

УДК 575.1

## ГЕНОГЕОГРАФИЧЕСКИЙ АТЛАС ДНК-МАРКЕРОВ, КОНТРОЛИРУЮЩИХ ЦВЕТ ГЛАЗ И ВОЛОС ЧЕЛОВЕКА<sup>1</sup>

© 2021 г. Е. В. Балановская<sup>1, 2, \*</sup>, И. О. Горин<sup>1, 3</sup>, С. М. Кошель<sup>1, 4</sup>, **О. П. Балановский**<sup>1, 2, 3</sup>

<sup>1</sup>Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, 115478 Россия

<sup>2</sup>АНО “Биобанк Северной Евразии”, Москва, 115215 Россия

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

<sup>4</sup>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 HIGisPlex 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 измененной 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-

<sup>1</sup> 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 взаимосвязанных 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)	<i>N</i>	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 рас-

считаны 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]:

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

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 1129038	HERC2 12913832	HERC2 12916300	TP63 6804480	SLC45A2 35407	MC1R 1805007	HERC2 2238289	DEF8 8051733	EDAR 3827760	SPATA33 71396951	SLC45A2 28777	RALY 6059655	LOC100996492 12915889	RUNX2 2820339	GLBI 72856153
Аллели	T	G	T	T	G	T	A <sup>R</sup>	G	A <sup>R</sup>	C	A	A <sup>R</sup>	T <sup>R</sup>	G	T <sup>R</sup>
Суммарный генетический вклад (C)	<b>2.03</b>	<b>0.68</b>	<b>0.45</b>	<b>0.56</b>	<b>0.63</b>	<b>0.61</b>	<b>0.40</b>	<b>0.39</b>	<b>0.38</b>	<b>0.47</b>	<b>0.45</b>	<b>0.32</b>	<b>0.31</b>	<b>0.31</b>	<b>0.40</b>
Вклад в светлые глаза (C <sub>E</sub> )	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 <sub>H</sub> )	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 <sub>R</sub> )	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
№	Частота, %														
1	51	3	3	21	18	0	64	9	9	0	18	1	59	75	46
2	48	1	1	31	13	0	68	21	1	2	13	0	48	86	51
3	54	2	2	19	12	0	56	40	12	0	12	0	71	74	50
4	53	5	5	13	19	3	65	43	22	0	19	3	46	60	49
5	50	17	15	33	27	0	70	27	49	0	31	1	54	82	67
6	50	6	4	30	22	1	58	23	33	4	24	4	66	74	57
7	51	46	46	24	74	3	69	23	50	3	73	4	57	64	61
8	72	13	14	32	44	1	61	21	49	1	42	2	55	69	66
9	52	17	17	55	68	1	61	28	97	1	69	8	52	65	70
10	61	27	28	31	90	7	69	23	93	8	91	4	59	63	81
11	50	36	36	44	90	4	72	27	99	2	90	3	49	67	83
12	49	30	30	35	71	2	73	30	97	2	71	0	55	60	88
13	51	79	79	44	100	5	97	35	96	2	100	5	61	64	88
14	55	83	81	44	91	4	91	27	97	6	91	8	72	68	88
15	50	49	49	47	94	5	79	30	99	5	93	2	62	73	87
16	50	89	93	39	98	7	100	34	95	6	98	1	69	66	93

