timing of human adaptation to agriculture // Mol. Biol. Evol. 2018. V. 35. P. 2957–2970. https://doi.org/10.1093/molbev/msy180
10. *Mathias R.A., Fu W., Akey J.M. et al.* Adaptive evolution of the FADS gene cluster within Africa // PLoS One. 2012. V. 7. e44926. https://doi.org/10.1371/journal.pone.0044926
11. *Amorim C.E., Nunes K., Meyer D. et al.* Genetic signature of natural selection in first Americans // Proc. Natl Acad. Sci. USA. 2017. V. 114. P. 2195–2199. https://doi.org/10.1073/pnas.1620541114
12. *Малярчук Б.А., Дереико М.В.* Полиморфизм генов метаболизма полиненасыщенных жирных кислот (FADS1 и FADS2) у коренного населения Сибири // Вестник Северо-Восточного науч. центра ДВО РАН. 2018. № 3. С. 106–111.
13. *Fumagalli M., Moltke I., Grarup N. et al.* Greenlandic Inuit show genetic signatures of diet and climate adaptation // Science. 2015. V. 349. P. 1343–1347. https://doi.org/10.1126/science.aab2319
14. *Harris D.H., Ruczinski I., Yanek L.R. et al.* Evolution of hominin polyunsaturated fatty acid metabolism: from Africa to the New World // Genome Biol. Evol. 2019. V. 11. P. 1417–1430. https://doi.org/10.1093/gbe/evz071
15. *Cardona A., Pagani L., Antao T. et al.* Genome-wide analysis of cold adaption in indigenous Siberian populations // PLoS One. 2014. V. 9. e98076. eCollection 2014. https://doi.org/10.1371/journal.pone.0098076
16. *Clemente F.J., Cardona A., Inchley C.E. et al.* A selective sweep on a deleterious mutation in the CPT1A gene in Arctic populations // Am. J. Hum. Genet. 2014. V. 95. P. 584–589. https://doi.org/10.1016/j.ajhg.2014.09.016
17. *Hsieh P., Hallmark B., Watkins J. et al.* Exome sequencing provides evidence of polygenic adaptation to a fat-rich animal diet in indigenous Siberian populations // Mol. Biol. Evol. 2017. V. 34. P. 2913–2926. https://doi.org/10.1093/molbev/msx226
18. *Reynolds A.W., Mata-Míguez J., Miró-Herrans A. et al.* Comparing signals of natural selection between three Indigenous North American populations // Proc. Natl Acad. Sci. USA. 2019. V. 116. P. 9312–9317. https://doi.org/10.1073/pnas.1819467116
19. *Mathieson I.* Limited evidence for selection at the FADS locus in Native American populations // Mol. Biol. Evol. 2020. V. 37. P. 2029–2033. https://doi.org/10.1093/molbev/msaa064
20. *Kothapalli K.S.D., Ye K., Gadgil M.S. et al.* Positive selection on a regulatory insertion-deletion polymorphism in FADS2 influences apparent endogenous synthesis of arachidonic acid // Mol. Biol. Evol. 2016. V. 33. P. 1726–1739. https://doi.org/10.1093/molbev/msw049
21. *Moltke I., Grarup N., Jørgensen M.E. et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes // Nature. 2014. V. 512. P. 190–193. https://doi.org/10.1038/nature13425
22. *Grarup N., Moltke I., Andersen M.K. et al.* Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes // Nat. Genet. 2018. V. 50. P. 172–174. https://doi.org/10.1038/s41588-017-0022-7
23. *Greenberg C.R., Dilling L.A., Thompson G.R. et al.* The paradox of the carnitine palmitoyltransferase type Ia P479L variant in Canadian Aboriginal populations // Mol. Genet. Metab. 2009. V. 96. P. 201–207. https://doi.org/10.1016/j.ymgme.2008.12.018
24. *Zhou S., Xiong L., Xie P. et al.* Increased missense mutation burden of fatty acid metabolism related genes in Nunavik Inuit population // PLoS One. 2015. V. 10. e0128255. https://doi.org/10.1371/journal.pone.0128255
25. *Pagani L., Lawson D.J., Jagoda E. et al.* Genomic analyses inform on migration events during the peopling of Eurasia // Nature. 2016. V. 538. P. 238–242. https://doi.org/10.1038/nature19792
26. *Mörseburg A.* Investigating the role of demography and selection in genome scale patterns of common and rare variant diversity in humans: Ph. D. thesis. Cambridge: Univ. Cambr., 2018. 438 p.
27. *Untergasser A., Cutcutache I., Koressaar T. et al.* Primer3 – new capabilities and interfaces // Nucl. Acids Res. 2012. V. 40. e115. https://doi.org/10.1093/nar/gks596

