

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ «ТОМСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
МЕДИЦИНСКИЙ ЦЕНТР РОССИЙСКОЙ АКАДЕМИИ НАУК»
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ МЕДИЦИНСКОЙ ГЕНЕТИКИ

На правах рукописи



Колесников Никита Александрович

**ИДЕНТИЧНЫЕ ПО ПРОИСХОЖДЕНИЮ БЛОКИ И РЕГИОНЫ
ВЫСОКОЙ ГОМОЗИГОТНОСТИ В ГЕНОМАХ КОРЕННОГО
НАСЕЛЕНИЯ СИБИРИ: ПРОИСХОЖДЕНИЕ, РАСПРОСТРАНЕНИЕ,
АДАПТИВНАЯ ЗНАЧИМОСТЬ**

1.5.7 генетика

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата биологических наук

Научный руководитель:
доктор биологических наук
Харьков Владимир Николаевич

Томск – 2022

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
ГЛАВА 1 ОБЗОР ЛИТЕРАТУРЫ	15
1.1 Изучение генетического разнообразия популяций человека.....	15
1.1.1 Системы генетических маркеров	16
1.2 Биоинформатические методы анализа генетической структуры.....	19
1.2.1 Анализ распределения частот аллелей с помощью метода главных компонент и алгоритмов t-SNE и Admixture.....	19
1.2.2 Регионы высокой гомозиготности	22
1.2.3 Коэффициент инбридинга.....	27
1.2.4 Блоки идентичные по происхождению	29
1.3 Поиск сигналов отбора в популяциях человека	30
1.4 Этногенез коренных народов Сибири	33
1.4.1 Антропологическая классификация.....	36
1.4.2 Языковые группы коренного населения Сибири	37
1.5 Заключение по обзору литературы	39
ГЛАВА 2 МАТЕРИАЛЫ И МЕТОДЫ ДИССЕРТАЦИОННОГО ИССЛЕДОВАНИЯ.....	41
2.1 Характеристика популяционных групп.....	41
2.2 Подготовка материала для генотипирования.....	47
2.3 Генотипирование.....	48
2.4 Биоинформатическая обработка данных.....	49
2.4.1 Оценка качества генотипирования.....	49
2.4.2 Фазирование	50

2.5 Анализ компонентного состава генофонда методами PCA, t-SNE и Admixture	50
2.6 Выявления общих по происхождению блоков сцепления (IBD-сегментов) .	50
2.7 Выявления регионов высокой гомозиготности (RON)	51
2.7.1 Идентификация регионов высокой гомозиготности с помощью программы PLINK.....	51
2.7.2 Расчет коэффициента геномного инбридинга	51
2.7.3 Идентификация регионов высокой гомозиготности с помощью программы GARLIC	52
2.7.4 Выявление горячих точек RON	52
2.8 Выявление сигналов естественного отбора в высокоплотных чиповых данных	53
ГЛАВА 3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	54
3.1 Анализ главных компонент.....	55
3.2 Анализ предковых компонент с помощью алгоритма Admixture.....	59
3.3 Распределение регионов высокой гомозиготности в исследуемых популяциях.....	67
3.3.1 Оценка распределения регионов высокой гомозиготности с помощью программы «Plink»	67
3.3.2 Оценка распределения регионов высокой гомозиготности с помощью программы «GARLIC»	69
3.3.3 Оценка уровня инбридинга F_{RON} в исследуемых популяциях.....	79
3.4 «Горячие точки» RON	83
3.5 Блоки идентичные по происхождению.....	87
3.6 Поиск сигналов направленного отбора с помощью тестов на протяженную гомозиготность гаплотипов	94

3.6.1 Сигналы направленного отбора, выявленные с помощью теста ХР-ЕНН	110
ЗАКЛЮЧЕНИЕ	113
ВЫВОДЫ	118
СПИСОК СОКРАЩЕНИЙ.....	119
СПИСОК ЛИТЕРАТУРЫ.....	120
ПРИЛОЖЕНИЕ А	136

ВВЕДЕНИЕ

Актуальность темы исследования

Одной из важнейших задач современных исследований в области популяционной генетики человека является характеристика variability генома как на видовом, так и на популяционном и индивидуальных уровнях. Генетическое разнообразие в геномах отдельных индивидов и генофонде человечества в целом представляет собой базу для решения задач эволюционной генетики, популяционной геномики, генетики мультифакториальных заболеваний и фармакогенетики. Анализ различий структуры ДНК является основой как для фундаментальных научных работ, так и для практической идентификации биологических образцов в судебно-медицинской экспертизе и ДНК-идентификации [1].

В настоящее время анализ структуры генофондов различных популяций является актуальным направлением научных исследований. За последние несколько лет развитие новых технологий масштабного генотипирования и технологий биоинформационного анализа позволили, с одной стороны, перейти на новый уровень изучения генетической структуры популяций человека, а с другой стороны, провести высокоразрешающие полногеномные исследования ассоциаций и сравнительный анализ генетического разнообразия индивидуальных геномов. Активно накапливающиеся данные по индивидуальным геномам из различных научных проектов появляются в свободном доступе и позволяют интегрировать эти результаты при биоинформационном анализе различных по масштабу массивов данных. При этом популяции России остаются пока относительно слабо вовлеченными в такого рода исследования. Популяции коренных этносов Сибири представляют значительный интерес как в силу относительно слабой их изученности с привлечением современных геномных технологий, так и по причине специфичности их генофондов, развивавшихся в условиях длительной генетической изоляции.

Генофонд коренного населения Сибири представляет собой уникальную систему с точки зрения исследования популяционно- и эволюционно-генетических процессов, анализа генетического разнообразия и реконструкции генетической истории популяций. Высокое этническое разнообразие является особенностью Сибири, как одного из регионов периферийного расселения современного человека. Огромные пространства этого региона и малочисленность аборигенного населения способствовали формированию значительной территориальной и генетической подразделенности. На территории сибирской историко-этнографической провинции расселены около сорока коренных народностей. Длительная изоляция населения Сибири от основных миграционных потоков на территории Евразии обеспечила сохранение в сибирском генофонде древних генетических пластов, хранящих информацию о важнейших этапах заселения человеком современного антропологического типа не только Сибири, но и Центральной Азии, Европы и Нового Света.

Как известно, генетико-демографические процессы в популяциях, колебания численности, метисационные события, миграции и естественный отбор, оказывают влияние на структурированность генетического разнообразия в геномах отдельных индивидов и популяциях в целом. В частности, генетико-демографические процессы приводят к тому, что формируются общие по происхождению блоки сцепления (IBD - identity by descent). Сегмент, имеющий идентичные нуклеотидные последовательности, является IBD у двух или более лиц, если они унаследовали его от общего предка без рекомбинации, то есть, у этих людей сегмент имеет общее происхождение. Ожидаемая длина блока IBD зависит от количества поколений с момента появления последнего общего предка в локусе сегмента [2].

Идентифицированные сегменты IBD могут использоваться для широкого круга задач. Одним из применений обнаруженных блоков IBD является количественная оценка степени родства, что может быть использовано в судебной генетике, и также может дополнить информацию в картографировании генетических связей между различными популяциями [3]. Сегменты IBD могут

помочь и в обнаружении естественного отбора в геноме человека. Отбор обычно увеличивает число сегментов IBD среди людей в популяции. При сканировании регионов с избытком обмена IBD могут быть идентифицированы регионы в геноме человека, которые находятся под очень недавним и сильным отбором [4,5]. Сегменты IBD могут использоваться для оценки демографической истории, включая эффект бутылочного горлышка и потоки генов в популяциях [6]. В недавних исследованиях были показаны различия в распределении IBD между африканскими, азиатскими и европейскими популяциями, а также сегменты IBD, общие с древними геномами, такими как геномы неандертальцев или денисовцев [7].

Помимо формирования идентичных по происхождению блоков, генетико-демографические процессы приводят к появлению регионов высокой гомозиготности (RON – runs of homozygosity). Структурные особенности распределения RON между отдельными людьми и популяциями также обеспечивают ценный и часто неиспользуемый ресурс для изучения генетического разнообразия человечества и его эволюционной истории.

Число и общая длина RON у каждого человека показывают значительные различия между отдельными людьми и группами населения. Сумма длин коротких и средних RON у индивида возрастает с увеличением географического расстояния популяций от Восточной Африки. Напротив, общий суммарный размер длинных RON имеет высокую индивидуальную изменчивость, которые, вероятно, отражают недавние события инбридинга, причем более высокие значения чаще встречаются в популяциях с известными высокими частотами кровнородственных браков. Распределение RON по геному неравномерно и имеет специфические особенности для разных географических регионов. Частота RON в геноме коррелирует со скоростью рекомбинации, а также с сигналами недавнего положительного отбора. Кроме того, длинные RON встречаются чаще в геномных регионах, содержащих гены, связанные с аутосомно-доминантными заболеваниями, чем в регионах, не участвующих в менделевских заболеваниях [8].

Выявленные закономерности распределения RОН в геноме человека на полногеномном уровне позволяют лучше понять генетико-демографическую историю популяций, оценить уровень инбридинга, выявить регионы подверженные отбору. Данные по распределению RОН в популяциях могут служить основой для улучшения специфичности идентификации регионов, связанных с заболеваниями, в контексте картирования гомозиготности [9].

На основании различных исследований показано, что генетическая структура коренного населения Сибири и Крайнего Севера сложилась под существенным влиянием естественного отбора в наиболее неблагоприятных условиях среды обитания: в арктическом климате Крайнего Севера и резко континентальном климате Сибири и Дальнего Востока. Длительная эволюция генофонда различных популяций Сибири, в том числе приспособительная, могла привести, с одной стороны, к формированию устойчивых адаптивных комплексов, а с другой – к специфическому спектру заболеваний как моногенной, так и мультифакторной природы. Ряд недавних исследований показывает, что генетические системы, связанные с жировым метаболизмом, устойчивостью к температурным шокам, тонусом сосудистого русла, изменением инфекционной нагрузки [10–13] могут играть существенную роль в адаптации к неблагоприятным климато-географическим условиям. Переход от охоты к животноводству, смена типа питания на фоне резкого роста численности популяций могли привести к изменению структуры генофонда популяций по различным генам, продукты которых участвуют в липидном обмене и усвоении различных типов пищевых продуктов.

В работах по популяционной генетике сибирских этносов с использованием нейтральных маркерных систем Y-хромосомной, митохондриальной ДНК, показано, что во многих популяциях наблюдается эффект основателя по отдельным гаплогруппам с низким генетическим разнообразием, которое свидетельствует о событиях экспансии численности в относительно недавнее время [14–18].

При анализе древней ДНК в популяциях Европы было показано по каким генам и в каком направлении за последние 8 тысяч лет действовал естественный отбор. Культурные и социальные изменения в неолите сопровождались изменениями частоты аллелей, связанных с различными признаками. Под отбором находились аллели толерантности к лактозе, пигментации кожи и глаз, метаболизма, а также роста и веса [19]. Население Сибири все еще остается очень слабо изученным в отношении генетического разнообразия на полногеномном уровне. Мы предполагаем, что генетическая структура коренного населения Сибири сложилась под существенным влиянием генетического дрейфа на фоне резкого изменения численности популяций, с одной стороны, и естественного отбора, сформировавшего устойчивые адаптивные генетические комплексы, с другой стороны. В совокупности с изоляцией расстоянием это привело к значительной генетической подразделенности населения Сибири.

Степень научной разработанности темы исследования

Несмотря на относительно неплохую изученность генофондов населения Сибири по классическим маркерам (различным системам аутосомных, митохондриальных [20–22] и Y-хромосомных маркеров [15,18,23,24]), генофонды этих народов относительно слабо охарактеризованы на уровне геномов. Полногеномные данные и данные по высокоплотным чипам имеются лишь для небольшого числа популяций, как правило, в очень ограниченных выборках [25–27]. Помимо этого, имеется относительно небольшое количество исследований, где популяции Сибири изучались на основании анализа идентичных по происхождению блоков и регионов высокой гомозиготности по данным генотипирования SNP-чипов высокой плотности с использованием современных биоинформационных подходов [27].

Учитывая все вышеперечисленное, были сформулированы цель и задачи настоящей работы:

Цель исследования

Охарактеризовать генетическую структуру популяций коренного населения Сибири на основании анализа идентичных по происхождению блоков и регионов высокой гомозиготности.

Задачи исследования

1. Выявить идентичные по происхождению блоки (IBD) и регионы высокой гомозиготности (ROH) в геномах представителей коренного населения Сибири на основе генотипов, полученных с помощью SNP-чипов.
2. Провести анализ внутри- и межпопуляционного распределения блоков идентичных по происхождению (IBD) и регионов высокой гомозиготности (ROH).
3. Выявить роль потенциально значимых генетико-демографических процессов в формировании генофонда сибирских популяций с помощью биоинформатических подходов.
4. Провести поиск сигналов естественного отбора в популяциях Сибири с помощью биоинформатического анализа регионов гомозиготности.

Научная новизна

В рамках данной работы впервые проведено масштабное популяционное исследование генофонда коренных народов Сибири на уровне высокоплотных чипов. Это позволило заполнить существенный пробел в геногеографической картине населения Евразии.

Впервые проведен подробный анализ состава генофонда и генетического разнообразия некоторых коренных народов Сибири с использованием методики выявления общих по происхождению блоков сцепления (IBD-сегментов) и регионов высокой гомозиготности (ROH), что позволило подробно описать генофонды популяций, выявить демографические экспансии на территории Сибири и определить причины, которые привели к тем или иным структурным изменениям в геноме.

В рамках выполненного исследования проведена характеристика структуры генофонда коренных сибирских популяций на уровне высокоплотных чипов. Впервые получены данные о генетической вариабельности ряда российских популяций. Выявлены гены и генетические маркеры, участвовавшие в адаптивной эволюции генетической структуры популяций Сибири. Выполнен анализ влияния естественного отбора в аспекте адаптации генофонда популяций к факторам окружающей среды.

Результаты работы расширяют новое направление в эволюционной и популяционной генетике человека и эволюционной медицине, связанное с поиском сигналов адаптивной эволюции генофонда популяций человека.

Теоретическая и практическая значимость работы

Основной научной проблемой исследования является характеристика генетического разнообразия коренного населения Сибири на уровне генома с использованием высокоплотных чипов. Исследование находится в рамках фундаментальных направлений популяционной и эволюционной генетики человека – выявления этноспецифичных особенностей структуры геномов и популяционных генофондов современных популяций человека, реконструкции древних миграций и экспансии численности населения, а также эволюции генофондов и механизмов генетической адаптации.

Полученные в ходе настоящего исследования данные на основании анализа идентичных по происхождению блоков и регионов высокой гомозиготности по данным генотипирования SNP-чипов высокой плотности расширили представления о структуре генофонда населения Сибири. Данные о сигналах естественного отбора, являются важным дополнением к существующим данным о эволюции генофондов и механизмам генетической адаптации населения Евразии. Практическая значимость работы может реализоваться в использования полученных данных для разработок в области арктической медицины с целью выявления физиологических, генетических и системных механизмов адаптации к экстремальным условиям среды обитания и разработке на этой основе

эффективных инновационных средств и методов оценки адаптивного потенциала, а также ДНК-идентификации, лечения и профилактики болезней человека.

Методология и методы исследования

Теоретическую и методологическую основу работы обеспечили исследования в области изучения популяционной и эволюционной генетики человека, а также разработки в области идентификации регионов высокой гомозиготности, блоков IBD, а также сигналов естественного отбора.

Молекулярно-генетические методы, использованные в работе, включали выделение ДНК фенол-хлороформной экстракцией и генотипирование на основе чипов высокой плотности.

Статистические методы и подходы, примененные в данной работе, включали анализ данных с помощью программного обеспечения, реализующего современные подходы для идентификации исследуемых регионов в геноме и поиска сигналов естественного отбора, а также разработанные ранее пакеты прикладных программ для статистической среды R.

В отношении объема исследуемого материала, технологического уровня генотипирования, степени детализации анализа генотипических данных и методов биостатистики исследование полностью соответствует современному мировому уровню.

Положения, выносимые на защиту

1. Популяции Сибири демонстрируют значительные различия в распределении суммы длин и количества ROH как на индивидуальном уровне, так и между исследуемыми географическими регионами и этносами. Для популяций Сибири характерно большее значение суммарной длины коротких и средних ROH на человека по сравнению с популяциями Кавказа (в том числе Дагестана), Средней Азии, Европы и Волго-Уральского региона.

2. В популяциях Сибири выявлены существенные отличия в значениях коэффициента геномного инбридинга по сравнению с популяциями Кавказа, Средней Азии, Европы и Волго-Уральского региона: большая доля популяций Сибири имеют высокие значения коэффициента геномного инбридинга для $RON > 1,5$ м.п.о.

3. Распределение блоков IBD в популяциях Сибири в большей степени согласуется с географическим расположением популяций, а в отдельных случаях – с их языковой принадлежностью.

4. В генофонде популяций коренного населения Сибири содержится существенное число сигналов направленного отбора, вероятно, обусловленных процессами долговременной адаптации к условиям проживания в резко континентальном или арктическом климате.

Степень достоверности результатов проведенных исследований.

Достоверность полученных результатов работы обосновывается большим объемом исследуемых популяционных выборок (1836 индивидов, в среднем 24 образца на популяцию), использованием современных молекулярно-генетических методов и статистических подходов для анализа данных.

Апробация диссертационной работы

Основные положения диссертационной работы были представлены на следующих конференциях:

- 1) Международная научная конференция молодых ученых «Актуальные проблемы медицинской генетики» (Томск, 2018);
- 2) Международная конференция «VII Съезд Вавиловского общества генетиков и селекционеров, посвященный 100-летию кафедры генетики СПбГУ, и ассоциированные симпозиумы» (Санкт-Петербург, 2019);
- 3) Всероссийский Конгресс молодых ученых «Актуальные вопросы фундаментальной и клинической медицины» (Томск, 2020);.

Декларация личного участия автора

Основные результаты диссертационной работы получены автором исследования. Анализ литературных данных по теме диссертации, проведение экспериментальной работы (выделение и проверка качества ДНК, подготовка образцов для генотипирования), биоинформатическая обработка данных, статистический анализ, описание и обобщение полученных результатов, а также подготовка публикаций по результатам исследования выполнены лично автором.

Публикации

По теме диссертационной работы опубликовано 8 работ (5 статей в журналах, рекомендованных ВАК, 1 статья в сборнике конференции, 2 тезиса в материалах зарубежных и отечественных конференций).

Объем и структура диссертации

Диссертация изложена на 176 страницах машинописного текста и содержит введение, обзор литературы, материалы и методы, результаты и обсуждение, заключение, выводы, список литературы, приложение. Работа иллюстрирована 18 таблицами и 22 рисунками. Список литературы состоит из 147 источников, из них 31 отечественный.

ГЛАВА 1 ОБЗОР ЛИТЕРАТУРЫ

1.1 Изучение генетического разнообразия популяций человека

Изменчивость есть фундаментальное свойство всех живых организмов, её же непосредственным выражением является генетическое разнообразие. Итогом всех генетических процессов, совершающихся на всех уровнях организации – от молекулярного до видового – является генетический полиморфизм [28]. Изучением механизмов формирования этого генетического разнообразия, а также исторических процессов, приведших к современной популяционно-генетической структуре, занимается популяционная генетика.

Совокупность геномов отдельных организмов и суммы аллелей генов их составляющих в пределах популяции формирует ее генофонд. В применении к человеческой популяции, “генофонд – это географически распределенное и исторически упорядоченное множество генов (и любых других реплицирующихся участков молекулы ДНК), воспроизводимое ею в поколениях и поддерживаемое систематическими и стохастическими силами эволюции в динамическом равновесии с состоянием вмещающей популяцию и изменяемой ею среды” [28].

Достижения последних лет в области развития технологий генотипирования ДНК способствовали бурному развитию этногеномики [29–31]. Согласно определению, данному Э. К. Хуснутдиновой, этногеномика – это раздел популяционной генетики, изучающий особенности геномного полиморфизма и геномного разнообразия отдельных популяций, этносов и реконструкция на этой основе их генетической истории [32]. В современном представлении вместо этногеномики также используется термин «популяционная геномика» [33].

Изучение изменчивости генетических маркеров на уровне отдельных популяций, этносов, этнических групп с целью выявления структуры генофондов локальных популяций и филогенетических взаимоотношений между ними является одной из основных задач этногенетики. До недавнего времени большинство этногеномных исследований были направлены на воссоздание

картины расселения человека по Земле в глобальном масштабе. В настоящее время все большую актуальность приобретают работы, связанные с детализацией эволюционной и демографической истории отдельных регионов или расово-этнических групп [34,35].

Популяция является основной эволюционной единицей. Именно особенности структуры локальных популяций определяют состав генофонда этносов и региональных групп населения. Структура генофонда во многом зависит от особенностей эволюционной истории тех или иных популяций и имеет популяционно-специфический характер.

К современным достижениям популяционной генетики и геномики человека относятся полное секвенирование большого числа полных геномов из различных популяционных выборок, которое привело к выявлению огромного множества полиморфных ДНК-маркеров, систематической каталогизации генов и генетических полиморфизмов. Выявление полногеномной структуры гаплотипов и новые статистические подходы позволяют проводить исследование генофондов на совершенно новом уровне.

1.1.1 Системы генетических маркеров

В исследованиях эволюции популяций человека используются различные системы генетических маркеров, отличающиеся локализацией в геноме, уровнем вариабельности и характером мутирования. Современная популяционная генетика человека имеет широкий выбор различных маркерных систем. Это «классические» аллоантигены клеточных поверхностей, ферментные системы и белки сыворотки крови [36,37], молекулярно-генетические маркеры, такие как аутосомные и однополовые ДНК-маркеры (гаплогруппы митохондриальной ДНК и Y-хромосомы). ДНК маркеры делят на диаллельные (SNP) и мультиаллельные (мини- и микросателлиты), каждый из которых имеет свои специфические особенности, достоинства и недостатки [38,39].