Таблица 2. Окончание

Гены	HERC2		T	HERC2	HERC2	T	HERC2	TP63	SLC45A2	MC1R	HERC2	DEF8	EDAR	SPATA33	SLC45A2	RALY	LOC100996492	RUNX2	GLBI
	HERC2	HERC2																	
rs	1129038	12913832	12916300	6804480	35407	1805007	2238289	8051733	3827760	71396951	28777	6059655	12915889	2820339	72856153				
Аллели	T	G	T	T	G	T	A <sup>R</sup>	G	A <sup>R</sup>	C	A	A <sup>R</sup>	T <sup>R</sup>	G	T <sup>R</sup>				
17	52	85	85	56	100	4	95	14	98	6	96	8	68	64	93				
18	50	78	78	42	96	0	88	32	99	1	96	4	66	59	90				
19	50	62	56	33	77	3	83	21	81	3	77	1	65	64	85				
20	50	74	72	31	83	2	86	28	90	1	83	3	63	60	76				
21	50	71	71	45	93	13	86	37	93	13	93	5	74	72	75				
22	53	38	38	36	64	2	73	23	73	2	64	2	60	61	73				
Показатели частоты																			
$q_{\text{MIN}}$	1	1	1	13	12	0	56	9	22	0	12	0	46	59	46				
$q_{\text{MAX}}$	89	93	91	56	100	13	100	43	99	13	100	8	74	86	93				
$q_{\text{MEAN}}$	41	41	41	36	66	3	76	27	70	3	66	3	60	68	73				
Генетическое разнообразие																			
$G_{\text{ST}}$	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				
$H_{\text{T}}$	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				
$H_{\text{S}}$	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 HigusPlex	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);  $q_{\text{MIN}}$  – минимальное значение частоты;  $q_{\text{MAX}}$  – максимальное значение частоты;  $q_{\text{MEAN}}$  – среднее значение частоты;  $G_{\text{ST}}$  – показатель межпопуляционного разнообразия;  $H_{\text{T}}$  – показатель общего генетического разнообразия;  $H_{\text{S}}$  – показатель среднего внутрипопуляционного разнообразия; H – ДНК-маркер входит в зарубежную криминалистическую панель HigusPlex; H-S – ДНК-маркер входит в зарубежную криминалистическую панель HigusPlex-S.

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

Гены	LC105370627	PRDM16	ANKRD11	AFG3L1P	-	OCA2	MC1R	SLC24A5	FANCA	TYR	HERC2	CACNA2D3	TYR	C19orf45	SLC24A4	-	
																	T
rs	12896399	1999528	3114908	4238832	34191540	2871886	2228479	1426654	75570604	1393350	7494942	11283625	1042602	685034	12588868	11228497	
Аллели	T	G	T <sup>R</sup>	C <sup>R</sup>	T	C	G <sup>R</sup>	A <sup>R</sup>	C	A	G	A	A	C	T <sup>R</sup>	A <sup>R</sup>	
Суммарный генетический вклад (C)	<b>0.27</b>	<b>0.26</b>	<b>0.25</b>	<b>0.18</b>	<b>0.18</b>	<b>0.16</b>	<b>0.12</b>	<b>0.10</b>	<b>0.10</b>	<b>0.10</b>	<b>0.06</b>	<b>0.03</b>	<b>0.03</b>	<b>0.02</b>	<b>0.02</b>	<b>0.01</b>	
Вклад в светлые глаза (C <sub>E</sub> )	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 <sub>H</sub> )	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 <sub>R</sub> )	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	
№	Частота, %																
1	51	27	49	40	1	69	74	92	20	1	0	45	28	1	60	24	69
2	48	33	32	33	0	19	76	78	5	1	0	50	25	1	50	16	82
3	54	30	34	37	3	38	83	64	9	3	1	25	24	1	49	19	76
4	53	11	46	48	6	49	50	67	44	10	4	42	28	3	67	16	68
5	50	28	30	33	2	53	49	74	66	4	2	55	35	8	54	27	78
6	50	29	36	27	3	27	57	83	24	4	4	34	38	7	55	19	77
7	51	33	47	17	8	68	57	83	76	3	8	58	26	9	77	36	87
8	72	28	42	31	11	49	49	90	52	4	11	44	29	19	51	39	83
9	52	33	42	20	22	57	49	93	92	1	18	38	29	28	52	36	88
10	61	37	59	33	20	67	45	98	95	7	18	49	43	48	61	40	92
11	50	8	45	34	19	100	39	90	99	5	26	50	53	35	70	46	94
12	49	22	45	31	23	67	42	95	99	2	18	47	49	47	61	45	93
13	51	40	56	36	29	76	30	88	99	2	24	93	52	23	59	53	92
14	55	42	57	34	25	64	25	88	96	5	23	84	38	22	61	60	91
15	50	37	64	21	22	63	25	90	100	6	19	70	40	43	68	49	93
16	50	60	48	35	25	73	19	86	99	7	31	98	50	18	57	60	97