28. *Tamura K., Peterson D., Peterson N. et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods // Mol. Biol. Evol. 2011. V. 28. P. 2731–2739. https://doi.org/10.1093/molbev/msr121
29. *Excoffier L., Laval G., Schneider S.* Arlequin (version 3.0): An integrated software package for population gene-tics data analysis // Evol. Bioinform. Online. 2007. V. 1. P. 47–50.
30. *Васильевский Р.С.* Происхождение и древняя культура коряков. Новосибирск: Наука, 1971. 250 с.
31. Народы Северо-Востока Сибири / Под ред. Батьяновой Е.П., Тураева В.А. М.: Наука, 2010. 773 с.
32. *Seo J.S., Ju Y.S., Lee W.C. et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma // Genome Res. 2012. V. 22. P. 2109–2119. https://doi.org/10.1101/gr.145144.112
33. *Martinelli N., Girelli D., Malerba G. et al.* FADS genotypes and desaturase activity estimated by the ratio of arachidonic acid to linoleic acid are associated with inflammation and coronary artery disease // Am. J. Clin. Nutr. 2008. V. 88. P. 941–949. https://doi.org/10.1093/ajcn/88.4.941
34. *Lattka E., Illig T., Heinrich J., Koletzko B.* Do FADS genotypes enhance our knowledge about fatty acid related phenotypes? // Clin. Nutr. 2010. V. 29. P. 277–287. https://doi.org/10.1016/j.clnu.2009.11.005
35. *Chen Y., Estampador A.C., Keller M. et al.* The combined effects of FADS gene variation and dietary fats in obesity-related traits in a population from the far north of Sweden: the GLACIER study // Int. J. Obes. (Lond.). 2019. V. 43. P. 808–820. https://doi.org/10.1038/s41366-018-0112-3
36. *Del Gobbo L.C., Imamura F., Aslibekyan S. et al.* ω-3 polyunsaturated fatty acid biomarkers and coronary heart disease: pooling project of 19 cohort studies // JAMA Intern. Med. 2016. V. 176. P. 1155–1166. https://doi.org/10.1001/jamainternmed.2016.2925

Adaptive Changes of Fatty Acid Desaturation Genes in Indigenous Populations of Northeast Siberia

B. A. Malyarchuk[a], *, M. V. Derenko[a], and G. A. Denisova[a]

[a] *Institute of Biological Problems of the North Far-East Branch of Russian Academy of Sciences, Magadan, 685000 Russia*
**e-mail: malyarchuk@ibpn.ru*

Since ancient times meat and fat of seals and whales, rich in omega-3 polyunsaturated fatty acids, prevailed in the traditional "Arctic" diet of the indigenous people of the far North-East of Siberia (Eskimos, Chukchi, Koryaks). This was reflected in the peculiarities of lipid metabolism of northern aborigines – for example, among them, variants of the genes for fatty acid desaturases (*FADS1* and *FADS2*) encoding enzymes with reduced activity are predominantly widespread. In this study, we investigated the distribution of 22 bp insertion of the *FADS2* gene (rs66698963), which affects the expression of the *FADS1* gene, in populations of the northeastern (Koryaks, Evens) and southern (Buryats) parts of Siberia. The Koryaks had a minimum frequency of 22 bp insertions (5.6%) – almost an order of magnitude less frequently than Buryats (45.3%). In addition, a high frequency (10.9%) of nonsense mutation in the *CYB5R2* gene (position 7694023 of chromosome 11) was detected in the Koryaks. This mutation leads to premature termination of the synthesis of NADH cytochrome b5 reductase, which is necessary for the functioning of desaturases encoded by *FADS* genes. The revealed genetic features of the indigenous population of the Far North are explained by long-term adaptation to a traditional diet very rich in lipids, which is a reason that prevents the need for additional synthesis of polyunsaturated fatty acids using fatty acid desaturases.

Keywords: fatty acid desaturase genes, *FADS1*, *FADS2*, *CYB5R2*, human populations, indigenous populations of Siberia.