Анализ генофонда популяции с использованием какой-либо одной маркерной системы является важным подходом в большинстве научных

исследований [28,31,40]. Однако для получения полной информации о структуре генофондов популяций необходимо использовать всю совокупность методов, включая классические маркерные системы и ДНК-маркеры аутосом, Y-хромосомы и мтДНК.

Маркеры на основе биохимического полиморфизма были первыми молекулярными маркерами. Использование сотен биохимических маркеров позволило разработать основные теоретические положения популяционной генетики и оценить уровень генетического полиморфизма у огромного количества биологических видов [28,37,41]. Ограничения в применении биохимических маркеров связаны с возможностью работы только с белок-кодирующими последовательностями у экспрессирующихся генов. При этом из анализа исключаются различные участки, расположенные вне генов, а также сайты регуляции, промоторные области, энхансеры, расположенные в интронах и нетранслируемых областях генов [38].

В начале 80-х годов появилось возможность использовать ДНК-маркеры, что дало сильный толчок в эволюционных и популяционных исследованиях. Более детальное изучение генетического полиморфизма стало доступно благодаря развитию методов выделения, клонирования и рестрикции генов. До 1988 г. разнообразие ДНК-маркеров было невелико (ПДРФ и минисателлиты). Решающая роль в осуществлении «ДНК-революции» и созданию новых типов ДНК-маркеров принадлежит изобретенному К. Мюллисом в 1983 г. методу амплификации *in vitro* определенных участков ДНК в процессе повторяющихся температурных циклов полимеразной реакции (ПЦР) – в настоящий момент очень широко используемого в генетических исследованиях.

Одним из методов изучения структуры генофондов современных популяций человека является исследование гаплоидных участков генома человека – митохондриальной ДНК и Y-хромосомы. Маркеры такого типа имеют значительные преимущества, такие как отсутствие рекомбинации и линейная передача в ряду поколений. ДНК-маркеры специфических участков генома человека позволяют исследовать отдельно материнскую и отцовскую линии, так

как Y-хромосома передается потомкам только от отца, митохондриальная ДНК наследуется только по материнской линии. Это дает новые возможности в этногенетических исследованиях: проследить и сопоставить историю женской и мужской части популяции и оценить их вклад в популяционный генофонд [42].

Разработка и применение технологии ДНК-микрочипов для обнаружения генетических вариаций начали активно развиваться в конце 1990-х годов. Этот метод позволяет за короткое время тестировать сотни и тысячи образцов с возможностью производить анализ мутаций во всех генах генома одновременно. Для анализа всех возможных мутаций во всех генах человека достаточно ДНК-чипа с количеством ячеек 100–200 млн, что технически достижимо, однако современные ДНК-микрочипы могут выявить около нескольких миллионов мутаций. ДНК-чипы разделяют на чипы «высокой плотности» (high density) и «низкой плотности» (low density).

В классическом исполнении метод ДНК-микрочипов основан на проведении моноплексной или мультиплексной ПЦР с последующей гибридизацией флуоресцентно-меченного фрагмента ДНК на микрочипе, содержащем оригинальный набор дифференцирующих олигонуклеотидов, а также процедуры регистрации и интерпретации результатов [43]. Успешное внедрение технологий микрочипов потребовало разработки многих методов и приемов изготовления микрочипов и определения зондов, проведения и обнаружения реакций гибридизации, а также информатики для анализа данных [44].

В настоящее время в популяционных научных исследованиях используются главным образом SNP-маркеры [45,46]. SNP – это однонуклеотидные позиции в геномной ДНК, для которых в популяции имеются различные варианты последовательностей, при этом аллель считается редким, если он встречается с частотой не более 1% [47]. Чаще всего SNP представлены двумя аллельными вариантами, но возможен и трехаллельный полиморфизм (примерно 0,1% от всех SNP в геноме человека), хотя они крайне редки [48]. Большинство SNP, как принято считать, появляются в результате уникальных мутаций. В геноме

человека насчитывается огромное количество SNP, что позволяет отобрать около 100000 SNP-маркеров с частотой редкого аллеля более 20% при среднем расстоянии между ними в 30 т.п.н. [48]. SNP имеют низкий уровень мутации на поколение (примерно 10^{-8}), что делает их удобными маркерами молекулярной эволюции при оценке крупномасштабных эволюционных событий [49].

Появившиеся биоинформатические методы анализа, а также развитие вычислительной техники дают возможность интерпретировать большие объёмы экспериментальных данных (что до развития этой области было практически невозможно) и позволяют более детально анализировать картину структуры генофондов взаимодействия популяций, генетико-демографических процессов, естественного отбора и реконструкции миграций.

1.2 Биоинформатические методы анализа генетической структуры

Современные методы исследования генетической структуры основаны на анализе сотен тысяч SNP-маркеров по всему геному. Одни из наиболее распространенных методов анализа аутосомных маркеров включают в себя алгоритмы Admixture, поиск идентичных блоков по происхождению и регионов высокой гомозиготности.

1.2.1 Анализ распределения частот аллелей с помощью метода главных компонент и алгоритмов t-SNE и Admixture

Анализ генетической структуры популяций человека необходим для исследования генетического разнообразия и дифференциации населения и позволяет реконструировать генетическую историю, миграции и расселение современного человека. Существует множество способов визуализации и анализа генетических маркеров, одними из наиболее распространенных статистических методов изучения структуры популяции являются метод главных компонент (PCA – Principal components analysis) [50,51] и метод, использующий алгоритм кластеризации Admixture [52].

Анализ с помощью метода главных компонент позволяет выявить интегральные характеристики распределения частот аллелей в исследованных популяциях. Главными целями такого подхода к анализу генотипических данных являются сокращение числа переменных и определение структуры взаимосвязей между переменными. Суть метода главных компонент состоит в замене первоначальных коррелированных признаков на некоррелированные компоненты, или факторы, которые могут быть упорядочены в соответствии с долей дисперсии, которую они определяют в совокупной изменчивости первоначальных признаков. Таким образом, можно выделить наиболее значимые факторы, определяющие основные закономерности изменчивости генофонда. Метод главных компонент позволяет выделить группы генетически связанных друг с другом этносов. Несмотря на то, что он впервые был представлен в 1978 г., этот подход долгое время остается востребованным [26,53]. Актуальность метода сохраняется благодаря возможности интерпретации данных, содержащих информацию о огромном массиве генотипических данных для тысяч образцов, в результате чего относительно небольшие, но реальные эффекты могут хорошо определяться при построении графиков между отдельными компонентами.

Для дополнительного анализа популяций также можно использовать метод стохастического вложения соседей с t -распределением (англ. t -distributed Stochastic Neighbor Embedding, t -SNE) [54] – это алгоритм машинного обучения для визуализации [55], набирающий популярность в анализе генетических данных человека [56]. Он использует нелинейное снижение размерности, хорошо подходящее для данных высокой размерности, для их визуализации в пространстве низкой размерности (двух- или трехмерное). Метод моделирует каждый объект высокой размерности двух- или трёхмерной точкой таким образом, что похожие объекты моделируются близко расположенными точками, а непохожие точки моделируются с большой вероятностью точками, отстоящими далеко друг от друга. В отличие от PCA, t -SNE более устойчив к наличию выбросов.

Биоинформатический подход к анализу распределения частот в популяциях, который реализован в компьютерной программе Admixture – ещё один инструмент, позволяющий строить предположения о происхождении популяции. Алгоритм позволяет выявлять смешанность состава популяционной выборки индивидов на основе данных о генотипах. Алгоритм кластеризации Admixture используется для идентификации потенциальных предковых компонентов среди геномов исследуемых популяций. Полногеномные данные и высокоплотные панели SNP позволяют на уровне индивидуальных геномов и популяций оценить вклад компонентов разного генетического происхождения. В последнее время моделирование с помощью Admixture является одним из основных методов анализа при исследовании генофондов современных и древних популяций человека. Этот метод позволяет анализировать одни и те же данные на разных иерархических уровнях – от индивидуального генома до генофонда человечества.

Основная суть анализа с помощью метода Admixture – проверка серии гипотез о разном числе предполагаемых предковых компонент на одном и том же массиве данных. В программе вносится допущение, что изучаемые популяции произошли от каких-то гипотетических предковых популяций, которые обозначаются как «предполагаемые предковые популяции» (ППП). Простейшая гипотеза, когда минимальное число предковых компонентов составляет 2 ($K=2$). При таком условии алгоритм исходит из предположения, что для всех индивидов в анализе существовало только две значительно различающиеся между собой предковые популяции, и рассчитывает вклад каждой из них в генотип каждого образца. Проверка следующей гипотезы – о трех предковых популяциях ($K=3$) для того же набора генотипов – происходит независимо и аналогично с последующими гипотезами.

Имея сведения о генотипах и предположение о количестве предполагаемых предковых популяций, Admixture строит свою модель того, каков вклад каждой из PPP для каждого отдельного индивида. В результате для каждого индивида возникают несколько значений (в зависимости от количества сравниваемых популяций): количественный вклад каждой из этих популяций. При этом может

быть и такая ситуация, что у конкретного индивида в составе отсутствует какая-то из ППП, даже возможно, что он принадлежит только к одной из ППП. Чем больше вклад каждой ППП в отдельного индивида, тем больше этот индивид является «носителем» данной популяции и ее представителем. Эти значения можно графически визуализировать. Далее индивиды группируются по «известным» популяциям, и в результате мы получаем генетические профили каждой из популяций. Затем возможные известные популяции сравниваются по этим генетическим профилям между собой. При сравнении важен как количественный состав популяций, так и качественный, т.е. какие «предполагаемые предковые популяции» присутствуют в генотипе каждого исследуемого образца и в каком соотношении. В итоге, в результате анализа Admixture таблица исходных генотипов используется для генерации таблицы с долями множества предковых компонентов (которые можно варьировать от 2 до 10-15 и более) для каждого образца. Практика применения этого подхода для анализа как чиповых, так и полногеномных данных показывает его высокую эффективность при анализе значительных массивов генотипических данных их различных популяций.

Алгоритм Admixture позволяет решить несколько задач. Во-первых, провести контроль качества, оценивая схожесть рисунка компонент у представителей одной популяции. Во-вторых, оценить близость популяций по порядку и особенности комбинаций компонент при постепенном их увеличении. В-третьих, выявить метисацию отдельных индивидов для дальнейшей фильтрации данных.

1.2.2 Регионы высокой гомозиготности

Существенная часть любого индивидуального генома человека представлена регионами высокой гомозиготности. Общая длина и распределение таких участков связаны с генетико-демографической историей популяции и отражают эффекты основателя, уровень инбридинга, а также могут быть следствием действия сильного направленного отбора [57,58]. Современные

методы анализа данных полных геномов или высокоплотных наборов SNP позволяют детально выявлять структуру и распределение ROH в индивидуальных геномах и популяциях с целью уточнения генетической истории популяций, оценки уровня инбридинга, поиска сигналов естественного отбора [57,59]. Данные по распределению ROH в популяциях могут служить основой для идентификации участков генома, связанных с заболеваниями, в контексте картирования гомозиготности [60]. Общая длина и структура ROH характеризуется существенной межиндивидуальной и межпопуляционной вариабельностью.

Предыдущие исследования показали, что у неродственных людей широко распространены короткие ROH, измеряемые несколькими десятками тысяч пар оснований (т.п.о.) [61]. Регионы высокой гомозиготности средних размеров – от сотен тысяч до нескольких миллионов п.о. – являются, как правило, результатом общего происхождения индивидов внутри популяции и отображают недавние демографические изменения в популяции, например, такие как эффект бутылочного горлышка. Длинные ROH, измеряющиеся десятками миллионов п.о., характерны для популяций с высоким уровнем близкородственных браков [62].

Как правило, сумма длин коротких и средних ROH у индивида возрастает от восточно-африканских к европейским и азиатским популяциям в силу накопления рекомбинационных событий. Напротив, общий суммарный размер длинных ROH показывает большие межиндивидуальные вариации, отражая, вероятно, эффекты недавнего инбридинга [8,62]. Кроме того, показано, что частота распределения ROH в геноме коррелирует со скоростью рекомбинации, а также с сигналами недавнего положительного отбора [63]. Шесть различных гипотетических демографических сценариев в популяциях представлены на рисунке 1. Как можно видеть, наличие ROH связано с размером популяции, при этом малочисленные популяции имеют большее количество и длину ROH. Биологическое смешение между индивидами различных этнических групп, приводящее к формированию новых популяций, объединяет разные гаплотипы и обычно снижает количество ROH до очень небольшого количества коротких ROH, тогда как эффект

бутылочного горлышка увеличивает количество ROH, которые обычно все еще относительно короткие. Инбридинг, с другой стороны, добавляет небольшое количество очень длинных ROH для тех, кто является потомком кровнородственных браков, таким образом также увеличивая разброс в сумме ROH, видимую как сдвиг вправо на графике NROH по сравнению с SROH. Для популяций с эффектом бутылочного горлышка и распространенными кровнородственными браками будет наблюдаться максимальное количество длинных ROH с тенденцией увеличения длины ROH.

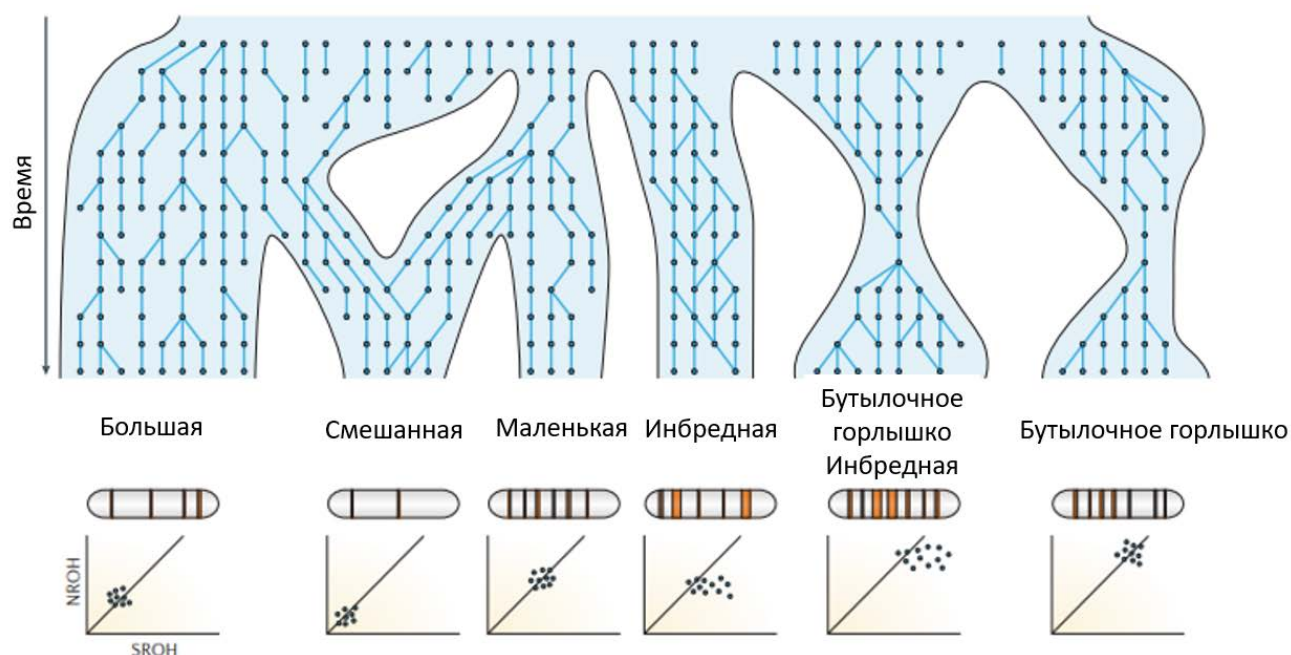


Рисунок 1 - Демографическая история шести различных гипотетических популяций [64]. Примечание: Последствия каждого демографического сценария проиллюстрированы ниже: схематическое распределение ROH по хромосомам, показывающие типичное распределение регионов высокой гомозиготности (ROH) для каждого сценария, ниже график зависимости общей длины ROH (SROH) от общего количества ROH (NROH) ожидаемый в каждом сценарии.

ROH могут быть обнаружены в данных генотипирования микрочипов или полногеномного секвенирования с использованием либо методов «наблюдения», либо подходов на основе моделей [8], реализованных в программах PLINK[65] и GARLIC [58], соответственно.

Анализ ROH в PLINK выполняется с использованием функции `homozyg`. Алгоритм PLINK для обнаружения ROH основан на подходе «сканирующего окна», который состоит из трех шагов. Во-первых, размер сканирующего окна определяется заранее определенным количеством SNP (`--homozyg-window-snp`) с максимальным количеством гетерозиготных SNP (`--homozyg-window-het`) и максимальным количеством отсутствующих SNP (`--homozyg-window-missing`). Окно пошагово сканирует геном человека и оценивает для каждого SNP пропорцию, которая появляется в гомозиготном окне.

Во-вторых, сегменты гомозиготных SNP идентифицируются по всему геному с использованием порога для этих оценок для каждого SNP – частоты совпадений окна сканирования (`--homozyg-window-threshold`). Для размера окна в 100 SNP и порога 0,05 SNP должен появиться как минимум в пяти гомозиготных окнах, прежде чем он будет идентифицирован как часть сегмента. Такие гомозиготные окна могут содержать отсутствующие или гетерозиготные SNP, в зависимости от настроек окна сканирования.

В-третьих, для этих гомозиготных сегментов устанавливаются дополнительные ограничения для идентификации конечных сегментов ROH. Проверяется максимальный интервал между двумя SNP в сегменте (`--homozyg-gap`), а также максимальное количество гетерозиготных SNP, разрешенных в конечном сегменте ROH (`--homozyg-het`). Затем сегменты ROH, которые не соответствуют этим двум требованиям, разделяются и повторно оцениваются. Это может привести к обнаружению сегментов ROH меньших, чем размер окна сканирования. После этого оценивается минимальная плотность SNP (в kb/SNP) на сегмент (`--homozyg-density`), а также минимальная длина и количество SNP (`--homozyg-kb` и `--homozyg-snp`). Сегменты ROH, которые не удовлетворяют ни одному из этих трех условий, удаляются из анализа.

В литературе нет единого мнения о том, следует ли исключать SNP, на предмет неравновесия по сцеплению (LD – англ. linkage disequilibrium) и/или частоты минорных аллелей (MAF – англ. minor allele frequency) перед анализом ROH. В большинстве исследований на животных применяется фильтрация MAF с

порогом 0,01–0,05, а в некоторых исследованиях также выполняется удаление LD [66].

В исследовании [66] рекомендуют исключить применение фильтров MAF и LD перед идентификацией ROH на основе генотипов со средней плотностью покрытия SNP. Однако исследования LD показали, что размеры гаплотипов редко превышают 100 Кб у людей и что общая длина отдельных ROH, измеренная с обрезкой LD или без нее, одинакова, если рассматривать ROH более 1,5 м.п.о. [57,67,68].

Модельный подход для идентификации ROH, реализованный в программе Beagle, использует скрытые марковские модели (НММ – англ. Hidden Markov model) для учета фоновых уровней LD [69]. Чтобы сделать вывод о статусе гомозиготности «скользящих» окон у каждого человека, другие подходы, основанные на правдоподобии, используют логарифм отношения вероятностей генотипа в соответствии с гипотезами об аутозиготности и неаутозиготности, включая оценки частоты аллелей для конкретной популяции. На основе этих оценок логарифмических шансов определяется порог для каждой конкретной популяции, выше которого найденные регионы называются ROH. Распределение длин ROH моделируется как смесь трех гауссовых распределений, классифицирующих ROH по трем классам длин. Несмотря на повышенную чувствительность этого подхода при обнаружении более коротких ROH, необходимость оценки частот аллелей является его ограничением. Данный подход к идентификации ROH реализован в GARLIC [58]. На практике результаты модели смеси гауссовых распределений очень сильно коррелируют с результатами из PLINK [8]; однако популяционно-специфический характер границ классов ROH усложняет метаанализ. С другой стороны, такой подход помогает более точно охарактеризовать классы ROH с биологической точки зрения.

Среди многочисленных групп населения Северной Евразии, имеющих различную генетико-демографическую историю, особый интерес для исследования ROH представляют популяции коренных народов Северного

Кавказа и Сибири, характеризующиеся, как правило, небольшой численностью и значительной географической изоляцией от других популяций. В частности, ранее для популяций Дагестана, говорящих на языках нахско-дагестанской семьи, были показаны высокие коэффициенты инбридинга, очень большое количество и большая суммарная длина гомозиготных блоков, по сравнению с окружающими популяционными группами Кавказа, Ближнего Востока, Европы, Центральной и Южной Азии [70]. Для популяций Сибири по данным низкоплотных чипов нами ранее было показано значительно большее число и суммарная длина ROH по сравнению с популяциями Волго-Уральского региона, Кавказа, Европы и Средней Азии [71].

1.2.3 Коэффициент инбридинга

Традиционно оценку уровня инбридинга популяции получали из генотипирования выборок с последующей оценкой аллельных частот и гетерозиготности (наблюдаемую и ожидаемую) или из анализа генеалогического древа [72]. Степень индивидуального инбридинга измеряется с использованием (F), вероятности того, что индивидуум получает два аллеля, которые идентичны по происхождению в данном локусе [73], что также является ожидаемой пропорцией генома, который является аутозиготным, например $F = 0,0625$ для потомства двоюродных братьев. Чем выше степень инбридинга, тем ближе к единице находится коэффициент инбридинга, и стремится к нулю в случае с индивидами, имеющими отдалённых общих предков.