Таблица 3. Окончание

Гены	12896399 LOC105370627		G	T <sup>R</sup>	C <sup>R</sup>	T	C	G <sup>R</sup>	A <sup>R</sup>	C	A	G	HERC2	CACNA2D3	TYR	C19orf45	T <sup>R</sup>	A <sup>R</sup>
	T	52																
ts			1999528	3114908	4238832	34191540	2871886	2228479	1426654	75570604	1393350	7494942	11283625	1042602	685034	12588868	11228497	—
Алели																		
17	Восточные славяне	50	52	24	18	77	25	93	99	5	22	86	51	26	59	54	91	
18	Мордовия	50	43	35	19	100	31	82	98	0	31	84	37	29	65	54	96	
19	Тюрки Приуралья	50	45	28	12	60	26	89	91	3	12	70	39	22	58	56	86	
20	Марий Эл	50	43	32	28	78	24	91	97	13	30	82	29	20	51	48	84	
21	Пермь	50	27	29	20	100	39	92	96	2	12	84	35	25	49	53	85	
22	Южный Урал	53	27	38	18	57	38	89	83	3	14	61	38	21	53	38	87	
Показатели частоты																		
$\eta_{\text{MIN}}$		8	30	17	0	19	19	64	5	0	0	25	24	1	49	16	68	
$\eta_{\text{MAX}}$		60	64	48	29	100	83	98	100	13	31	98	53	48	77	60	97	
$\eta_{\text{MEAN}}$		32	46	32	15	64	43	86	75	4	15	61	37	21	58	40	86	
Генетическое разнообразие																		
$G_{\text{ST}}$		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	
$H_{\text{T}}$		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	
$H_{\text{S}}$		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 NIrusPlex		<b>H-S</b>		<b>H</b>				<b>H-S</b>	<b>H</b>		<b>H-S</b>			<b>H-S</b>				

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 “К”, минимальное “min”, максимальное “max” и среднее “avg” значения признака, рассчитанные по цифровой модели карты. Аналогичные статистические показатели рассчитаны по информации табл. 2 и 3 и приведены в этих таблицах: минимальные  $q_{\text{MIN}}$ , максимальные  $q_{\text{MAX}}$  и средние  $q_{\text{MEAN}}$  частоты, показатель межпопуляционного разнообразия  $G_{\text{ST}}$ , общего генетического разнообразия  $H_{\text{T}}$ , среднего внутривнутрипопуляционного разнообразия  $H_{\text{S}}$ . 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

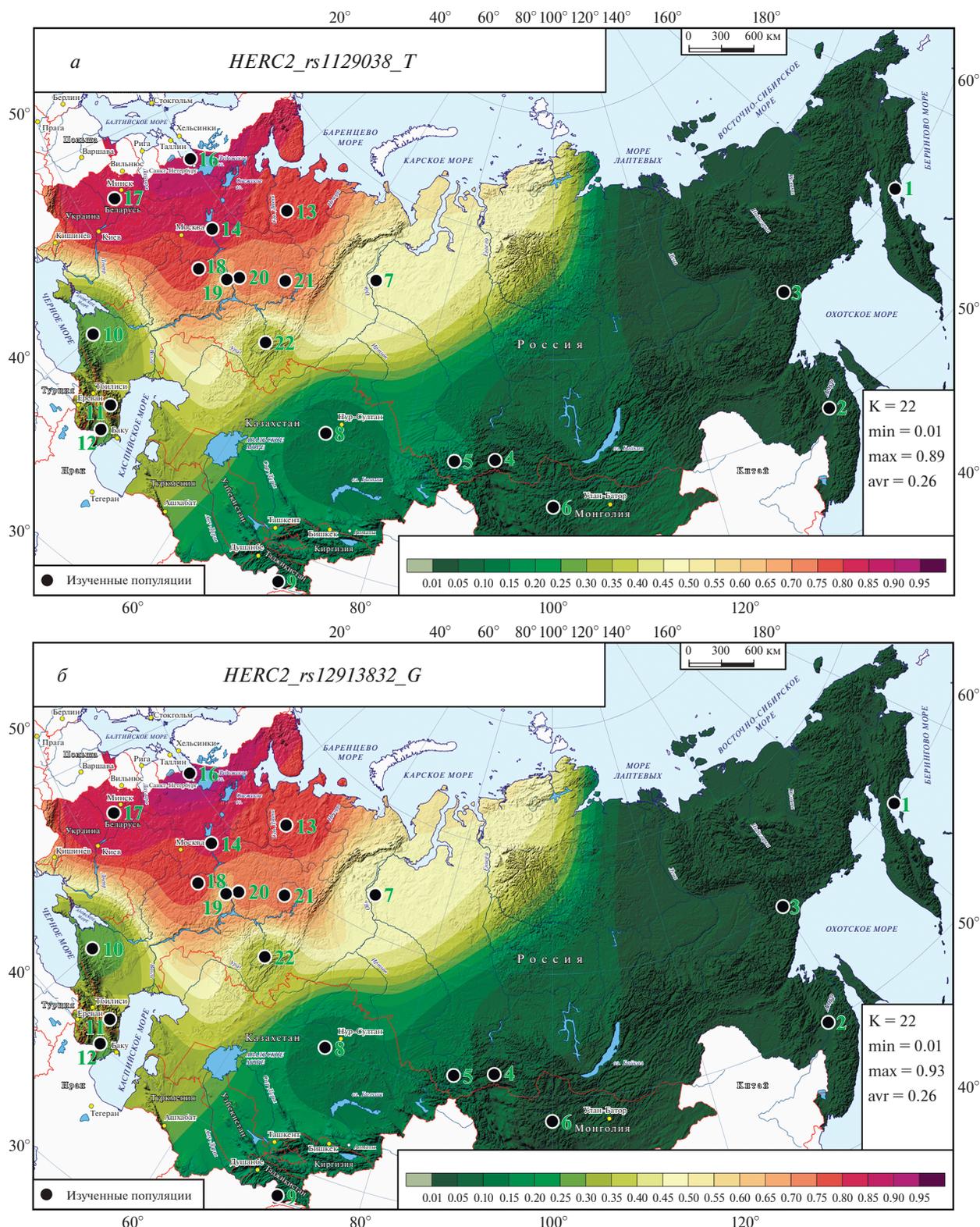
Для изучения распространения генетических маркеров цвета глаз выбраны 22 метапопуляции, охватывающие основное разнообразие народонаселения Северной Евразии, и сформирована выборка из 1153 образцов ДНК (в среднем 53 образца из метапопуляции). Их генотипирование по 31 SNP-маркеру, входящему в отечественную 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 ( $G_{\text{ST}} = 0.39$ ). 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 ( $C_{\text{E}} = 0.68$ ), то первый не только 2 раза мощнее контролирует светлый цвет глаз ( $C_{\text{E}} = 1.31$ ), но крайне велико и его влияние на формирование светлого цвета волос ( $C_{\text{H}} = 0.71$ ). 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 ( $1 \leq q \leq 6\%$ ) охватывают коренное население Дальнего Востока, Монголии и Южной Сибири. На Алтае-Саянах, в Казахстане и Таджикистане они несколько выше ( $13 \leq q \leq 17\%$ ). Еще выше частоты “осветления” на Кавказе, в Закавказье и в степях Южного Урала ( $27 \leq q \leq 38\%$ ). 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 ( $62 \leq q \leq 74\%$ ). Максимальные значения охватывают ареал от Мордовии ( $q = 78\%$ ) до западных финноязычных популяций карел и ижоры ( $q = 93\%$ ). Популяции всех восточных славян занимают промежуточное положение ( $79 \leq q \leq 85\%$ ) между западными и приуральскими финно-уграми.

Третий ДНК-маркер этого гена — *HERC2\_rs12916300\_T* — обладает только вкладом в формирование светлых глаз ( $C_{\text{E}} = 0.45$ ) 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 К – число метапопуляций; min, max, avg – 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.