Для получения наиболее точного значения коэффициента инбридинга необходимо анализировать родословные не менее десяти поколений [74]. Однако на практике в редких случаях можно получить точную информацию о родословной более чем на три или четыре поколения назад, особенно если это касается популяционных выборок.

Ситуация резко изменилась в связи с возможностью использовать геномные данные, полученные с помощью высокоплотных чипов или полных геномов,

позволяющих идентифицировать как длинные, так и короткие регионы высокой гомозиготности в различных популяциях человека.

Исследования, использующие коэффициент инбридинга для анализа популяций, выявили высокий уровень аутозиготности даже в космополитических неинбредных популяциях, показав увеличение инбридинга и значительное сокращение генетического разнообразия, пропорционально географическому расстоянию от Африки [8,62,75]. Распределение ROH неравномерно по геному и для популяций из различных регионов выявлены разные значения коэффициента инбридинга [62,76].

Коэффициент геномного инбридинга (F_{ROH}) измеряет фактическую долю аутомного генома, определяемую как долю общей длины ROH идентифицируемых выше определенного порога от общей длины генома [57]. F_{ROH} определяет общий коэффициент инбридинга индивидуума независимо от изученности его родословной или ее отсутствия. F_{ROH} ограничен только плотностью покрытия генома и, следовательно, размерами ROH, которые можно идентифицировать. Недавние исследования показали, что у потомков двоюродных братьев с аутомно-рецессивным заболеванием среднее значение F_{ROH} составляло 11%, что было значительно выше прогнозируемого, вероятно, из-за наличия в поколениях близкородственных браков [77]. Анализ с использованием родословных из изолированной популяции показал, что F_{ROH} , рассчитанный с $ROH > 1,5$ м.п.о., наиболее сильно коррелирует ($r = 0,86$) с коэффициентами инбридинга, посчитанными из родословных по шести поколениям (F_{Ped} - Pedigree based inbreeding coefficient) [57]. Данные родословных обеспечивают только значения ожидаемой аутозиготности, в то время как ROH определяют фактическую аутозиготность.

Также было показано, что братья и сестры различаются в среднем на 10 м.п.о. по суммарной длине ROH. Неродственные индивидуумы без близкородственных браков, по крайней мере, от пяти до десяти поколений несли ROH до 4 Мб в длину, соответственно: эти ROH имеют более древний возраст происхождения [57]. Фактически, в разных выборках среднее значение

инбридинга, определенного по родословным, также хорошо коррелирует с F_{ROH} с использованием $\text{ROH} > 5$ м.п.о ($r = 0,87$), но не с F_{ROH} , рассчитанным по $\text{ROH} < 5$ м.п.о. [62].

Таким образом, уровни инбридинга являются информативными в отношении истории популяции, связанных с метисацией, изменением численности, а также могут быть связаны с ее генетической нагрузкой и распространенностью генетических заболеваний.

1.2.4 Блоки идентичные по происхождению

Сегменты генома, идентичные по происхождению, могут быть обнаружены в популяциях с использованием высокоплотных чипов и полных геномов. Сегмент, имеющий идентичные нуклеотидные последовательности, является IBD у двух или более лиц, если они унаследовали его от общего предка без рекомбинации, то есть у этих людей сегмент имеет общее происхождение. Ожидаемая длина сегмента IBD зависит от количества поколений с момента появления последнего общего предка в локусе сегмента.

Идентифицированные сегменты IBD могут использоваться для широкого круга задач. Одним из применений обнаружения сегмента IBD является количественная оценка степени родства, что может быть использовано в судебной генетике, и также может дополнить информацию о картографировании генетических связей на межиндивидуальном и межпопуляционном уровнях внутри различных групп [3].

Сегменты IBD используются для обнаружения естественного отбора в геноме человека. Отбор обычно увеличивает число сегментов IBD среди людей в популяции. Участки с избытком IBD помогают идентифицировать регионы в геноме человека, которые находятся под сильным недавним отбором [4,5].

Сегменты IBD могут использоваться для оценки демографической истории, включая эффект бутылочного горлышка и поток генов в популяциях [6]. В недавних исследованиях были показаны различия в распределении IBD между африканскими, азиатскими и европейскими популяциями, а также сегменты IBD,

общие с древними геномами, такими как геномы неандертальцев или денисовцев [7].

Близкие биологические родственники имеют идентичные друг с другом достаточно длинные фрагменты ДНК и, соответственно, в большинстве хромосом находятся идентичные по происхождению блоки значительной длины ($> 66,7$ сМ), в результате чего ожидаемая длина общих IBD составляет около 1700 сМ. Двоюродные и троюродные родственники, вероятно, будут иметь несколько регионов IBD (более 2,5 ожидаемых сегментов) из-за наличия недавних предков, определяющих их родство. Ожидается, что для двоюродных родственников каждый IBD будет иметь общую длину больше 62 сМ и для троюродных – 25 сМ. Более дальние родственники также с большой вероятностью наследуют один или несколько регионов IBD от своего ближайшего общего предка. Такие пары индивидов составляют большинство людей в популяции и обычно называются «неродственными», поскольку ожидается, что наличие IBD по всему геному будет незначительным. Также обычно, когда используется только один сегмент IBD, количество поколений до ближайшего общего предка отдаленных родственников не может быть надежно оценено [6].

1.3 Поиск сигналов отбора в популяциях человека

Покинув Африку около 50-100 тыс. лет назад, человек современного типа, быстро заселил чрезвычайно разнообразную среду обитания. Например, археологические исследования свидетельствуют о том, что люди жили в условиях, значительно отличающихся от условий Африки к югу от Сахары, таких как холодный климат в арктической Сибири [78] или высокогорные районы на Тибетском плато [79] уже 27 и 30 тыс. лет назад, соответственно. Помимо климатических отличий, культурные изменения, такие как внедрение сельского хозяйства и скотоводства, также способствовали расхождению в условиях обитания человека [80]. Фенотипическое разнообразие человеческих популяций в различных географических регионах является результатом дрейфа генов, метисаций, естественного отбора на протяжении всей эволюции человека.

Исследованию роли естественного отбора в генетическом разнообразии популяций были посвящены и работы ряда ведущих отечественных биологов в 60-е – 90-е годы прошлого века (работы Ю.П. Алтухова, Ю.Г. Рычкова, Л.А. Животовского, Л.Л. Соловечука и др.). В НИИ медицинской генетики исследования по оценке вклада естественного отбора в динамику генофондов популяций жителей Сибири активно велись со времени становления института на основе генетико-демографических и маркерных подходов [81,82].

Генетическая изменчивость человека, лежащая в основе биологической адаптации к селективному давлению, не полностью изучена. С появлением методов популяционно-генетического моделирования, использующих высокоплотные геномные данные, полученные для различных популяций, наблюдается значительный прогресс в обнаружении действия естественного отбора у человека [83–88]. Предполагаемые сигналы отбора могут быть использованы для идентификации функциональных аллелей-кандидатов, лежащих в основе предполагаемых адаптивных фенотипов. Точное картирование с привязкой к фенотипу облегчают идентификацию предполагаемых адаптивных аллелей. Многие адаптивные признаки, например такие как персистенция лактазы [89–93], пигментация кожи [94–98], адаптация к большим высотам над уровнем моря, развивались параллельно во многих человеческих популяциях, и, вероятно, сильно варьируют даже при аналогичном селективном давлении. Помимо обоснования эволюционных процессов, опосредующих адаптацию человека к меняющейся среде обитания, открытие и функциональная аннотация адаптивных аллелей также может иметь значение для поиска генов, вовлеченных в патогенез заболеваний человека.

Несмотря на высокую скорость накопления данных о генах, подверженных отбору, в общедоступных базах данных недостаточно информации в отношении многих популяций человека, что ограничивает изучение биологической изменчивости человека. Понимание генетической основы полезных признаков является одной из основных задач генетики человека [99].

Предположительно, генетическая структура коренного населения Сибири и Крайнего Севера сложилась под существенным влиянием естественного отбора: в наиболее неблагоприятных условиях среды обитания в арктическом климате Крайнего Севера и резко континентальном климате Сибири и Дальнего Востока. Длительная эволюция генофонда популяций Сибири, в том числе приспособительная, могла привести, с одной стороны, к формированию устойчивых адаптивных комплексов, а с другой – к специфическому спектру заболеваний как моногенной, так и мультфакторной природы.

Ряд недавних исследований показывает, что генетические системы, связанные с жировым метаболизмом, устойчивостью к температурным шокам, тонусом сосудистого русла, изменением инфекционной нагрузки [10,12,100–102] могут играть существенную роль в адаптации к неблагоприятным или экстремальным климато-географическим условиям. Переход от охоты к животноводству, смена типа питания на фоне резкого роста численности популяций могли привести к изменению структуры генофонда популяций по различным генам, продукты которых участвуют в липидном обмене, усвоении различных продуктов.

Многие из существующих методов для определения сигналов положительного отбора используют выявление гомозиготности гаплотипов [99,103–105], например, гомозиготность по расширенному гаплотипу (ЕНН, extended haplotype homozygosity) и интегрированную оценку гаплотипов (iHS, integrated haplotype score). Причина этого заключается в том, что по мере увеличения частоты предпочтительного аллеля, регион, в котором он находится, будет увеличивать гомозиготность и испытывать меньшую внутриаллельную рекомбинацию на уровне популяции. Следовательно, аллели, находящиеся под отбором, частоты которых увеличиваются, будут располагаться на более длинных чем ожидалось гаплотипах, с большей частотой в популяции. Аллели на разных гомологичных хромосомах идентичны по происхождению, если они являются прямыми копиями одного и того же предкового аллеля. Тест iHS оценивает степень гомозиготности гаплотипа (ЕНН) вокруг тестируемого SNP при

сравнении гаплотипов для предкового и производного аллелей. В пакете SelScan реализован усовершенствованный вариант теста iHS - статистика nSL (number of Segregating sites by Length). Ключевое отличие теста nSL от классического iHS заключается в том, что nSL сравнивает длину гомозиготных фрагментов на двух гаплотипах (для предкового и производного аллелей) в терминах числа мутаций (полиморфных сайтов), а не пар нуклеотидов [10,106]. Тем самым, nSL не зависит от привязки к генетической карте и является более устойчивым к эффектам вариабельности в частоте рекомбинации или темпе мутирования.

Методы обнаружения отбора на основе гомозиготности гаплотипов определяют регионы, где подмножество индивидуумов несет IBD, которые длиннее, чем можно было бы ожидать при отсутствии какого-либо отбора [107–109]. Поэтому были разработаны методы, которые позволяли делать выводы о блоках идентичных по происхождению, общих между людьми в популяции без какой-либо информации о их родословных, с использованием высокоплотных чиповых данных. Первоначальная цель этих методов была в том, чтобы идентифицировать регионы, приводящие к различным заболеваниям с наследственной компонентой, в которых IBD встречается чаще среди больных, чем среди здоровых людей. Кроме того, эти методы также можно использовать для изучения генетико-демографической истории популяций. Они обеспечивают выявление следов очень недавнего сильного отбора в геноме, и не только отбора, действующего на новый аллель, но также отбора, действующего на ближайшие расположенные аллели, частота которых увеличивается в популяции благодаря неравновесному сцеплению.

1.4 Этногенез коренных народов Сибири

Сибирь — обширный географический регион в азиатской части России, ограниченный с запада Уральскими горами, с востока — водораздельными хребтами, идущими вдоль Тихого океана, с севера — Северным Ледовитым океаном, с юга — государственной границей России (Казахстана, Монголии, Китая) [110]. Также Сибирь подразделяют на Западную, Восточную, Южную,

Северо-Восточную и Среднюю (Центральную) Сибирь. Как историческое понятие в своих широких границах Сибирь включает в себя и северо-восток Казахстана и весь Российский Дальний Восток.

В конце двадцатого века численность коренного населения Сибири не превышала 32 млн. человек, из них около 2 млн. являлись коренными жителями края. В настоящий момент по данным всероссийской переписи населения 2010 г. наиболее многочисленными народами Сибири являются якуты (478 тыс.), буряты (461 тыс.), тувинцы (265 тыс.), алтайцы (81 тыс.), хакасы (73 тыс.). Малочисленные народы представлены ненцами (включая европейские группы (44,6 тыс.)), хантами (30,9 тыс.), эвенками (37,8 тыс.), эвенами (22,4 тыс.), чукчами (15,9 тыс.), шорцами (12,9 тыс.), манси (12,2 тыс.), нанайцами (12 тыс.), коряками (7,9 тыс.), долганами (7,8 тыс.), сибирскими татарами (6,8 тыс.), нивхами (4,6 тыс.), селькупам (3,6 тыс.), ительменами и ульчи (около 3 тыс. каждый), кетами, юкагирами, эскимосами и удэгейцами (менее 2 тыс. каждый), чулымцами, нганасанами, тофаларами, энцами, алеутами, орочами, негидальцами и уйльта/ороками (менее 1 тыс. каждый) [111].

Коренные народы Сибири отличаются друг от друга антропологически, лингвистически, генетически, а также по культурным особенностям. Согласно концепции хозяйственно культурных типов и историко-культурных областей (ИКО), на территории России выделяются три большие историко-этнографические провинции (Восточно-Европейская, Сибирская и Дальневосточная).

Сибирскую историко-этнографическую провинцию разделяют на четыре историко-культурных области: 1) ЯмалоТаймырская ИКО (тундра и лесотундра Западной и Центральной Сибири, заселённые придерживающимися традиционных верований самодийскими народами – ненцами, энцами, нганасанами, а также тюркоязычными долганами); 2) Западно-Сибирская ИКО – бассейн среднего течения Оби, заселённый сохранившими традиционные верования народами уральской семьи (хантами, манси и селькупам) и уникальными в языковом отношении кетами; 3) Южно-Сибирская или Алтайско-

Саянская ИКО (территория проживания тюркских народов – алтайцев, хакасов, тувинцев, тофаларов и шорцев на Алтае и в Саянах, а также монголоязычных бурят вокруг озера Байкал, в значительной степени придерживающихся буддизма ламаистского направления); 4) Восточно-Сибирская ИКО, занимающая обширные пространства от Енисея до Охотского моря и заселённая первоначально юкагирами, а затем тунгусскими народами (эвенками и эвенами), частично сохранившими традиционные верования, и тюркоязычными якутами (самоназвание – саха), принявшими православную религию.

Дальневосточная историко-культурная провинция включает две историко-культурные области: 1) Чукотско-Камчатская ИКО, заселённая народами чукотско-камчатской семьи (чукчи, коряки, ительмены), и эскимосо-алеутской семьи (эскимосы и алеуты) – приверженцами традиционных верований; 2) Амуро-Сахалинская ИКО, где проживают тунгусо-манчжурские народы, частично сохранившие традиционные верования (нанайцы, ульчи, орочи, ороки, удэгейцы и негидальцы), а также своеобразные в языковом отношении нивхи.

В выделении некоторых ИКО природная зональность сыграла решающую роль. Например, Северная ИКО занимает тундровые и лесотундровые пространства Европейской части России, где проживают народы, традиционным занятием которых является кочевое оленеводство. В основе культуры Центрально-русской ИКО лежит хозяйственнокультурный тип «земледельцев лесной зоны и лесостепи». По этой же причине была специально выделена Южно-русская подобласть, которую отделяет от Центрально-русской области чёткий природный рубеж – граница лесостепной и степной зон. Таким образом, даже в самом выделении историко-этнографических областей и хозяйственно-культурных типов нашла отражение географическая зональность, определяемая преимущественно природными факторами и поэтому имеющая в основном широтный характер. Можно найти очевидные соответствия между природными зонами и основными территориями проживания коренных народов России, а также степенью их ассимиляции со стороны русского населения [112].

1.4.1 Антропологическая классификация

Специфической особенностью Сибири, как одного из регионов периферийного расселения современного человека, является высокое этническое разнообразие. Этногенез народов Сибири по данным антропологии и археологии начался в эпоху неолита или позднего палеолита [113]. Сибирская историко-этнографическая провинция охватывает огромную территорию, на которой расселены около сорока коренных народностей.

В настоящее время среди коренного населения Сибири выделяют следующие антропологические варианты [114].

Монголоидные варианты, не обнаруживающие посторонних примесей:

центральноазиатский, преобладающий среди бурят, большей части якутов и части населения Саяно-Алтайского нагорья (тувинцев-скотоводов, южных алтайцев; за пределами России – монголов);

байкальский (палеосибирский), характерен для эвенов, эвенков, юкагиров, нанайцев, тофаларов, ороков, оленных тувинцев; исторически он связан, по-видимому, с древним юкагирским населением Восточной Сибири;

арктический, характерен для эскимосов, чукчей, коряков, ительменов.

Монголоидные варианты с примесью других рас:

южносибирский, образовавшийся в результате смешения носителей центрально-азиатского типа с древним европеоидным населением; встречается у южных алтайцев, а за пределами Сибири – у казахов.

уральский, довольно разнородный антропологический тип, образовавшийся в результате смешения иных монголоидных и европеоидных форм; главным образом, распространен у хантов, манси, ненцев, селькупов, барабинских татар, шорцев, северных алтайцев и хакасов;

сахалино-амурский тип, преобладает у нивхов и частично у ульчей.

1.4.2 Языковые группы коренного населения Сибири

Языки Сибири — условное собирательное название для различных языков, на которых говорят коренные обитатели Сибири и Дальнего Востока России. В лингвистике отсутствует понятие «Сибирская языковая семья» или подобные, так как эти языки отличаются как генетическим, так и типологическим разнообразием.

Языки Сибири относятся к 7 языковым семьям: уральской, енисейской, алтайской, юкагиро-чуванской, эскимосско-алеутской, чукотско-камчатской, айнской; нивхский язык является изолированным.

Алтайская языковая семья является одной из самых многочисленных языковых семей на территории Сибири. В нее входят три языковые ветви: тюркская, монгольская и тунгусо-манчжурская. Тюркская группа включает языки народов Саяно-Алтая. Алтайцы расселяются с запада на восток Южной Сибири. По материалам переписи 2002 г. впервые учтены в качестве самостоятельных этносов телеуты, тубалары, теленгиты, кумандинцы и другие. На востоке Южной Сибири находятся шорцы, хакасы, тувинцы, тофалары [115].

На Западной Сибири расселяются западносибирские татары, в составе которых выделяются группы барабинских, тарских, томских и прочих татар.

Народы, говорящие на языках урало-юкагирской языковой семьи, расселяются на территории Западной Сибири. В самодийскую группу входят ненцы (зона лесотундры и тундры от Полярного Урала на западе до Енисейской губы на востоке), энцы (правобережье Енисейской губы) и нганасаны (Таймырский полуостров). Селькупы, коренной малочисленный народ, проживают в западносибирской тайге на Средней Оби и в бассейне р. Таз.

Угорская группа в Западной Сибири представлена народами ханты и манси. Ханты расселены в бассейне Оби и ее притоках от лесотундры до лесостепи. Этническая территория манси простирается от Урала до левобережья Оби. Недавно в состав уральской языковой семьи был добавлен язык юкагиров, проживающих в Восточной Сибири в бассейне р. Колыма. Несмотря на

территориальную удаленность, еще в XIX в. лингвисты отмечали уралоидный субстрат в языке этого народа, как отражение древних переселений уралоязычных народов.

Якуты заселяют значительную часть Восточной Сибири (бассейны Лены, Анабары, Оленека, Яны, Индигирки). Самый северный тюркоязычный народ мира, проживающий на юге Таймыра — долганы. Буряты и сойоты принадлежат к монголоязычным народам Сибири.

Тунгусо-маньчжурские языки широко распространены в таежной зоне Восточной Сибири от Енисея до Камчатки и Сахалина. Это языки северных тунгусов — эвенков и эвенов. В бассейне р. Амур, живут нанайцы, ульчи, ороки (остров Сахалин), говорящие на языках, относящихся к южной, амурской или маньчжурской ветви тунгусо-маньчжурской группы. Негидальцы расселяются по берегам левого притока Амура р. Амгуни. Удэгейцы и орочи живут в горах Сихотэ-Алиня и к побережью Японского моря Приморского края.

Палеоазиатские народы (чукчи, коряки и ительмены) населяют северо-восток Сибири, Чукотку и Камчатку. С учетом ряда признаков сходства они были выделены в чукотско-камчатскую языковую семью. Внутри неё большее родство наблюдается между языками чукчей и коряков. Ительменский язык по отношению к ним демонстрирует не столько генетическое, сколько ареальное соответствие.

Представители эскимосско-алеутской семьи в основном расселяются за пределами России (Канада, США). На Северо-Востоке Сибири небольшие группы азиатских эскимосов проживают на побережье Чукотского моря, Анадырского залива и острова Врангеля. На Командорских островах проживает коренное население Алеутских островов — алеуты.

Язык нивхов, проживающих на Амурском лимане и на севере острова Сахалин и кетов, проживающих в бассейне р. Енисей, относят к категории изолированных. Нивхский язык ранее относили к группе палеоазиатских языков в связи с нечеткой выраженностью генеалогического начала. Кетский язык лингвисты относят к единственному представителю енисейской языковой семьи.

Другие носители енисейских языков (арины, асаны, яринцы и др.) в течение XVIII—XIX вв.. переселившиеся в верховья Енисея и его притоков, были ассимилированы соседними народами [116].

1.5 Заключение по обзору литературы

Успехи популяционной генетики в области изучения генетического разнообразия обусловлены как обнаружением различных типов информативных генетических маркеров, так и активным развитием технологий их генотипирования. Новые подходы в детекции различных типов ДНК-полиморфизма (в особенности SNP) разрабатываются очень быстро, и на сегодняшний день существует большое количество протоколов генотипирования разных маркерных систем. В течение двух последних десятилетий произошел настоящий прорыв в исследовании генетической вариабельности человеческих популяций как по нейтральным маркерам, так и в области медицинской генетики, где в настоящее время рутинными стали работы по полногеномному поиску ассоциаций различных заболеваний с сотнями тысяч SNP. Результатом этого становится все увеличивающийся объем фактической информации по индивидуальным генотипам и целым геномам, для адекватной статистической обработки которого подчас требуются мощности высокопроизводительных компьютеров и большие системы хранения данных.