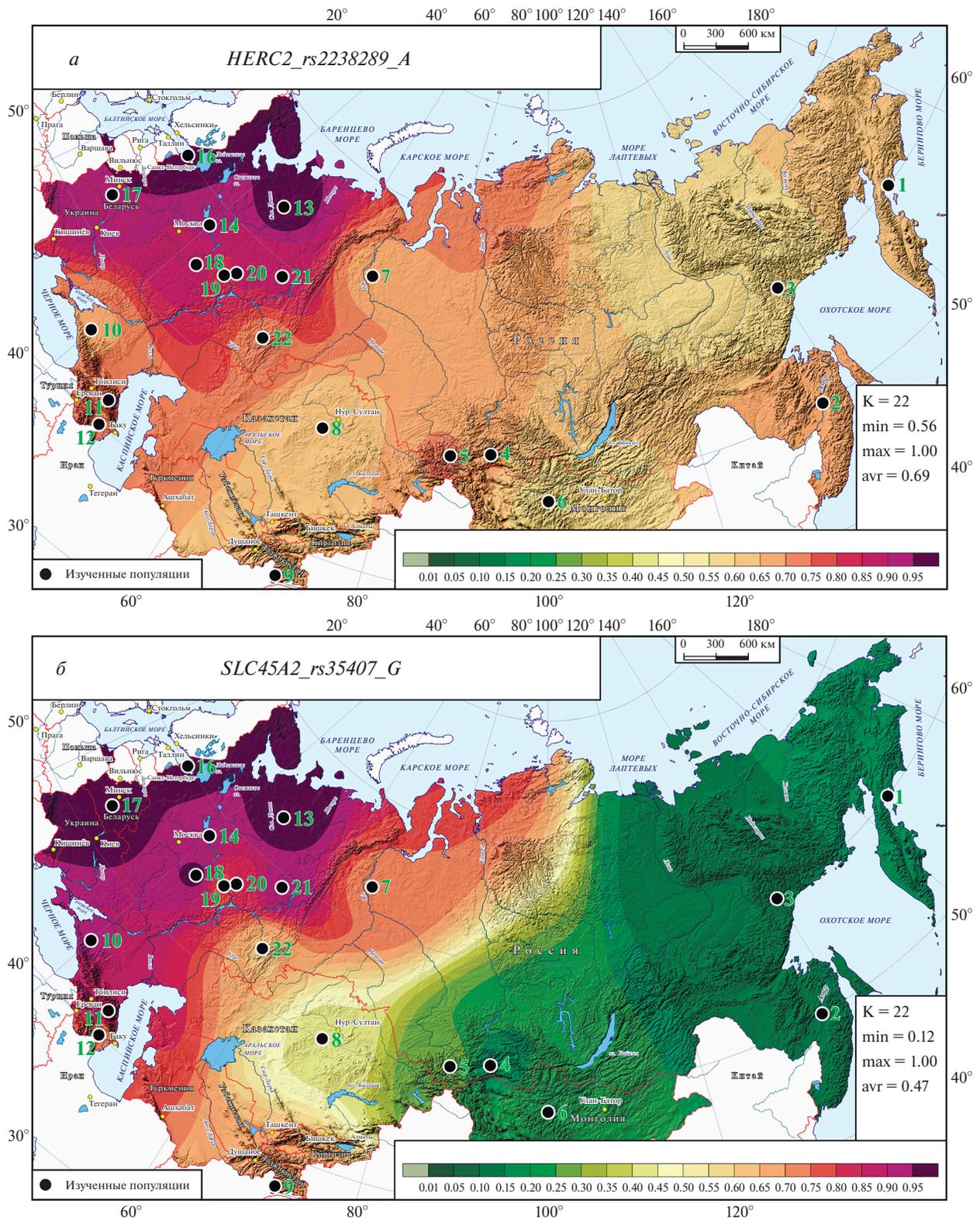


Рис. 2. Карты распространения ДНК-маркеров *HERC2\_rs2238289\_A* (а) и *SLC45A2\_rs35407\_G* (б) в метапопуляциях России и сопредельных стран.