Современные методы геномики и биоинформатики дают богатейшие и разнообразные возможности для анализа популяционного генофонда и позволяют вывести этногенетические исследования на совершенно новую ступень. В частности, с помощью современных методологических подходов мы можем подробно охарактеризовать структуру, компонентный состав, подразделенность, популяционную и этническую историю народов, населяющих различные регионы. На основе широкогеномного генотипирования SNP имеется возможность очень детально изучить паттерны гаплотипического разнообразия, маркирующие различные субстратные и суперстратные пласты генофонда популяций, степень метисации с пришлым населением на различных уровнях – от индивидуального

до родового и этнического, – провести детальный анализ эволюционной и демографической истории этносов. Имеющаяся практика применения молекулярно-генетических исследований даёт возможность реконструировать и достаточно точно датировать во времени генетико-демографические события, произошедшие в прошлом. Применение современных биоинформационных подходов на широком массиве собственных и доступных из открытых источников генотипических данных по полным геномам, высокоплотным биочипам позволяет точно и достоверно реконструировать формирование структуры генофондов всех исследуемых этносов и генетико-демографические процессы, которые привели к ее формированию.

Одним из наиболее актуальных направлений этногеномных исследований является изучение региональной и внутриэтнической структурированности генетического разнообразия коренного населения разных регионов. Генофонд населения Сибири представляет собой уникальную систему с точки зрения исследования популяционно- и эволюционно-генетических процессов, анализа генетического разнообразия и реконструкции генетической истории популяций. Огромные пространства этого региона и малочисленность аборигенного населения способствовали формированию его значительной территориальной и генетической подразделенности. Относительная изоляция населения отдельных регионов, таких как Сибирь и Дальний Восток от основных миграционных потоков на территории Евразии обеспечила сохранение в их генофонде древних генетических пластов, хранящих информацию о важнейших этапах заселения человеком современного типа не только Сибири, но и Европы, Центральной Азии и Америки. Большинство народов несут в своих геномах различные предковые компоненты, которые различаются по времени появления на той или иной территории, направлениям миграций их носителей, их антропологической и языковой принадлежности. Современные методы, применяемые в молекулярно-генетических исследованиях, и новые биоинформационные подходы позволяют достоверно выявлять различные предковые генетические компоненты в составе генофонда различных народов и отдельных людей, реконструировать миграционные события, генетико-демографические процессы, эффекты основателя и адаптацию к условиям обитания.

ГЛАВА 2 МАТЕРИАЛЫ И МЕТОДЫ ДИССЕРТАЦИОННОГО ИССЛЕДОВАНИЯ

2.1 Характеристика популяционных групп

Материал исследования составили образцы ДНК, выделенные из лейкоцитов крови, принадлежащих к 20 популяциям коренного населения Сибири включающих 15 этнических групп. На рисунке 2 приведена локализация исследованных этнических выборок коренного населения Сибири. Для изучения взаимодействия и влияния популяций из близлежащих регионов на популяции Сибири, а также в качестве групп сравнения были взяты образцы из Средней Азии, Волго-Уральского региона, Восточной Европы, Северного Кавказа и Дагестана. В совокупности охарактеризовано 76 популяционных выборок, принадлежащих к 62 этническим группам. Проведение исследования одобрено Комитетом по биомедицинской этике НИИ медицинской генетики Томского НИМЦ (№4 от 21 ноября, 2017). Все эксперименты были проведены на базе Центра коллективного пользования научно-исследовательским оборудованием с использованием ресурсов биологической коллекции «Биобанк населения Северной Евразии» (НИИ медицинской генетики Томского НИМЦ, г. Томск).

Большая часть материала была собрана в ходе экспедиций, организованных НИИ медицинской генетики, проводившихся сотрудниками института, а также в рамках договоров о совместных научных исследованиях с другими российскими научными коллективами. При заборе образцов крови соблюдалась процедура информированного согласия доноров. Образцы крови отбирались в вакуумные пробирки с ЭДТА. На каждого донора составлялась анкета с краткой родословной, указанием его национальности и мест рождения предков. Собранные образцы определяли к различным этническим группам на основании национальности донора и национальности родителей.

Изначальная общая численность суммарной выборки исследованных популяций составляет более 2100 образцов. После исключения родственных

индивидуумов, в результате анализа IBD и родословных, а также «выбросов» в результате анализа t-SNE осталось 1836 индивидуумов.

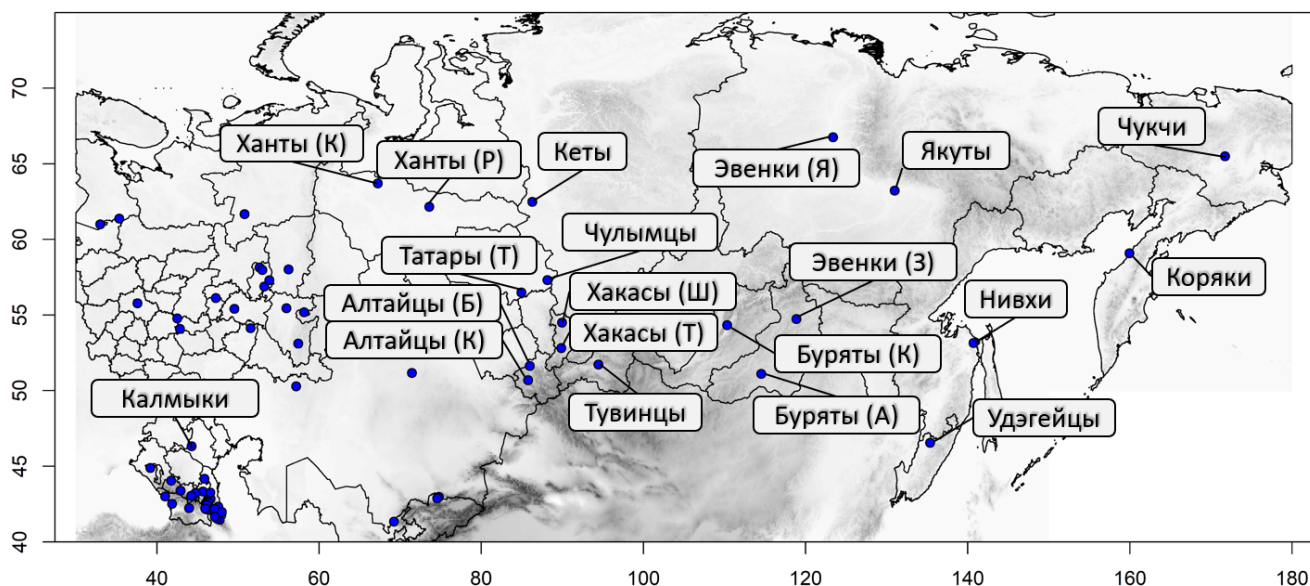


Рисунок 2 - Локализация исследованных этнических выборок коренного населения Сибири.

Популяции коренного населения Сибири ($N = 477$) представлены алтайцами (Б – с.Бешпельтир Чемальский муниципального района, $N = 24$ и К – с.Кулада Онгудайский района, $N = 25$) (материал собран сотрудником республиканской детской больницы Республики Алтай канд. мед. наук В.Н. Тадиновой, а также в ходе экспедиций НИИ МГ СО РАМН), бурятами (А – п.Агинское Агинского района, $N = 23$ и К – с.Курумкан Курумканского района, $N = 28$) (материал собран канд. мед. наук Е.Р. Ереминой), калмыками ($N = 29$), кетами ($N = 15$) (материал собран в ходе экспедиционных работ НИИ МГ СО РАМН студентом Харьковым В.Н. и аспирантом Хитринской И.Ю. в 2001г.), коряками ($N = 20$), чукчами ($N = 25$) (популяционные выборки коряков и чукчей предоставлены директором НИИ терапии СО РАМН, член-корр. РАМН М.И. Воеводой (Новосибирск). Материал коряков собран в Корякском автономном округе Камчатской области. Популяционная выборка чукчей, образцы крови которых собраны в пос. Лорино, Сиреники, Янарыкот и Новое Чаплино Чукотского автономного округа, относятся к береговой группе), нивхами ($N = 13$), татарами (Т – г.Томск $N = 20$), тувинцами (

N = 28) (материал собран в ходе экспедиций НИИ МГ СО РАМН, а также сотрудниками многопрофильной научной лаборатории по медико-биологическим проблемам Министерства здравоохранения Республики Тува, руководитель д-р. мед. наук Э.А. Ондар), удэгейцами (N = 15) (материал собран канд. биол. наук М.Г. Спиридоновой и канд. биол. наук А.В. Марусиным), хантами (К – с.Казым Белоярского района, N = 30 и Р – д.Русскинская Сургутского района = 26) (материал собран автором, канд. биол. наук М.Г. Спиридоновой и канд. биол. наук А.В. Марусиным, канд. биол. наук В.Н. Харьковым), хакасами (Т – сагайцами Таштыпского р-на N = 29 и Ш – качинцами Ширинского р-на N = 26) (материал был собран в ходе совместной научно-практической медицинской экспедиции, проводившейся в 2007-2008 гг. в рамках Договора о совместной деятельности между НИИ медицинской генетики СО РАМН и ГОУ ВПО Хакаский государственный университет), чулымцами (N = 22) (материал собран в ходе экспедиционных работ НИИ МГ СО РАМН в 2019 г.), эвенками (З – забайкальские (п.Чара Каларского района, с. Моклакан и с.Тупик Тунгиро-Олёкминского района) N = 25 и Я – якутские N = 28) и якутами (N = 26). Материал якутских выборок собран в ходе совместных экспедиционных работ сотрудниками НИИ МГ СО РАМН (канд. биол. наук Л.А. Конева, канд. биол. наук Н.П. Бабушкина), НИИ кардиологии ТНЦ СО РАМН (К.В. Пузырев) и сотрудниками Якутского научного центра РАМН и Правительства Республики Саха (Якутия) (д-р. мед. наук Н.Р. Максимова, канд. мед. наук А.Н. Ноговицына).

Локализация изученных популяций коренного населения Сибири, число обследованных индивидов, а также лингвистические и антропологические характеристики этнических групп приведены в таблице 1.

Таблица 1 - Характеристики изученных популяционных выборок коренного населения Сибири

Популяции	Локализация	Объем выборки	Языковая принадлежность (Семья/группа)	Антропологический тип
Алтайцы (Б)	Республика Алтай	24	Алтайская/тюркская	Монголоидный (центральноазиатский)
Алтайцы (К)	Республика Алтай	25	Алтайская/тюркская	Монголоидный (центральноазиатский)
Буряты (А)	Забайкальский край	23	Алтайская/монгольская	Монголоидный (центральноазиатский)
Буряты (К)	Республика Бурятия	28	Алтайская/монгольская	Монголоидный (центральноазиатский)
Чукчи	Чукотский автономный округ	25	Чукотско-камчатские языки	Монголоидный (арктический)
Чулымцы	Томская область	22	Алтайская/тюркская	Уральский, Монголоид-ный (южносибирский)
Эвенки (Я)	Якутия	28	Алтайская/Тунгусо-маньчжурские языки	Монголоидный (байкальский)
Эвенки (З)	Забайкальский край	25	Алтайская/Тунгусо-маньчжурские языки	Монголоидный (байкальский)
Калмыки	Республика Калмыкия	29	Алтайская/монгольская	Монголоидный (центральноазиатский)
Кеты	Красноярский край	15	Енисейская	Енисейский
Хакасы (Ш)	Республика Хакасия	26	Алтайская/тюркская	Уральский, Монголоид-ный (южносибирский)
Хакасы (Т)	Республика Хакасия	29	Алтайская/тюркская	Уральский, Монголоид-ный (южносибирский)
Ханты (К)	Ханты-Мансийский	30	Уральская/Обско-угорская	Уральский

Продолжение таблицы 1

Популяции	Локализация	Объем выборки	Языковая принадлежность (Семья/группа)	Антропологический тип
	автономный округ — Югра			
Ханты (Р)	Ханты- Мансийский автономный округ — Югра	26	Уральская/Обско-угорская	Уральский
Коряки	Камчатский край	20	Чукотско-камчатские языки	Монголоидный (арктический)
Нивхи	Хабаровский край	13	Палеоазиатских язык.	Монголоидный (сахалиноамурский)
Татары (Т)	Томская область	20	Алтайская/тюркская	Уральский, Монголоид-ный (южносибирский)
Тувинцы	Республика Тыва	28	Алтайская/тюркская	Монголоидный (центральноазиатский)
Удэгейцы	Приморский край	15	Алтайская/Тунгусо- маньчжурские языки	Монголоидный (байкальский)
Якуты	Якутия	26	Алтайская/тюркская	Монголоидный (центральноазиатский)

Коренные популяции Дагестана ($N = 518$) включают: аварцев ($N = 24$), даргинцев ($N = 28$), лакцев ($N = 24$), табасаран ($N = 21$), лезгин ($N = 28$), агулов ($N = 24$), рутульцев ($N = 22$), цахуров ($N = 24$), арчинцев ($N = 24$), андийцев ($N = 17$), ахвахцев ($N = 24$), тиндинцев ($N = 18$), каратинцев ($N = 24$), багулалов ($N = 23$), ботлихцев ($N = 16$), хваршин ($N = 15$), цезов ($N = 24$), бежтинцев ($N = 22$), чамалинцев ($N = 24$), гунзибцев ($N = 17$), гинухцев ($N = 19$), кумыков ($N = 37$) и караногайцев ($N = 19$). Материал на территории Дагестана собран в ходе экспедиций с 2001 по 2019 года сотрудниками НИИ МГ СО РАМН (д.б.н. В.Н. Харьков), а также сотрудниками ФГБОУ ВО ДГМУ Минздрава России под руководством к.б.н. М. О. Раджабова.

Волго-Уральский географический регион и популяции Европы ($N = 419$) представлены: коми ($N = 30$), марийцами ($N = 30$)(коми и марийцы предоставлены директором Института общей генетики им. Н.И.Вавилова РАН чл.-корр. РАН, д-ром. биол. наук., проф. Н.К. Янковским (Москва)), вепсами ($N = 30$), русскими ($N = 33$), удмуртами (удмурты $N = 30$, Б – п. Балезино $N = 28$ и Ш – с. Шаркан $N = 18$ (материал собран в ходе экспедиции сотрудниками НИИ МГ СО РАМН канд. биол. наук М.Г. Спиридоновой и канд. биол. наук А.В. Марусиным)). Материал выборки чувашей ($N = 26$), башкир (Б – Бурзянский р-н $N = 34$, П – пермские $N = 15$ и С – Салаватский р-н $N = 15$), бесермян ($N = 16$), карелов ($N = 29$), мордвы (Э – эрзя $N = 16$, М – мокша $N = 30$, Ш – шокша $N = 14$), татаров (К – казанские $N = 33$) предоставлен сотрудниками Уфимского федерального исследовательского центра Российской академии наук (д.б.н. Э.К. Хуснутдинова, к.б.н. С.С. Литвинов).

Коренные популяции Кавказского региона (за исключением Дагестана) ($N = 294$) включают: абхазов ($N = 30$), адыгейцев ($N = 10$), балкарцев ($N = 50$), ингушей ($N = 30$), карачаевцев ($N = 22$), мегрелов ($N = 28$), ногайцев ($N = 20$), осетин (С – северные $N = 30$, Ю – южные $N = 17$), черкесов ($N = 30$) и чеченцев ($N = 27$). Материал выборки Кавказского региона предоставлен сотрудниками Уфимского федерального исследовательского центра Российской академии наук (д.б.н. Э.К. Хуснутдинова, к.б.н. С.С. Литвинов).

Популяции северных киргизов ($N = 28$) (северные киргизы представлены жителями г. Бишкек и села Кегеты Чуйского района Республики Киргизия (материал собран сотрудником Киргизской государственной медицинской академии канд. биол. наук Ж.О. Солтобаевой), узбеков ($N = 22$), казахов (М – младший жуз $N = 29$, С – старший жуз $N = 26$) и дунган ($N = 23$), относятся к популяциям Средней Азии ($N = 128$).

2.2 Подготовка материала для генотипирования

Выделение геномной ДНК из периферической крови проводилось с помощью метода фенол-хромового экстракции. Хранение образцов геномной ДНК производилось в ТЕ-буфере при -20°C . Оценка качества и количества геномной ДНК выполнена с помощью спектрофотометра NanoDrop 1000 (Thermo Scientific, США). Отношение поглощения при длинах волн 260 нм и 280 нм (260нм/280нм) показывает чистоту препарата ДНК. Препарат считается чистым, если отношение значений 260нм/280нм приблизительно равно 1,8. В случае меньших значений показателя 260нм/280нм препарат содержит большое количество примесей белка, фенола или иных контаминирующих агентов, имеющих значительное поглощение при 280 нм. В качестве первичной оценки качества геномной ДНК был проведен гель-электрофорез (1% агарозный гель с содержанием бромистого этидия). Производитель микрочипов Illumina рекомендует более точные флуориметрические методы количественного определения двунитевой ДНК. В настоящем исследовании для данной задачи использован прибор Qubit® Fluorometer и набор Qubit® dsDNA HS Assay Kit по протоколу производителя набора. Было проведено разведение выделенного материала до необходимой концентрации, соответствующей требуемому производителем микрочипов количеству ДНК (200 нг). Далее были проведены этапы в микропланшете включающие амплификацию ДНК (Pre-Amp), инкубацию ДНК (Post-Amp), фрагментацию ДНК (Post-Amp), преципитацию (Post-Amp) и ресуспендирование ДНК (Post-Amp). Следующим этапом образцы наносились на микрочипы BeadChip и разделялись изолятором IntelliHyb. Загруженный

микрочип BeadChip инкубировался в течение ночи в термостате Illumina Hybridization Oven. Образцы амплифицированной и фрагментированной ДНК отжигались на локус-специфические 50-мерные олигонуклеотиды во время гибридизации. Негибридованная и неспецифически гибридизованная ДНК отмывалась и микрочипы BeadChip подготавливались к окрашиванию и однонуклеотидному удлинению (ОНУ). Однонуклеотидное удлинение олигонуклеотидов на микрочипе с использованием гибридизованной ДНК в качестве матрицы встраивало детектируемые метки на микрочип BeadChip, что позволило установить генотип образца. Процесс XStain проводился в специальной капиллярной проточной ячейке.

Сканирование микрочипа BeadChip выполнялось с помощью системы Illumina HiScan, использующей лазер для возбуждения флюорофора в составе продукта однонуклеотидного удлинения на микробусинах и детектирующей свет, испускаемый флюорофором в каждой точке микрочипа.

2.3 Генотипирование

Полногеномное генотипирование ДНК-маркеров было проведено с использованием платформы Infinium Multi-Ethnic Global BeadChip (Illumina). Данная микрочиповая платформа была создана для анализа различных по этническому происхождению образцов с целью решения широкого спектра задач в области эпидемиологии, популяционной генетики, охраны здоровья и геномики.

Платформа Infinium Multi-Ethnic Global BeadChip включает свыше 1,7 миллиона отобранных экспертным анализом SNP-маркеров, и это позволяет проводить идентификацию генетической ассоциации с широко распространенными и редкими фенотипами, с учетом генетической архитектуры популяций.

Микрочипы Infinium Multi-Ethnic Global BeadChip выполнены по технологии Infinium LCG Assay. Для детекции изображений и считывания сигналов в каждой точке микрочипов BeadChip использовался специальный ридер чипов iScan System (Illumina). Оперативные процедуры получения изображений

микрочипов, детекции и обработки сигналов выполнялись в соответствии со стандартным руководством производителя (iScan System User Guide).

2.4 Биоинформатическая обработка данных

2.4.1 Оценка качества генотипирования

Кластеризацию массива генотипов и контроль качества выполняли с использованием протокола, опубликованного в статье Guo Y. [117] с использованием GenomeStudio (модуль генотипирования v2.0.3), программного пакета, который Illumina разработала для анализа полученных геномных данных.

На данном этапе исследования проводилась оценка качества генотипирования SNP с помощью микрочипов. Было использовано два типа контроля качества: 1) относительно SNP – то есть, в пределах прогенотипированных SNP во всей выборке; 2) относительно образцов – то есть, все прогенотипированные SNP в пределах одного образца

При биоинформатической обработке полученных данных из анализа были исключены SNP с более чем 10% отсутствующими генотипами. Были исключены все позиции делеций и инсерций. Данные были предварительно отфильтрованы по минимальной частоте редкого аллеля (MAF, minor allele frequency > 0.01). В итоге, после фильтрации исходных данных по 1779819 маркерам, в финальный массив данных вошло 886889 аутомных SNP. Кроме этого, были исключены образцы, имеющие более 5% пропущенных SNP, а также образцы метисов первого, второго и третьего поколения, выявленных с помощью методов определения общих IBD, построения t-SNE и Admixture. Данная фильтрация позволяет повысить качество выборки повысив точность за счет исключения близкородственных индивидов.

2.4.2 Фазирование

Техническое определение понятия фазирования генотипа в контексте биоинформатики – это выявление или выводение (инференция) гаплотипной фазы массива генотипных данных. Основным преимуществом фазированных генотипов является более высокая степень вероятности общего происхождения гаплоблоков, которые в случае с идентичными по происхождению сегментами предполагают единую генеалогию. Одним из применений обнаружения сегмента IBD является количественная оценка степени родства, что также может дополнить информацию о генетических связях популяций. Фазирование генотипов проведено с использованием программного обеспечения Beagle 4.1[118].