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 *NIrusPlex* (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 ( $56 \leq q \leq 100\%$ ) 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 ( $12 \leq q \leq 100\%$ ), 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 ( $12 \leq q \leq 27\%$ ) 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 ( $77 \leq q \leq 83\%$ ), 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 вы-

сокую 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 ( $91 \leq q \leq 100\%$ ), 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 ( $13 \leq q \leq 56\%$ ), 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 ( $0 \leq q \leq 13\%$ ,  $G_{ST} = 0.032$ ). 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 *NIrusPlex-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 ( $9 \leq q \leq 43\%$ ) 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 ( $22 \leq q \leq 99\%$ ) 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 ( $1 \leq q \leq 12\%$ ), а уже начиная с Приуралья и далее на запад ДНК-маркер достигает максимальной частоты ( $90 \leq q \leq 99\%$ ). Но не менее яркой особенностью распространения этого ДНК-маркера, контролирующего светлый цвет волос, являются максимальные частоты на Кавказе ( $93 \leq q \leq 99\%$ ) 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 ( $0 \leq q \leq 13\%$ ,  $G_{ST} = 0.032$ ), 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 ( $12 \leq q \leq 100\%$ ;  $G_{ST} = 0.433$ ). 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* ( $0 \leq q \leq 8\%$ ;  $G_{ST} = 0.02$ ) 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 ( $74 \leq q \leq 86\%$ ) и на юге (монголы, алтайцы;  $74 \leq q \leq 82\%$ ) Северной Евразии. В то время как в Европе наибольшие значения обнаруживаются лишь у македонцев и коми-пермяков ( $72 \leq q \leq 73\%$ ). 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 ( $46 \leq q \leq 66\%$ ), максимальные у всех восточных славян, в Мордовии и у западных финно-угров ( $90 \leq q \leq 93\%$ ), причем область высоких частот полностью захватывает и весь Кавказ ( $81 \leq q \leq 88\%$ ).

*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 ( $5 \leq q \leq 9\%$ ) максимальные частоты ( $95 \leq q \leq 100\%$ ) 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 ( $0 \leq q \leq 13\%$ ) уже практически исчезают.

Многие из 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 ( $43 \leq q \leq 48\%$ ). 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 ( $61 \leq q \leq 70\%$ ). 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 ( $35 \leq q \leq 48\%$ ) разбросаны географически хаотично, приходясь не только на западных фин-

но-угров, русские и мордовские популяции, но и на Южную Сибирь, и на Дальний Восток.