2.5 Анализ компонентного состава генофонда методами PCA, t-SNE и Admixture

Для понижения размерности и пространственного анализа был использован метод главных компонент (PCA – Principal components analysis) [50,51]. Для выполнения задачи по выявлению и оценке индивидуальных компонентов и количества примесей в генофонде изучаемых популяций была использована программа Admixture [52]. Для визуальной дифференцировки и оценки кластеров была использован метод стохастического вложения соседей с t-распределением (англ. t-distributed Stochastic Neighbor Embedding, t-SNE) [119].

2.6 Выявления общих по происхождению блоков сцепления (IBD-сегментов)

Для анализа идентичных по происхождению блоков генома использовали алгоритм Refined IBD [120]. Использовалась программа версии refined-ibd.16May19.ad5.jar. Для сравнения популяций были получены суммы средних длин сегментов IBD между парами индивидов. Тепловая карта с дендрограммой на основании логарифма от суммы средних длин сегментов IBD между парами

индивидов для $IBD > 1.5$ cM, 1-4 cM и 4-10 cM построена с использованием пакета heatmap.2 в программной среде R.

2.7 Выявления регионов высокой гомозиготности (ROH)

2.7.1 Идентификация регионов высокой гомозиготности с помощью программы PLINK

Для нахождения участков, которые с большей вероятностью представляют аутозиготность, перед поиском ROH в программе «PLINK v1.9» [65] мы удалили SNP с высоким неравновесием по сцеплению (LD) (англ. LD – linkage disequilibrium). Для этого использовали сканирование генома в рамке из 50 SNP со скользящим окном 5 SNP и порогом попарного коэффициента сцепления r^2 порог = 0,8). В результате контроль прошли 585 715 SNP. Анализ был выполнен с учетом трех различных наборов ROH с учетом минимальной длины 0.5 м.п.о., 1,5 м.п.о. и 5 м.п.о. (малые, средние и большие ROH).

Для межпопуляционных сравнений были рассчитаны средние значения количества и совокупной длины ROH на геном для каждой популяции. Статистический анализ проводился в программной среде R. Статистическую значимость межпопуляционных различий оценивали при помощи теста Вилкоксона. Для коррекции уровня значимости по причине множественных сравнений использовалась поправка Бонферрони (уровень значимости = 0,05).

2.7.2 Расчет коэффициента геномного инбридинга

Для расчета коэффициента геномного инбридинга F_{ROH} у каждого индивидуума размеры аутосомных областей ROH, превышающих 1,5 м.п.о., суммировали и затем делили на общую длину аутосом, за исключением центромер (2708,613716 м.п.о. для сборки генома hg19). Как показано ранее F_{ROH} для 1.5 м.п.о. лучше коррелирует с оценками инбридинга, полученными из анализа родословных [57].

2.7.3 Идентификация регионов высокой гомозиготности с помощью программы GARLIC

Для оценки возраста ROH применяли программу GARLIC [58], использующую метод, основанный на вероятностной модели для ROH в классах с «коротким» (А, до 0,5 м.п.о.), «средним» (В, 0,5 – 1,5 м.п.о.) и «длинным» (С, более 1,5 м.п.о.) размерами. Короткие ROH отражают гомозиготность по древним гаплотипам, средние ROH являются результатом недавних демографических изменений в популяции (например, эффект бутылочного горлышка). Длинные ROH отражают гомозиготность возникшую в результате недавних кровнородственных браков.

Мы позволили GARLIC автоматически выбирать размер окна (--auto-winsize), начиная со значения, указанного параметром «--winsize 40» и увеличиваясь с шагом 2 до завершения, и также указали коэффициент ошибок 0,001. Мы использовали всех индивидуумов в популяции для оценки уровня детализации KDE (--kde-subsample 0). Поскольку анализируемые популяции различаются по размеру выборки, мы использовали одно и то же рекомендуемое значение для количества повторных выборок для оценки частот аллелей для каждого анализа (--resample 40). Все остальные параметры были установлены по умолчанию.

2.7.4 Выявление горячих точек ROH

Для изучения свойств горячих точек ROH мы сосредоточились на областях генома, которые содержали SNP в «крайних» хвостах (> 99,5%) эмпирического распределения общих частот ROH, рассчитанных для трех классов размеров ROH. SNP с частотами ROH выше 99,5-го перцентиля попадают в 69 областей генома, идентифицированных как горячие точки ROH (таблица А.1), при этом, если расстояние между SNP было меньше 300 т.п.н., регионы были объединены в один регион.

Данные о генах, расположенных в горячих точках ROH, были получены с помощью ANNOVAR (версия: 2020-06-08 00:46:07 -0400 (понедельник, 8 июня 2020 г.))

2.8 Выявление сигналов естественного отбора в высокоплотных чиповых данных

Поиск сигналов направленного отбора в популяциях Сибири, Северной Азии и севера европейской части России на основе анализа данных по полногеномному набору SNP проводили в пакете SelScan с помощью геномных тестов на протяженную гомозиготность гаплотипов (iHS, nSL и XP-EHH). Ожидаемая доля ложных отклонений FDR (англ. False discovery rate) была больше или равна 0.05.

Белок-белковые взаимодействия продуктов генов под отбором были визуализированы и проанализированы с помощью ресурса STRING (Search Tool for the Retrieval of Interacting Genes/Proteins). Обогащенность KEGG-путей генами под отбором была оценена с помощью ресурса webgestalt (WEB-based Gene Set AnaLysis Toolkit).

ГЛАВА 3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ структуры генофондов популяций человека на уровне геномов с использованием чипов является одним из современных и результативных подходов к изучению генетического разнообразия различных популяций который позволяет проводить подробное изучение истории популяций, адаптации, влияния естественного отбора. Накопление данных по различным популяциям привело к быстрому появлению новых аналитических и биоинформационных подходов в популяционной геномике. Эти подходы значительно расширили возможности анализа эволюционных, демографических и адаптивных механизмов в популяциях. Полученные результаты исследования существенно дополняют общемировую картину по популяционной геномике человека, закрывая значительный пробел для популяций Северной Евразии.

Выполненная работа посвящена решению некоторых фундаментальных проблем общей генетики человека и генетики популяций: изучения генетического разнообразия и структуры генофондов популяций человека, механизмов ее формирования, реконструкции миграций и расселения человека в аспекте анализа данных генотипирования высокоплотных биочипов с применением последних достижений в области обнаружения и генотипирования новых SNP-маркеров. В результате был выполнен анализ структуры генофондов коренных народов Сибири и реконструкции генетико-демографических процессов, приведших к особенностям генофонда локальных и региональных групп этносов, а также реконструкции демографической истории популяций: роста численности, древних и исторических миграций, анализу компонентного состава, уровню геномного инбридинга, реконструкции миграций и взаимоотношений между популяциями и различными народами с применением современных биоинформатических методов анализа. Для реализации задач исследования и аналитической обработки данных в работе были использованы современные методы биоинформатической обработки данных.

3.1 Анализ главных компонент

Для понижения размерности данных и выявления генетических компонент в составе генофондов и структуре исследуемых популяций мы остановились на трех алгоритмах: PCA, Admixture и t-SNE.

Полученные данные о частотах SNP в исследованных популяционных выборках были использованы для выяснения генетических взаимоотношений между различными этносами и региональными группами. Анализ с помощью метода главных компонент (PCA) позволил выявить интегральные характеристики суммарного генофонда. Метод главных компонент – классический и достаточно универсальный алгоритм, который демонстрирует для первых компонент картину, значительно схожую с географическим распределением популяций. Анализ частот распределения SNP с помощью метода главных компонент позволил выявить интегральные характеристики их распределения в исследованных популяциях.

На две первые главные компоненты приходится 60,27% и 9,97% вариабельности частот SNP у изученных этносов. Первая и вторая главные компоненты эффективно дифференцируют исследованные выборки. В пространстве PC1-PC2 выявляется несколько кластеров популяций (рисунок 3). Наблюдается соответствие картины распределения образцов с их географической локализацией. Первая компонента согласуется с географической долготой, вторая с широтой. Распределение отдельных образцов на графике хорошо согласуется с их национальной принадлежностью, территориальной и этнической близостью исследованных выборок.

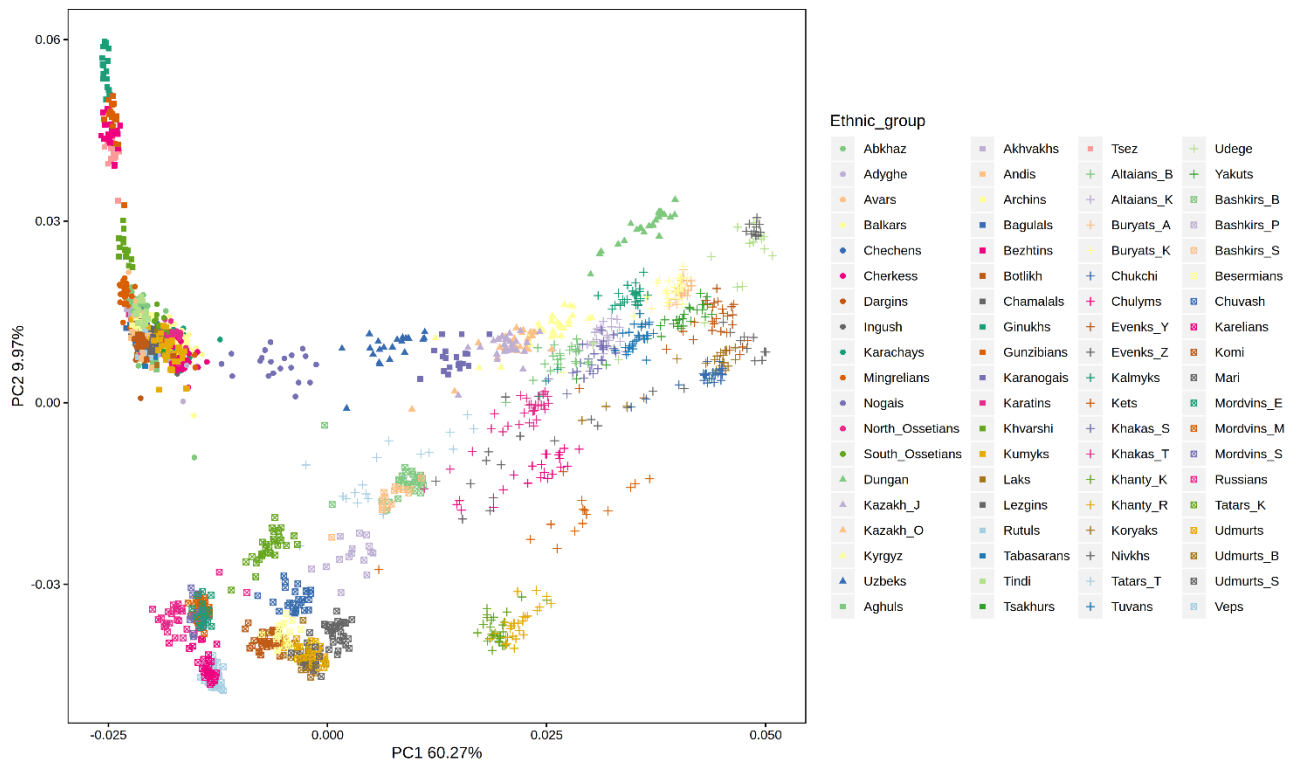


Рисунок 3 - Две первые главные компоненты массива геномных данных по популяциям Северной Евразии.

В целом, метод PCA дифференцирует все этносы на несколько групп генетически и географически близких популяций. Южносибирские этносы со значительной долей западноевразийских гаплогрупп (северные и южные алтайцы, телеуты, томские татары, хакасы) кластеризуются вместе. Хакасы при этом наиболее генетически удалены от остальных. Буряты, эвенки, якуты также близки друг к другу. Чукчи, коряки, нивхи и удэгейцы составляют дальневосточную группу генетически близких популяций, при этом коряки, нивхи и удэгейцы демонстрируют наибольшее родство.

Все образцы из популяций Северного Кавказа и Дагестана образуют отдельную тесную группу, из которой выбиваются бежтинцы, гинухцы, цезы и гунзибцы, с одной стороны, и ногайцы с караногайцами, с другой стороны.

Популяции Волго-Уральского региона имеют более значительный градиент, который разделяет их на территориальные этнические группы. Наиболее гетерогенными являются образцы этнических башкир, что полностью совпадает с

их максимальным генетическим разнообразием и наибольшим числом компонент, выявляемых методом Admixture. Популяции Восточной Евразии кластируются в отдельный кластер, из которого выбиваются карелы и вепсы.

Наиболее сильным алгоритмом для визуальной дифференцировки и оценки кластеров, показал себя t-SNE. Это разделение хоть и является максимально точным из используемых алгоритмов, но интерпретировать биологический смысл компонент довольно трудно (рисунок 4, 5).

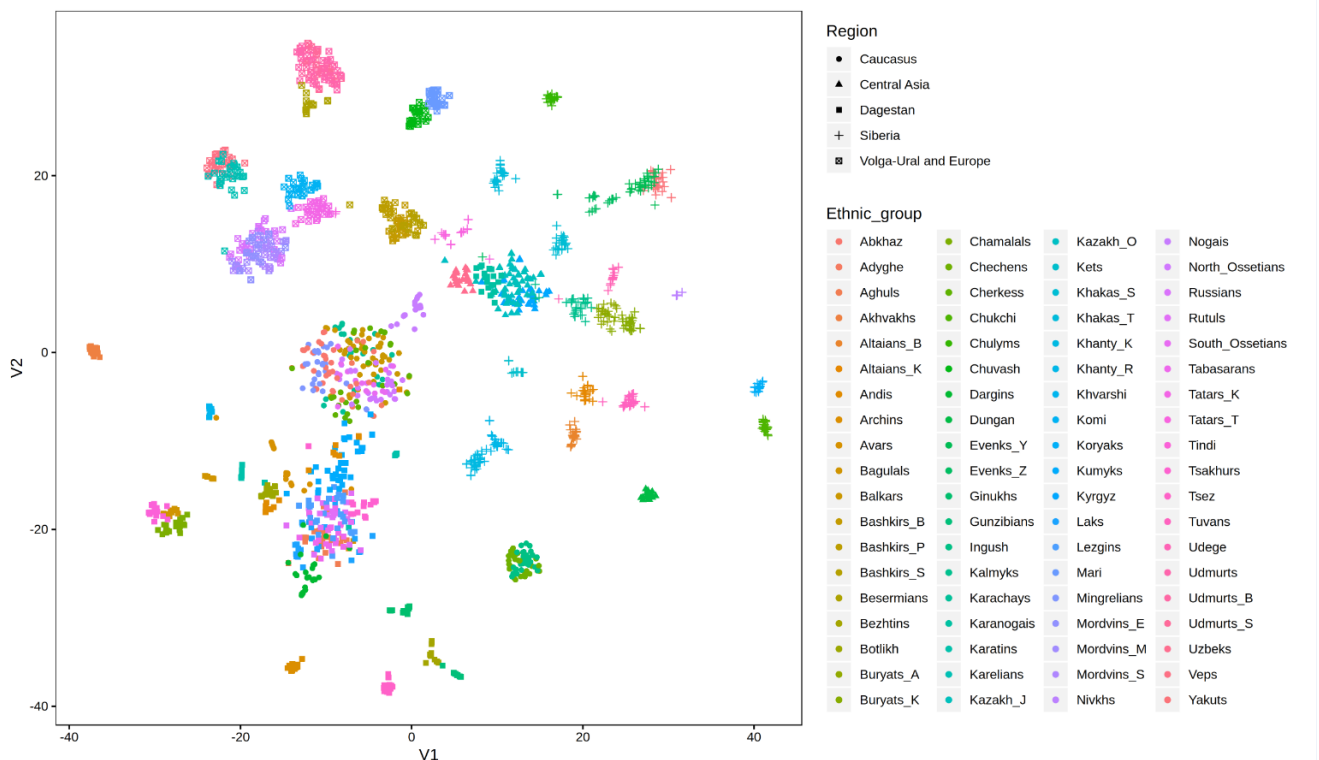


Рисунок 4 - Дифференциация геномов населения Северной Евразии по двум компонентам t-SNE.

С помощью этого подхода тот же массив данных четко разделяется на отдельные этноспецифичные группы образцов. Этнически и территориально близкие друг другу выборки располагаются довольно близко, но каждый образец входит в отдельный этнический кластер. Например, популяции коряков и чукчей располагаются близко друг к другу, но значительно удалены от остальных популяций Сибири. Относительно более гомогенно выглядит Северо-Кавказский регион с отделением ингушей и чеченцев. Дагестанский с сильной дисперсией по расположению отдельных популяций. При анализе t-SNE трех компонент и

построении трехмерной картины обособленно расположен кластер удмуртов (рисунок 5). Эти результаты хорошо согласуются с приведенными выше данными анализа частот SNP-маркеров методом PCA, но и дополняют его.

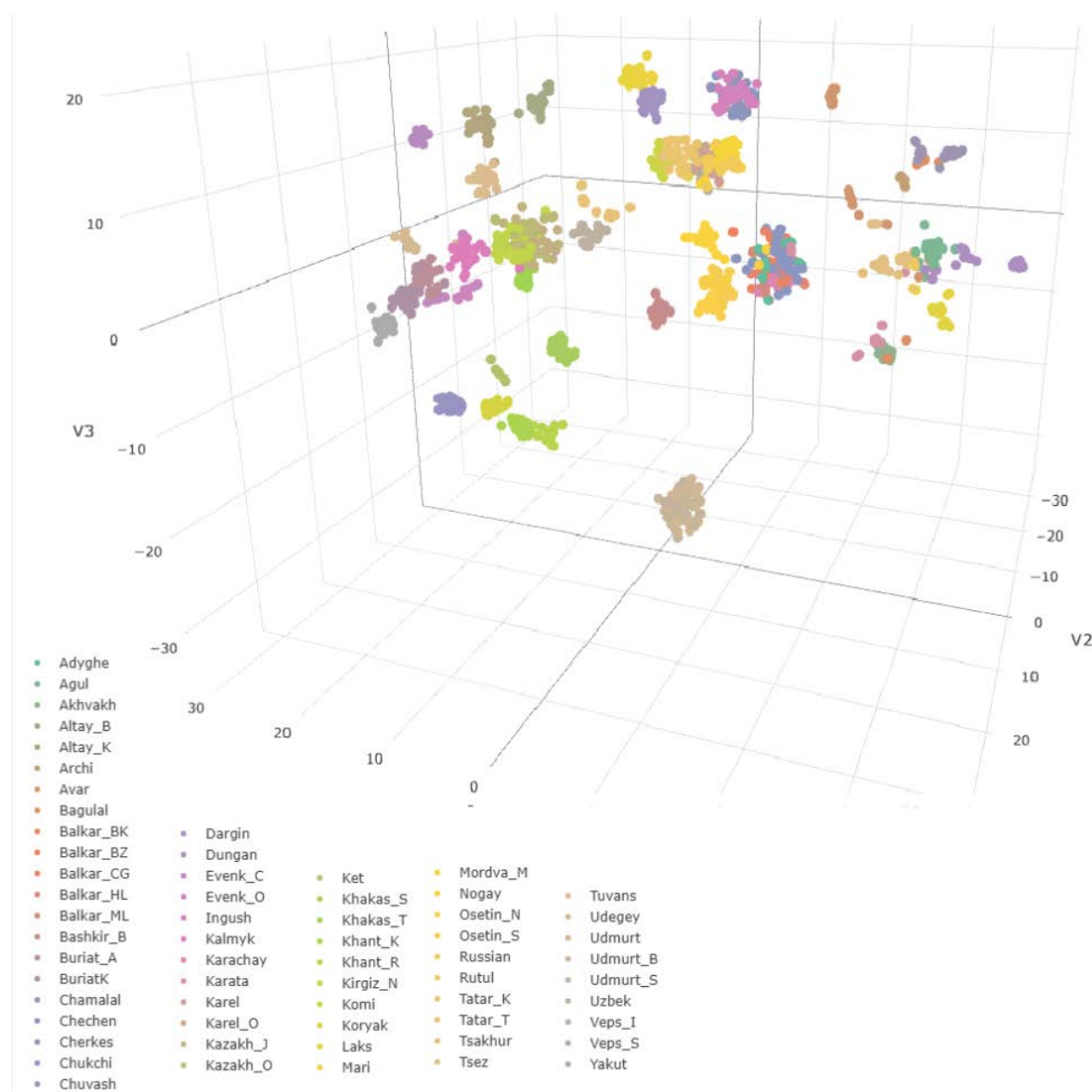


Рисунок 5 - Дифференциация геномов населения Северной Евразии по трем компонентам tSNE.

Генетические расстояния между популяциями Сибири и Восточной Европы значительно больше, по сравнению с популяциями Дагестана и Кавказа. Наиболее гетерогенными среди популяций Сибири являются томские татары, максимально близкие европейским башкирам и также имеющие минимальное значение F_{RON} . Близкое расположение киргизов с популяциями алтайцев и хакасов подтверждает Алтае-Саянские истоки их генофонда. Дальневосточные образцы разделены в

полном соответствии с географической локализацией на северную группу, чукчей и коряков и южную, включающую нивхов и удэгейцев.

3.2 Анализ предковых компонент с помощью алгоритма Admixture

По результатам анализа накопленного нами массива данных с помощью алгоритмов Admixture показано, что общая картина распределения компонент хорошо согласуется с географическим местоположением исследуемых популяций, привязкой к конкретному региону и антропологическим различиям. При относительно небольшом числе компонент $K=5-8$ большинство географических регионов характеризуются преобладанием отдельного компонента (рисунок 6). Начиная от 8 компонент и далее, некоторые популяции выделяются в отдельные этноспецифичные компоненты. Выделение таких отдельных этноспецифичных компонент связано с эффектами основателя и сдвигом частот аллелей для этих популяций. В качестве наиболее ярких примеров выступают выборки удмуртов, хантов, чукчей, коряков и некоторых популяций горной части Дагестана. Башкиры и томские татары занимают промежуточное положение между группами европейских и сибирских популяций, показывая наибольшее число генетических компонент в составе их генофонда. Отдельную группу формируют ханты, кеты и чулымцы с преобладанием собственного западно-сибирского компонента. Почти все остальные сибирские этносы объединяются в одну группу с популяциями Средней Азии. Отдельный специфичный компонент доминирует у северо-восточно-евразийских коряков и чукчей.

Помимо этого, также можно сосредоточить внимание на отдельных наиболее выраженных в популяциях генетических компонентах: монгольском, енисейском и тюркском.

Монгольский по происхождению компонент в генофонд различных этносов Сибири, Средней Азии и Волго-Уральского региона, присутствует у монголоязычных и тюркоязычных этносов. Начиная от $K=8$ в общем массиве

анализируемых популяций, отчетливо выделяется компонент, практически полностью доминирующий у монголоязычных бурятов (63%) и калмыков (54%) (рисунок 6). Наиболее четко он проявляется при $K=12$ (рисунок 7, таблица А.3). Именно этот генетический компонент, вероятно, маркирует вклад в генофонд населения Сибири, Средней Азии и Восточной Европы монгольских мигрантов. В коренных популяциях Сибири он занимает значительную долю у хакасов (41%), томских татар (24%) и тувинцев (12%). Практически все остальные исследованные популяции Сибири и Дальнего Востока – тюркоязычные якуты, шорцы, хакасы-сагайцы и чулымцы – демонстрируют практически полное отсутствие этого компонента. Не выявляется он и у эвенков, эвенов, хантов, кетов, ненцев, чулымцев, чукчей, коряков и нивхов. В генофонде среднеазиатских киргизов, казахов и узбеков он составляет более 40%. В Волго-Уральском регионе с наибольшей частотой выявлен у башкир (более 20%) и казанских татар (10%). Во всех остальных выборках восточноевропейских этносов он полностью отсутствует, за исключением калмыков. Общая картина распределения этого генетического компонента хорошо согласуется с антропологическими и этнографическими данными о влиянии монгольской экспансии на этногенез исследованных этносов.

В единую группу с преобладанием общего компонента входят практически все популяции Восточной Европы. Можно отметить выявление пермского компонента, выраженного у удмуртов 94,9%, бессермян 67%, коми 21,2%, башкир 13,7%, казанских татар 7%, и томских татар 6,1%. Его распределение по популяциям Волго-Уральского региона полностью согласуется с данными об антропологической близости и принадлежности к различным языковым группам внутри уральской языковой семьи. Удмурты и бесермяне формируют отдельную группу, что также объясняется довольно сильным эффектом основателя и изменением частот множества SNP.

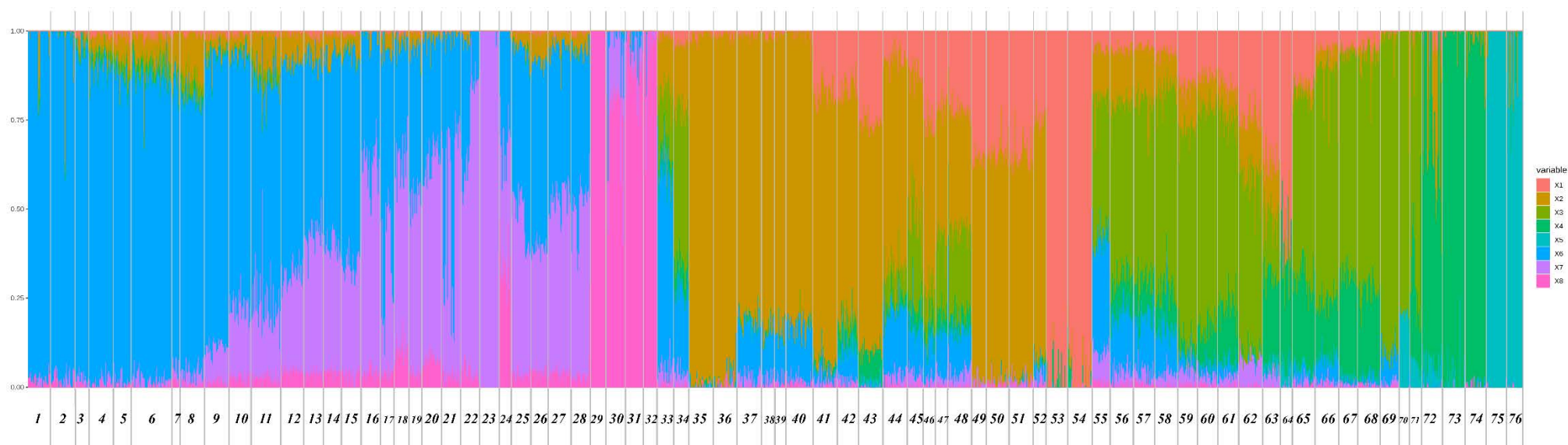


Рисунок 6 - Упорядоченная картина компонент Admixture при ранжировании популяций Евразии с запада на восток (K=8) Примечание: 1 – мегрелы, 2 – абхазы, 3 – осетинцы (Ю), 4 – осетинцы (С), 5 – карачаевцы, 6 – балкарцы, 7 – адыги, 8 – черкесы, 9 – ингуши, 10 – чеченцы, 11 – кумыки, 12 – лезгины, 13 – агулы, 14 – рутульцы, 15 – цахуры, 16 – арчинцы, 17 – андийцы, 18 – тиндинцы, 19 – ботлихцы, 20 – чамалалы, 21 – каратинцы, 22 – багулалы, 23 – ахвахцы, 24 – хваршины, 25 – лакцы, 26 – табасараны, 27 – даргинцы, 28 – аварцы, 29 – гинухцы, 30 – цезы, 31 – бежтинцы, 32 – гунзибцы, 33 – ногойцы, 34 – караногойцы, 35 – вепсы, 36 – карелы, 37 – мордва (М), 38 – мордва (Э), 39 – мордва (Ш), 40 – русские, 41 – коми, 42 – чувашы, 43 – марийцы, 44 – татары (К), 45 – татары (Т), 46 – башкиры (П), 47 – башкиры (С), 48 – башкиры (Б), 49 – удмурты (Ш), 50 – удмурты (Б), 51 – удмурты, 52 – бесермяне, 53 – ханты (Р), 54 – ханты (К), 55 – узбеки, 56 – казахи (М), 57 – казахи (С), 58 – киргизы, 59 – алтайцы (Б), 60 – алтайцы (К), 61 – хакасы (Ш), 62 – хакасы (Т), 63 – чулымцы, 64 – кеты, 65 – тувинцы, 66 – калмыки, 67 – буряты (А), 68 – буряты (К), 69 – дунгане, 70 – нивхи, 71 – удэгейцы, 72 – эвенки (З), 73 – эвенки (Я), 74 – якуты, 75 – чукчи, 76 – коряки.

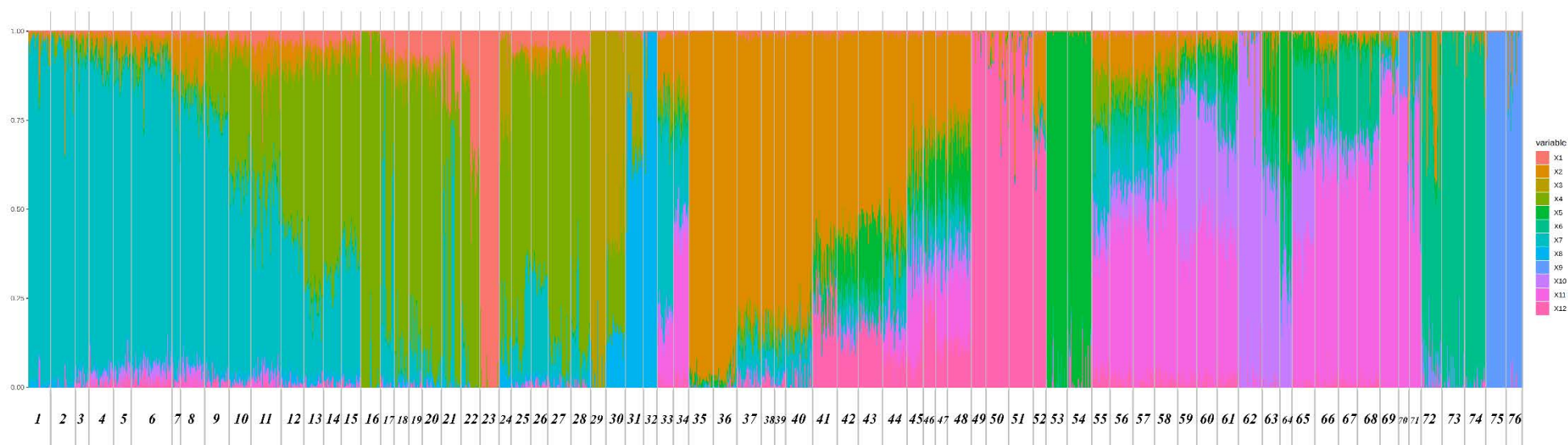


Рисунок 7 - Упорядоченная картина компонент Admixture при ранжировании популяций Евразии с запада на восток (K=12) Примечание: 1 – мегрелы, 2 – абхазы, 3 – осетинцы (Ю), 4 –осетинцы (С), 5 – карачаевцы, 6 – балкарцы, 7 – адыги, 8 – черкесы, 9 – ингуши, 10 – чеченцы, 11 – кумыки, 12 – лезгины, 13 – агулы, 14 – рутульцы, 15 – цахуры, 16 – арчинцы, 17 – андийцы, 18 – тиндинцы, 19 – ботлихцы, 20 – чамалалы, 21 – каратинцы, 22 – багулалы, 23 – ахвахцы, 24 – хваршины, 25 – лакцы, 26 – табасараны, 27 – даргинцы, 28 – аварцы, 29 – гинухцы, 30 – цезы, 31 – бежтинцы, 32 – гунзибцы, 33 – ногойцы, 34 – караногайцы, 35 – вепсы, 36 – карелы, 37 – мордва (М), 38 – мордва (Э), 39 – мордва (Ш), 40 – русские, 41 – коми, 42 – чувашы, 43 – марийцы, 44 – татары (К), 45 – татары (Т), 46 – башкиры (П), 47 – башкиры (С), 48 – башкиры (Б), 49 – удмурты (Ш), 50 – удмурты (Б), 51 – удмурты, 52 – бесермяне, 53 – ханты (Р), 54 – ханты (К), 55 – узбеки, 56 – казахи (М), 57 – казахи (С), 58 – киргизы, 59 – алтайцы (Б), 60 – алтайцы (К), 61 – хакасы (Ш), 62 – хакасы (Т), 63 – чулымцы, 64 – кеты, 65 – тувинцы, 66 – калмыки, 67 – буряты (А), 68 – буряты (К), 69 – дунгане, 70 – нивхи, 71 – удэгейцы, 72 – эвенки (З), 73 – эвенки (Я), 74 – якуты, 75 – чукчи, 76 – коряки.

При задании числа предковых компонент больше 7 в большинстве южносибирских популяций выявляется специфичный для них генетический компонент (Рисунок 8А), наиболее отчетливо проявляющийся на анализируемом массиве популяционных выборок при $K = 10$ (Рисунок 8Б), который можно интерпретировать как “енисейский” генетический пласт в генофонде современных популяций.

Его максимальное значение 92% показано у чулымцев. Существенную долю этот компонент занимает у кетов 64,7%. Вероятно, ассимиляция чулымцев проходила медленнее, чем кетов, значительно метисированных за последние несколько поколений. Однако, несмотря на значительную примесь европеоидного происхождения, кеты сохранили значительную часть собственного енисейского генетического компонента.

На уровне отдельных индивидов по результатам расчета доли различных генетических компонент видно, что протестированные образцы кетов отличаются друг от друга по доле енисейского компонента больше, чем чулымцы. Для каждого конкретного образца из выборки кетов доля енисейского компонента дополняется только компонентами, доминирующими в популяциях европейской части России, что полностью соответствует их метисации за последнее время в основном с пришлыми европейскими переселенцами. Таким образом, до метисации с пришлым населением генофонд кетов, судя по распределению компонент в исследованной выборке был достаточно гомогенен и в нем доминировал енисейский генетический компонент.

Около трети генофонда тувинцев 28,4% имеют в своем генофонде элементы енисейского компонента. По данным лингвистов выявляется связь между древним племенем чиков, населявших центральную территорию Тувы по обе стороны от Енисея до столкновений с тюркским каганатом и принимавшим активное участие в этногенезе тувинцев. При исследовании кето-енисейских языков в праенисейском языке обнаружены схожие слова с этнонимом «чики», чего не обнаружено в других языках того времени [121]. Можно предположить, что данные этносы проживали предположительно в одно время и на одной

территории, и складывание енисейского компонента происходило в Туве при их активном взаимодействии.

Существенную долю енисейский компонент занимает также в генофонде популяций Алтая-Саян – качинцев, северных и южных алтайцев, сагайцев, которые получили этот компонент не от ассимиляции с кетами, а от более ранних енисейских народов, которые проживали на территории Южной Сибири. Практически четверть генофонда хакасов качинцев 24,6% также представлены енисейским компонентом, что согласуется с антропологическими данными о взаимодействии енисейских народов с хакасами [116], а так же лингвистическим данным: енисейский языковой субстрат отмечен во многих современных тюркских языках Южной Сибири, таких как алтайский, хакасский, шорский, тувинский и в монгольском бурятском языке [122].

Также встречается енисейский компонент в генофонде алтайцев (В) 13,4% и алтайцев (К) 18,4%, помимо этого он отмечен в генофонде томских татар 9,2%. С меньшей частотой енисейский компонент представлен у других представителей алтайской языковой семьи: бурятов, калмыков, башкир, киргизов, казахов и др. Полученные результаты хорошо согласуются с данными антропологии и лингвистики о вкладе енисейского компонента в формирование различных народов Алтая-Саян и историческими ареалами енисейских языков. Данные о доле предковых компонент в других популяциях представлены в Таблице П.3

В выборках алтайцев, тувинцев, качинцев, и бурятов доля енисейского компонента на индивидуальном уровне варьирует значительно меньше, чему у кетов и чулымцев. Это свидетельствует об отсутствии метисации и достаточно большой древности енисейского субстрата в составе их генофондов.

В исследовании при изучении данных по менее плотным биочипам (около 160 тыс. SNP) наличие общего с кетами компонента было также выявлено у селькупов и энцев, а также у тюркоязычных популяций на Алтае (шорцы, хакасы, алтайцы и телеуты) [26].

Европеоидный компонент присутствует у некоторых народов средней Азии, у казахов, узбеков киргизов (K=12) (рисунок 8В). Среди сибирских популяций он

выявлен у томских татар 40,3%, и группы индивидуумов внутри популяции эвенков, что также подтверждается наличием длинных блоков IBD европейского происхождения и значительного снижения длины ROH и коэффициента геномного инбридинга, что говорит о метисации и неоднородности данной выборки.

При анализе данных аутосомных генотипов в качестве группы сравнения с наибольшей долей еще одного компонента (угорского) были использованы две различные выборки этнических хантов из Сургутского и Белоярского районов ХМАО. Угорский компонент доминирует у хантов и составляет 95%. Существенную долю этот компонент занимает также в генофонде популяций Волго-Уральского региона – башкир (до 25%), марийцев (до 20%), удмуртов и чувашей (до 15%). Отмечается его наличие у кетов (16%), башкир (11,4%), находящихся на территории расселения древних гуннов. В меньшей доле представлен у коми 8,4%, вепсов 3,7% и томских татар 3%. Значительна доля этого компонента также у кетов (при анализе до 8 компонент). С небольшой частотой он представлен практически во всех популяционных выборках Алтая-Саян – тувинцев, алтайцев, хакасов (от 5 до 10%).

Одним из наиболее широко распространенных на территории России и сопредельных стран является тюркский генетический компонент, маркирующий довольно позднюю волну миграций различных тюркских народов. Так, восточно-тюркский компонент можно выделить у хакасов сагайцев Таштыпского района 92,4%, хакасов качинцев 12,3%, алтайцев из с.Бешпельтир 32,9% и с.Кулада 17,9%, тувинцев 9%, а также представителей Средней Азии: казахов 4,6%, узбеков 5,2% и киргизов 5,6%.

Практика применения подхода, реализуемого в Admixture, показала его высокую эффективность при анализе значительных массивов генотипических данных их различных популяций, что позволило провести доскональный анализ генетической субструктуры в исследуемых группах популяций. В итоге, на индивидуальном и популяционном уровне реконструирован подробный состав генетических компонент генофонда различных народов на уровне популяций и

отдельных образцов. Эта информация позволяет более точно судить о сходстве или различии сравниваемых групп, структуре генетической дифференциации различных групп по составу предковых компонент и степени этих различий, а также о процессе формирования их генофонда.

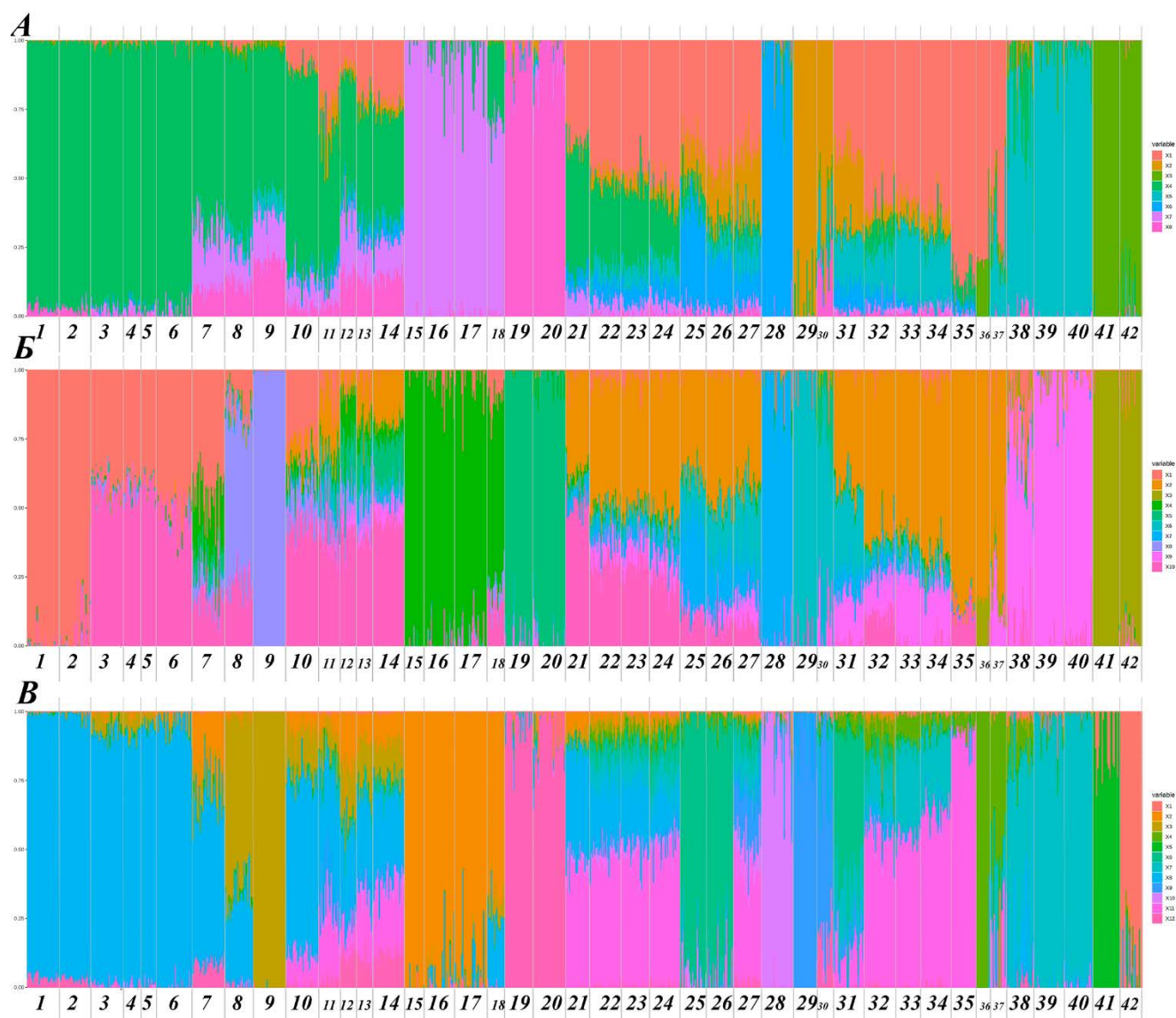


Рисунок 8 - Упорядоченная картина компонент Admixture при ранжировании популяций Северной Евразии с запада на восток для восьми (А), десяти (Б) и 12 (В) предковых компонент.

Примечание: 1 – вепсы, 2 – карелы, 3 – мордва (М), 4 – мордва (Э), 5 – мордва (Ш), 6 – русские, 7 – коми, 8 – чуваша, 9 – марийцы, 10 – татары казанские, 11 – татары томские, 12 – башкиры (П), 13 – башкиры (Ш), 14 – башкиры (Б), 15 – удмурты (Ш), 16 – удмурты (Б), 17 – удмурты, 18 – бесермяне, 19 – ханты (Р), 20 –

ханты (К), 21 – узбеки, 22 – казахи (М), 23 – казахи (С), 24 – киргизы, 25 – алтайцы (Б), 26 – алтайцы (К), 27 – хакасы (Ш), 28 – хакасы (Т), 29 – чулымцы, 30 – кеты, 31 – тувинцы, 32 – калмыки, 33 – буряты (А), 34 – буряты (К), 35 – дунгане, 36 – нивхи, 37 – удэгейцы, 38 – эвенки (З), 39 – эвенки (Я), 40 – якуты, 41 – чукчи, 42 – коряки.