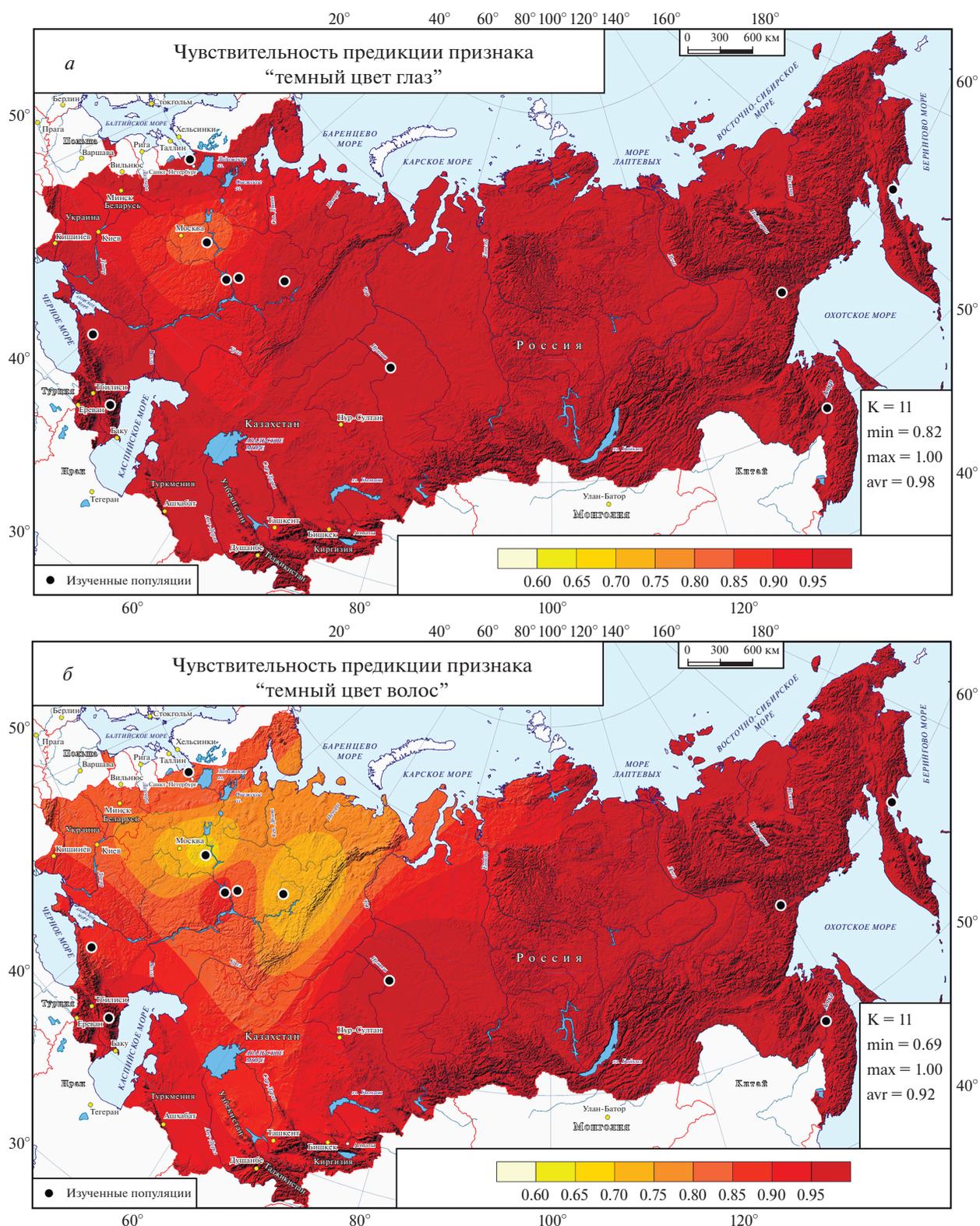
На карте *SACNA2D3\_rs11283625\_A* (рис. П.4) область максимальных частот своеобразно охватывает западной дугой популяции от Русского Севера до Восточного Кавказа, оставляя Поволжье и Приуралье в области более низких частот. А на карте *rs34191540\_T* (рис. П.22), демонстрирующей значительное генетическое разнообразие ( $19 \leq q \leq 100\%$ ;  $G_{ST} = 0.18$ ), 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 ( $p = 76\%$ ). Немногом выше показатель у тюрков 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 тюрков 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 К — 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 приходится вновь на коми-пермяков ( $p = 60\%$ ) и на популяции русских и Мордовии ( $p = 67\%$ ). Немного снижен ( $71 < p < 85\%$ ) показатель вновь у тюрков 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. Ожидается, 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 пигмента-

ции 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  $r = 0.2$  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

- Morgan M., Païro-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>
- 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>
- 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>
- 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>
- Lazaridis I., Mitnik 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- Балановская Е.В., Рычков Ю.Г. Этническая генетика: Этногеографическое разнообразие народов мира // *Генетика*. 1990. Т. 26. № 1. С. 114–121.
- 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>
- 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>
- 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.
- Берлянт А.М. Картография: Учебник для вузов. М.: Аспект Пресс, 2002. 336 с.
- 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>
- 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.
- 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.
- 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>
- 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<sup>a, b, \*</sup>, I. O. Gorin<sup>a, c</sup>, S. M. Koshel<sup>a, d</sup>, and O. P. Balanovsky<sup>a, b, c</sup>**<sup>a</sup>*Scientific Institution Research Centre of Medical Genetics, Moscow, 115478 Russia*<sup>b</sup>*Biobank of North Eurasia, Moscow, 115215 Russia*<sup>c</sup>*Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, 119991 Russia*<sup>d</sup>*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 inter-population 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.