3.3 Распределение регионов высокой гомозиготности в исследуемых популяциях

Проведен анализ распределения блоков гомозиготности (RON) для детального описания особенностей генофондов исследованных популяций, который позволяет проводить реконструкцию генетико-демографических процессов для всех изучаемых популяций.

Для идентификации регионов высокой гомозиготности используют различные статистические подходы. Один из наиболее популярных методов – метод «наблюдения», реализованный в широко используемой для генетического анализа популяций программе Plink. Другой метод, реализуемый в программном обеспечении GARLIC, использует подход на основе математических моделей, но в отличие от метода «наблюдения», этот анализ зависит от оценки частоты популяционных аллелей, что является как преимуществом, так и недостатком при сравнительном анализе с другими исследованиями. Для применения этого подхода мы остановились на программном обеспечении GARLIC, дополнительно позволяющем разделить длины регионов высокой гомозиготности на классы, что помогает более точно выполнить анализ RON с биологической точки зрения.

3.3.1 Оценка распределения регионов высокой гомозиготности с помощью программы «Plink»

Общая длина и количество RON для каждого индивидуального генома колеблется от 41,2 до 687,7 м.п.о. и от 61 до 394, соответственно. Медианное число RON на индивидуальный геном в исследованных образцах составило 135,4, граница первого квартиля 121,8, граница третьего квартиля 151,5. Медианная

длина RОН 128,4 м.п.о., граница первого квартиля 103,7 м.п.о. третьего 139,3 м.п.о.. Среднее значение совокупной длины RОН на индивидуума варьирует от 79,9 м.п.о. в выборке ногайцев, до 289,1 м.п.о. у гинухцев. Минимальное среднее популяционное значение числа RОН на человека наблюдается у узбеков (103,7), максимальное – среди коряков 312.

Доля RОН различной длины значительно колеблется в зависимости от географического региона и языковой группы. Выборки из Сибири, особенно носители языков чукотско-камчатских и нивхского языков, разительно отличаются от таковых из других географических регионов. Среднее значение общей длины RОН на человека в нашей выборке представителей чукотско-камчатских и нивхского языков (308,6 м.п.о.) сравнимо со значением в дагестанских популяциях, говорящих на языках дидойской подгруппы (305,03 м.п.о.).

По доле образцов с одним и более RОН длиной больше 5 м.п.о. по регионам выделяются три различные группы, которые явно связаны с эндогамией и изоляцией: большая часть сибирских (75.5%) и дагестанских образцов (78.1%) имеет длинные RОН, далее идут Северо-Кавказский, Волго-Уральский регионы и Восточная Европа 57.9% и 38%, соответственно. Для Средне-Азиатского региона наблюдается минимальная доля геномов, имеющих длинные RОН (15.6%). Это полностью соответствует полученным ранее данным о наибольшем генетическом разнообразии и многокомпонентности генофонда среднеазиатских популяций в сравнении с другими этносами Северной Евразии. При этом наблюдаются значительные различия между популяциями внутри каждого региона. Например, у казахов старшего жуза всего 7% образцов имеют RОН больше 5 м.п.о.. Для двух человек из эвенкийской (З) и казахской (С) популяций не было найдено RОН длиннее 1,5 м.п.о..

3.3.2 Оценка распределения регионов высокой гомозиготности с помощью программы «GARLIC»

Общая длина и количество ROH, идентифицируемые с помощью пакета «GARLIC», для каждого индивида колеблется от 124 до 788,6 м.п.о. и от 319 до 1237, соответственно. Медианное число ROH на индивидуальный геном в исследованных образцах составило 774.5, граница первого квартиля – 754,1, граница третьего квартиля – 798,5. Медианная длина ROH – 261,4 м.п.о., граница первого квартиля – 243,9 м.п.о. третьего – 282,8 м.п.о.. Среднее значение совокупной длины ROH на индивидуума варьируется от 167,3 м.п.о. в выборке башкир (С), до 480,2 м.п.о. у коряков. Минимальное среднее популяционное значение числа ROH на человека наблюдается у гинухцев (356,8), из-за большей длины ROH за счет сильного инбридинга, максимальное – среди коряков (1136).

Для количественной оценки закономерностей, наблюдаемых в распределениях суммарной длины ROH на уровне популяции, мы провели парные сравнения между суммарной длиной ROH класса А, класса В и класса С в отдельных геномах, используя `cor.test` в R для вычисления коэффициента корреляции Спирмена. Суммарные длины классов В и А имеют среднюю корреляцию ($r = 0,690$, $p < 2,2e-16$), однако оба класса имеют очень слабую корреляцию с суммарной длиной ROH класса С (С/А ($r = 0,127$, $p < 5,012e-08$), С/В ($r = 0,169$, $p < 2,741e-13$) (рисунок 9). Можно предполагать, что класс С и класс А и В ROH, вероятно, свидетельствуют о разных процессах: класс С о недавнем инбридинге, а класс А и В ROH о неравновесном сцеплении (англ. linkage disequilibrium (LD)) на уровне популяции в разных эволюционных масштабах [58].

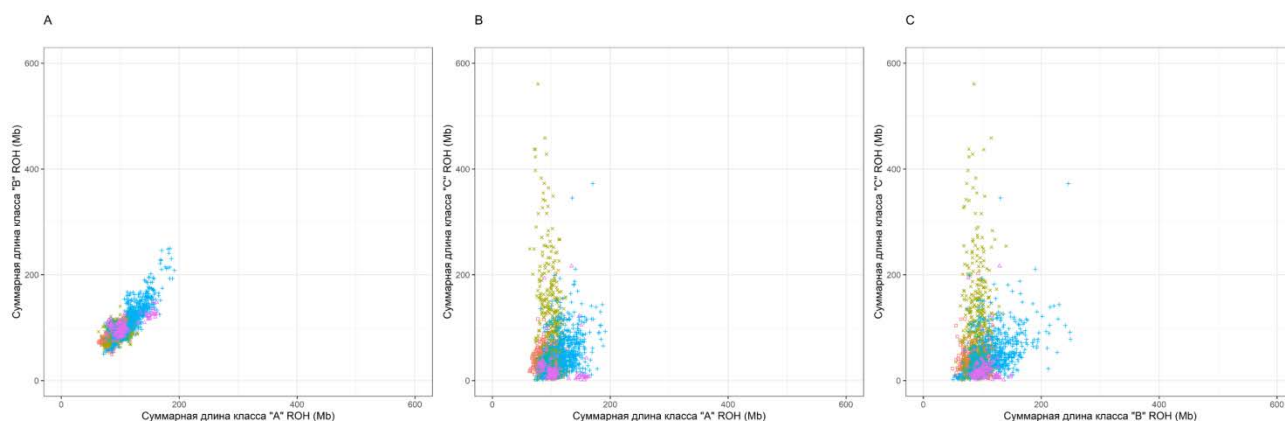


Рисунок 9 - Сравнение общей длины ROH для каждого индивидуума по разным классам. (А) Класс В против класса А; (В) Класс С против класса А; (С) Класс С против класса В.

Средние размеры ROH классов А и В одинаковы среди популяций одного географического региона, за исключением того, что популяции Дагестана и Сибири имеют большую изменчивость (Таблица 2). Максимальные значения наблюдаются у Дагестанских популяций (гинухцев, гунзибцев, ахвахцев, цезов).

Таблица 2 - Суммарная статистика для значений границ классов длин (кб) ROH по регионам

Статистика	Сибирь	Средняя Азия	Дагестан	Кавказ	Волго-уральский регион и Европа	Все
Классы А и В						
Минимальное	305,54	287,59	315,23	279,41	301,07	279,41
Максимальное	417,37	369,87	668,22	370,51	370,79	668,22
Среднее	366,96	333,15	390,07	321,02	338,75	358,55
SD	7,28	15,00	17,94	8,89	4,35	6,67
Классы В и С						
Минимальное	942,65	777,099	981,114	787,342	1010,64	777,10
Максимальное	1553,83	1402,82	2598,01	1200,04	1257,98	2598,01
Среднее	1260,15	1115,65	1380,02	999,93	1112,28	1214,99
SD	38,33	120,00	77,19	44,52	20,96	31,39

Наибольшие средние значения класса С также наблюдаются у популяций Дагестана и Сибири, далее у популяций Средней Азии, Волго-Уральского региона и Европы. Минимальные среднее значения обнаруживаются в кавказских

популяциях. У народов, имеющих смешанное происхождение (казахов (М, С), бурятов (А), калмыков, караногайцев и кумыков) средние размеры RОН аналогичны кавказским популяциям.

Для сибирских чукчей, коряков, нивхов для всех классов длин наблюдается значительное увеличение суммарной длины и количества RОН на человека по сравнению с другими популяциями, что говорит о продолжительном инбридинге, а также эффекте «бутылочного горлышка» (Рисунок 10).

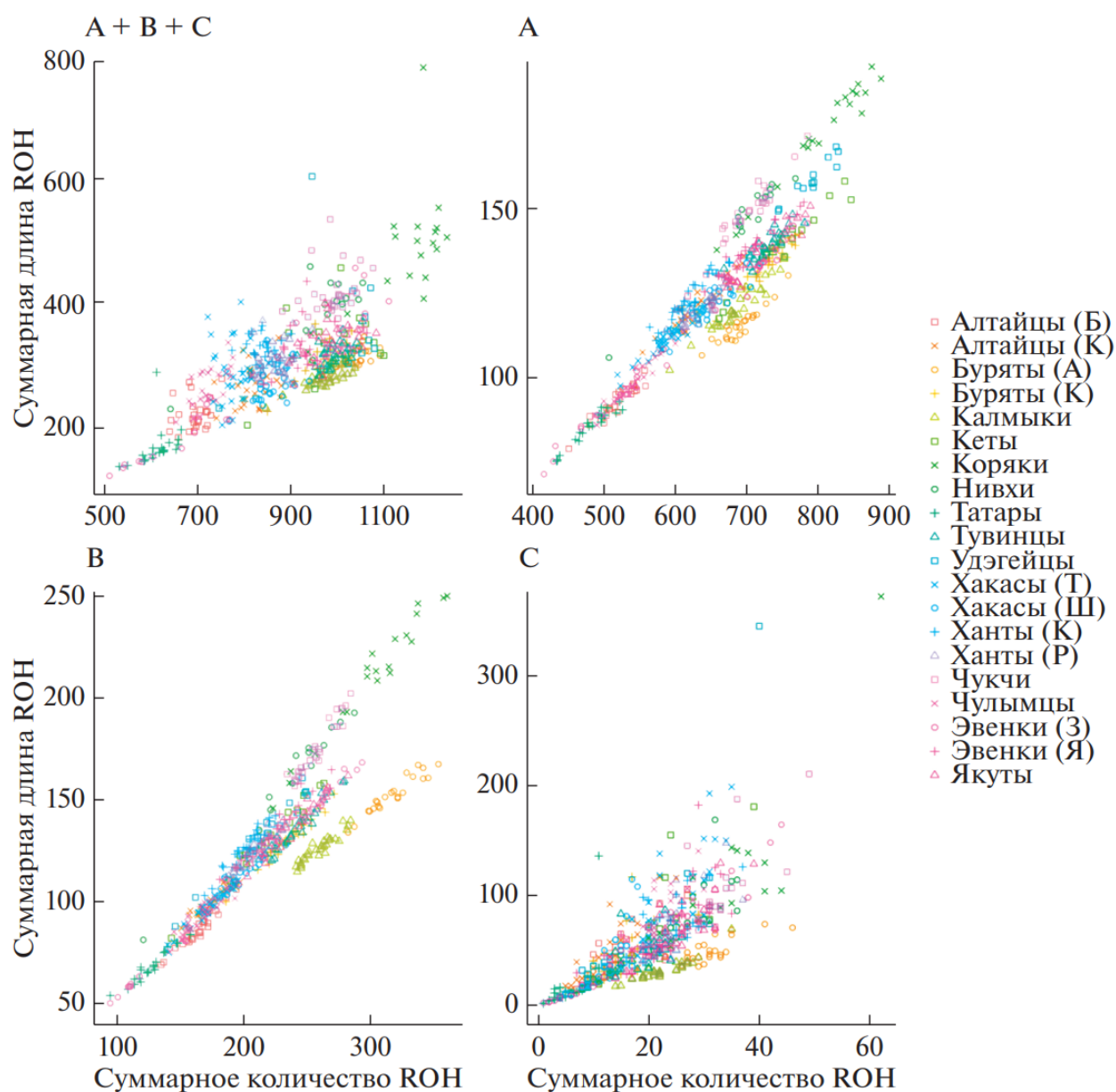


Рисунок 10 - Отношение суммарной длины RОН к суммарному количеству RОН в исследуемых популяциях Сибири.

Для чукчей, коряков, нивхов также наблюдается значительное увеличение суммарной длины среднего класса ROH на человека, как по сравнению с остальными популяциями, так и по сравнению с классом коротких ROH внутри популяции, что подтверждает продолжительность инбридинга и небольшую эффективную численность популяций. Далее, по коротким и средним длинам ROH выделяются кеты, якуты, эвенки (Я), удэгейцы, тувинцы и буряты (К). Для якутов и тувинцев, несмотря на большой эффективный размер популяции, большое количество и длина коротких и средних ROH, вероятно, связано с сильным эффектом основателя, который выявляется на маркерах мтДНК и Y-хромосомы [14,15]. В популяции эвенков (З) обнаружены образцы с признаками метисации, что может значительно занижать значения по длинам и количеству ROH в этой выборке. При анализе Y-хромосомных линий у них наблюдается типичные по происхождению европейские гаплогруппы, не характерные для коренных сибирских популяций. При анализе аутосомных блоков идентичных по происхождению, у эвенков также наблюдается значительное количество длинных блоков (4 м.п.о.) общих с европейскими популяциями, что говорит об относительно недавней метисации этой группы людей с пришлыми европейцами. Для бурятов (А) также характерно увеличение суммарной длины среднего класса ROH, но в большей степени не из-за длины фрагментов, а из-за их количества, что говорит об эффекте «бутылочного горлышка» в прошлом (Рисунок 10) [15,123]. У бурятов (А) и калмыков наблюдается увеличение количества ROH, но без увеличения суммарной длины наиболее ярко выражено для ROH средней длины, что говорит о небольшой численности предковой популяции с последующим резким ростом, без увеличения инбридинга. По средней суммарной длине и количеству для средних и коротких ROH максимальные значения наблюдаются у сибирских популяций (Таблица А.1).

У кетов, имеющих высокие значения по уровню инбридинга, сопоставимые со значениями хантов, чулымцев, якутов, удэгейцев и эвенков(Я). Схожая картина наблюдается и в распределении значений по количеству и длине ROH. По распределению коротких ROH кеты располагаются в группе с эвенками(Я),

якутами, тувинцами, бурятами(К). Для средних размеров RON добавляются удэгейцы и ханты. Отражая высокий уровень инбридинга внутри популяции, популяция кетов в распределении длинных RON попадают в группу с эвенками(Я), якутами, хантами, чулымцами (Рисунок 10).

Распределение RON связано с размером популяции и их эффективной численностью, при этом относительно небольшие популяции Сибири имеют RON большего количества и размера, чем большие по численности популяции этих регионов. Аналогичная картина образуется на территории Дагестана, где для популяций имеющих небольшую численность (гинухцы, гунзибцы, бежтинцы, ахвахцы и цезы), для всех классов RON наблюдается распределение соотношения суммарной длины и количества характерное для популяций подверженных сильному инбридингу (Рисунок 11). Инбридинг добавляет небольшое количество очень длинных RON для тех, кто является потомком кровнородственных браков, увеличивая дисперсию суммы RON , видимую как сдвиг влево на графике длин RON по сравнению с количеством RON .

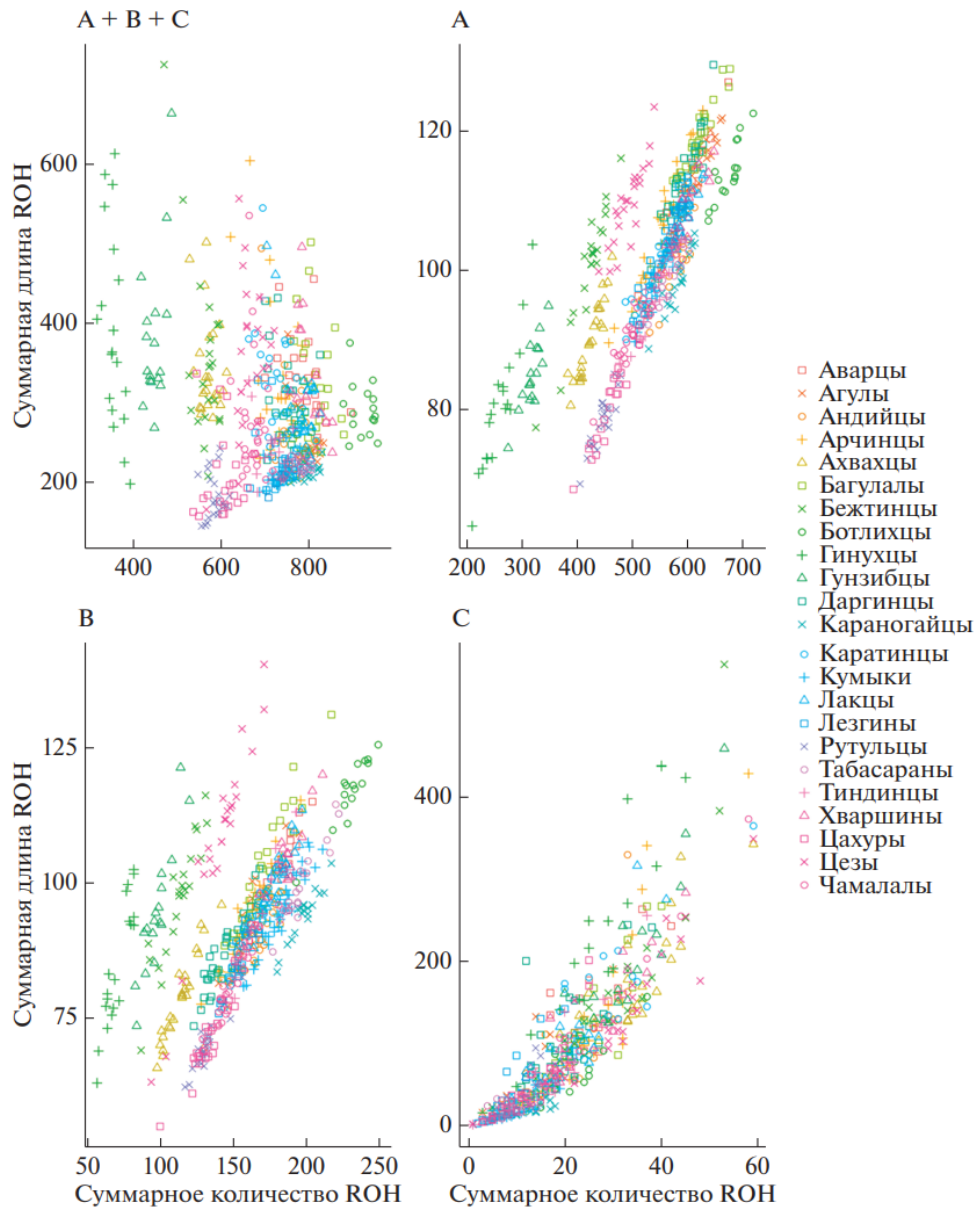


Рисунок 11 - Отношение суммарной длины ROH к суммарному количеству ROH в исследуемых популяциях Дагестана.

В популяциях, смешанных по происхождению и характеризующихся высоким уровнем генетического разнообразия, объединяются различные гаплотипы, что обычно значительно снижает количество и длину ROH. Это обнаружено у томских и казанских татар, узбеков и башкир. Для Волго-Уральского региона и Европы наблюдается разделение на две группы по количеству ROH, где основной вклад вносят короткие и средние ROH. В первую группу входят марийцы, коми, вепсы, чувашы, карелы, удмурты и башкиры (С),

во вторую: бесермяне, мордва (М, Ш, Э), удмурты (Ш, Б), татары (К), башкиры (Б, П) и русские (Рисунок 12).

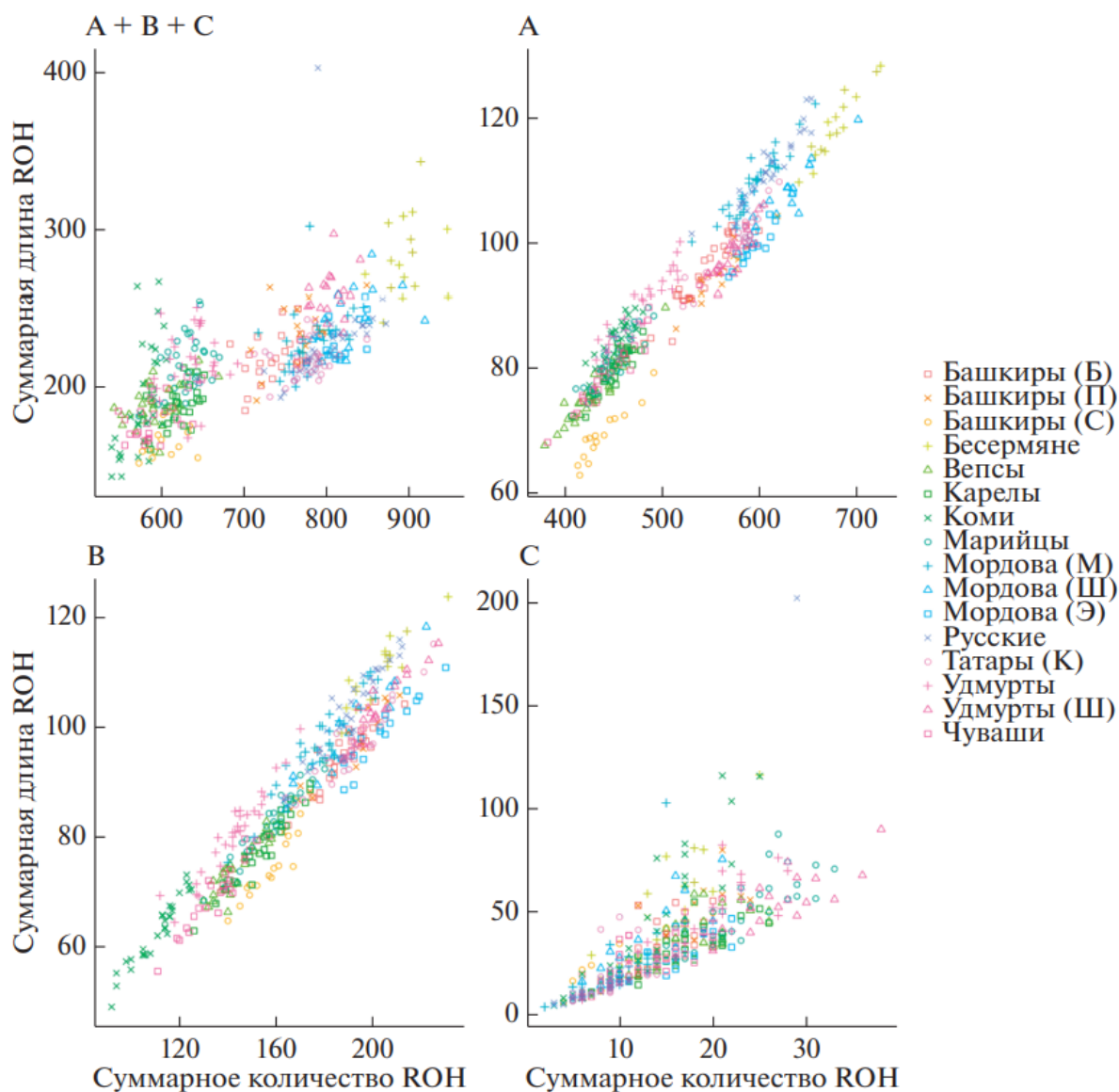


Рисунок 12 - Отношение суммарной длины ROH к суммарному количеству ROH в исследуемых популяциях Волго-Уральского региона и Европы.

Эффект бутылочного горлышка в демографической истории конкретного этноса увеличивает количество ROH, которые обычно все еще относительно короткие, например у популяций калмыков и бурятов (А). У некоторых популяций, имеющих как эффект бутылочного горлышка, так и распространенность близкородственных браков, наблюдаются большее количество и длина ROH с увеличением дисперсии суммы ROH (коряки, чукчи и нивхи).

Наибольшими отличиями среди популяций Средней Азии характеризуется популяция дунган: в ней наблюдается высокие значения по количеству и суммарной длине для средних и коротких ROH и минимальные значения для длинных (рисунок 13).

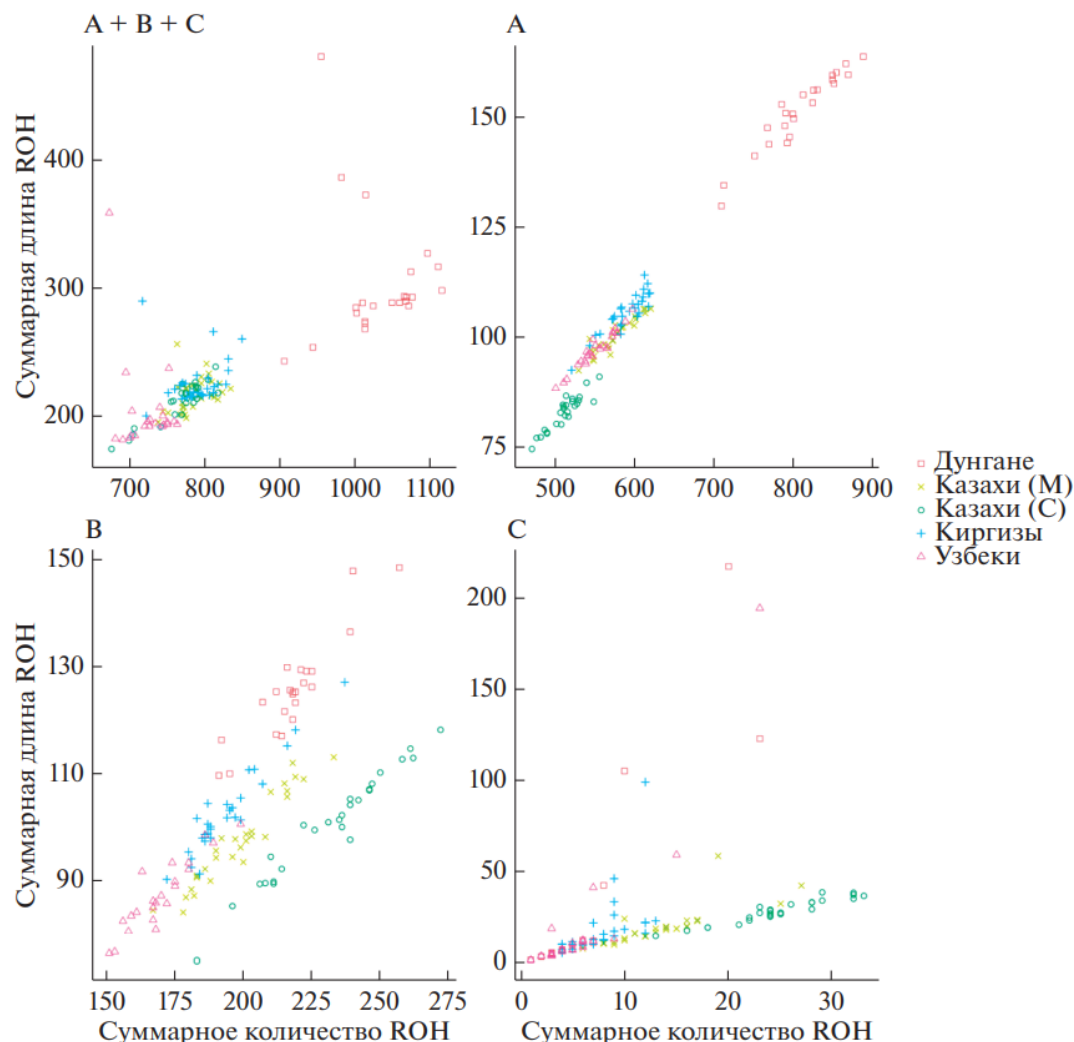


Рисунок 13 - Отношение суммарной длины ROH к суммарному количеству ROH в исследуемых популяциях Средней Азии.

Суммарная доля разных по длине и количеству ROH на уровне индивидуальных геномов сильно варьирует по популяциям и регионам (рисунок 14). Наши результаты хорошо согласуются с рядом недавних исследований, проведенных с использованием высокоплотных чипов [124,125]. Уровень гомозиготности по всему геному показывает самый высокий уровень инбридинга в сибирских и дагестанских группах, говорящих на дидойских, чукотко-камчатских и нивхском языках, по сравнению с остальными северо-евразийскими

популяционными выборками. Сибирские популяции наравне с дагестанскими популяциями обладают самыми длинными геномными участками гомозиготности для всех категорий длины RОН и более высокую долю людей с длинными областями аутозиготности.

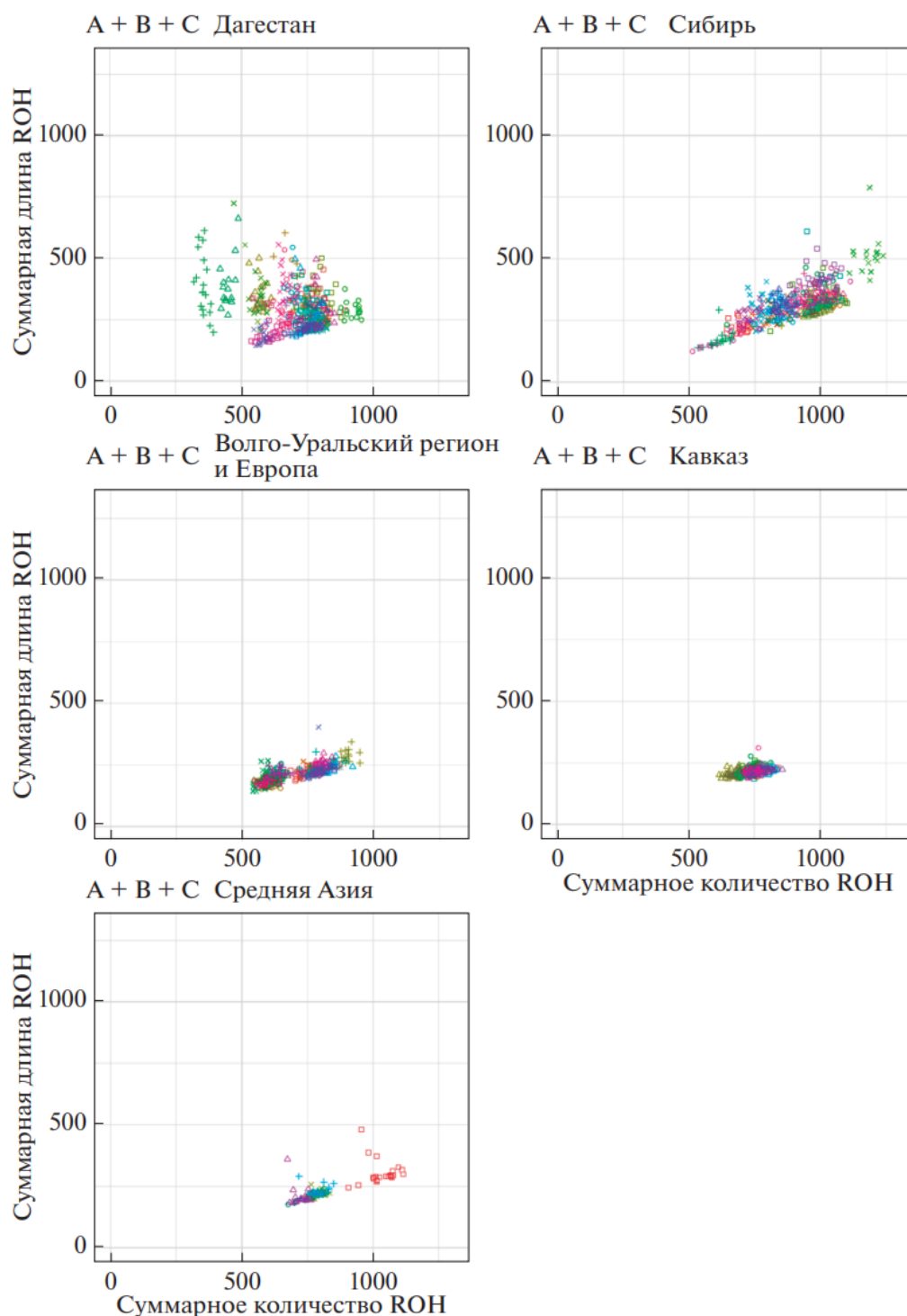


Рисунок 14 - Отношение суммарной длины RОН к суммарному количеству RОН в исследуемых популяциях для каждого географического региона.

Для работ по популяционной генетике сибирских этносов с использованием нейтральных маркерных систем Y-хромосомной, митохондриальной ДНК показано, что во многих популяциях наблюдается эффект основателя по отдельным гаплогруппам с низким генетическим разнообразием, которое свидетельствует о событиях экспансии численности в относительно недавнее время [14,123,126]. Интересно, что полученные нами оценки существенно не отличаются от тех, которые были получены в предыдущем исследовании нашей группы [71] с использованием гораздо меньшего числа маркеров на меньших выборках сибирских популяций.

Для популяций Сибири характерно большее значение суммарной длины класса A и B на человека, что соответствует данным об увеличении длин классов A и B с удалением популяции от Африки. Эта тенденция похожа на наблюдаемое сокращение разнообразия гаплотипов с увеличением расстояния от Африки. Суммарная длина класса C R_{OH} более изменчива и значительно больше в большинстве популяций Сибири и Дагестана по сравнению с популяциями Кавказа, Средней Азии, Европы и Волго-Уральского региона. Это также свидетельствует о том, что большая доля популяций из Сибири и Дагестана характеризуются высокой степенью кровного родства. Это подтверждается высокими значениями коэффициента геномного инбридинга (F_{ROH}).

Результаты проведенного анализа показывают, что генофонд населения Сибири демонстрирует значительные различия в распределении суммы длин и количества по всем классам R_{OH} как между исследуемыми географическими регионами, этносами, так и внутри локальных популяций. Особенно это характерно для территориально удаленных популяций чукчей и коряков, сопоставимыми по значениям с культурно и географически изолированными горными популяциями Дагестана, показывающими максимальные значения инбридинга, сохранившегося на высоком уровне до настоящего времени. Большее количество и длина коротких и средних R_{OH} у большинства сибирских популяций указывает на более древний и более продолжительный инбридинг. Получено подтверждение значительной генетической гетерогенности сибирских

татар, обладающими одними из самых низких значений по количеству и длине ROH. Полученные результаты позволяют существенно уточнить имеющиеся данные об особенностях структуры генофонда всех исследованных этносов. Данные о аутозиготных блоках сцепления в конкретных популяциях также могут быть использованы при изучении их компонентного состава предрасположенности к различным наследственным заболеваниям.

3.3.3 Оценка уровня инбридинга F_{ROH} в исследуемых популяциях

Коэффициенты инбридинга F_{ROH} для всех индивидов из 76 популяционных выборок оценивались отдельно и группировались по географическим регионам и языковым группам для трех классов ROH (длиной больше пороговых значений 0,5, 1,5 и 5 м.п.о.). Коэффициенты F_{ROH} для трех диапазонов длин ROH, среднее суммарное количество и средняя суммарная длина гомозиготных участков трех классов в исследованных популяциях представлены в таблице П.1. Примечательно, что $ROH > 1.5$ м.п.о. не было выявлено лишь у четырех индивидов (кумыка, лезгина, эвенка и узбека).

Среди коренного населения Сибири максимальные значения $F_{ROH > 1,5}$ обнаружены у дальневосточных коряков, чукчей и нивхов (0,0422, 0,0409, 0,0363), что вполне закономерно, в связи с их относительной малочисленностью и территориальной изолированностью. Среди Южно-Сибирских популяций включающих алтайцев, томских татар, тувинцев и хакасов, максимальное значение выявлено и для выборки хакасов-сагайцев из предгорного Таштыпского района (0,0318), вдвое превосходящее хакасов-качинцев равнинного Ширинского района. Минимальное значение характерно для томских татар (0,0071). Для Западно-Сибирских популяций максимальные значения имеет популяция чулымцев (0,0292), близкие по значению ханты(К) (0,0280), ханты(Р) (0,0266) и кеты (0,0259).

Наблюдается высокая корреляция для $F_{ROH > 1,5}$ с суммой средних длин сегментов $IBD > 1,5$ сМ между парами индивидов в сибирских популяциях ($r =$

0,9246, $p < 5,612e-09$) (Рисунок 15). Для вычисления коэффициента корреляции Спирмена использовался `cor.test` в программе R.

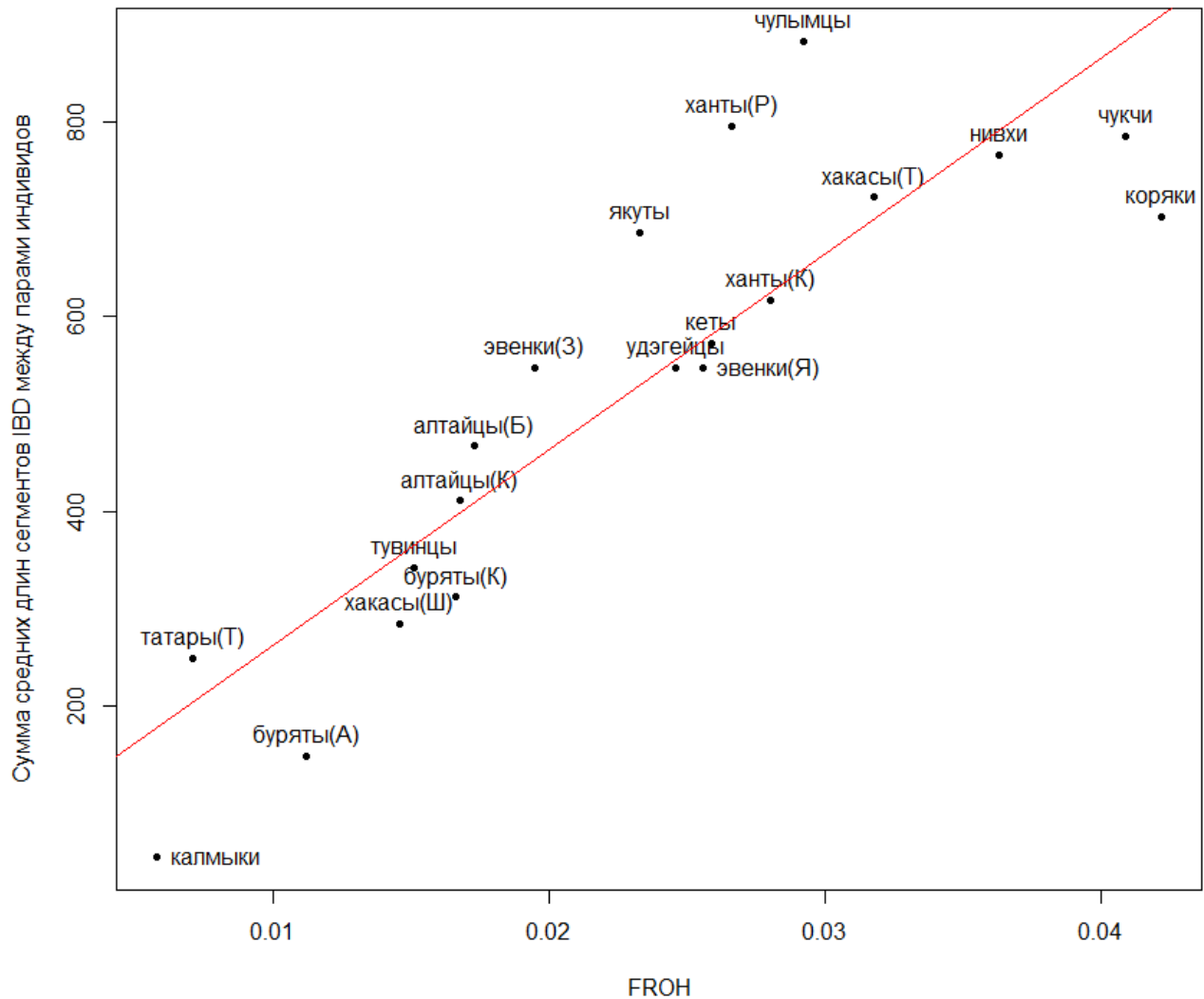


Рисунок 15 - Отношение суммы средних длин сегментов IBD ($IBD > 1,5$ cM) между парами индивидов в сибирских популяциях к коэффициенту геномного инбридинга ($F_{ROH > 1,5}$).

Несмотря на то, что $F_{ROH > 5}$ и $F_{ROH > 1,5}$ в популяциях Дагестана значительно превышает значения F_{ROH} в популяциях Сибири, по более коротким участкам гомозиготности ($F_{ROH > 0,5}$) наблюдаются схожие значения (Таблица А.1). Это может свидетельствовать о значительных эффектах инбридинга для этих

популяций Сибири в прошлом с постепенным уменьшением доли кровнородственных браков к настоящему времени.

Максимальные значения по $F_{ROH>1.5} = (0,0376)$ и $F_{ROH>5} (0,0239)$ наблюдаются в популяциях Дагестана (Таблица А.1). Наибольшим коэффициентом геномного инбридинга для длин $ROH>1,5$ м.п.о. обладают популяции гинухцев и гунзибцев ($F_{ROH} = 0,1039$; $F_{ROH} = 0,0887$). Этот показатель для гинухцев является одним из самых высоких в мире и схож лишь с уровнем инбридинга в изолированной популяции индейцев каритиана из бразильской Амазонии общей с численностью в 320 человек ($F_{ROH>1.5} = 0,1067$)[127]. Также по уровню гомозиготизации генома значительно выделяется и ряд других малочисленных народов Дагестана - бежтинцы, ахвахцы, и цезы (F_{ROH} 0,0689, 0,0678 и 0,0571 соответственно). Эти результаты хорошо соотносятся с предыдущим и данными по популяциям Дагестана [128], где для народов нахско-дагестанской языковой семьи, были выявлены более высокие коэффициенты инбридинга по сравнению с популяциями из других географических регионов и не нахско-дагестанскими популяциями Дагестана.

Внутри нахско-дагестанской языковой семьи по F_{ROH} этносы разделяются на три группы, в полном соответствии с их языковой классификацией. Максимальный средний уровень F_{ROH} выявлен для популяций дидойской (цезской) группы (0,0727), проживающих в высокогорных регионах в течение сотен поколений в относительной изоляции и имеющих также значительно более низкий уровень разнообразия по гаплогруппам Y-хромосомы, по сравнению с другими северокавказскими популяциями [129]. Для андийской группы показано гораздо меньшее значение (0,0378), находящееся на одном уровне с выборками чукчей, коряков и нивхов (0,0360). Внутри лезгинской группы меньшие значения (0,0193) коэффициентв геномного инбридинга характерны для популяций, занимающих низменные территории и предгорья южной части Дагестана и отличающейся гораздо большей, по сравнению с дидойской и андийской группами численностью представляющих ее этносов. Наименьший коэффициент инбридинга выявлен для предствителей тюрской языковой группы (кароногайцы

(0,0039), кумыки (0,0036)), занимающим равнинные территории северной части Дагестана. Эти результаты полностью согласуются с данными об очень высокой частоте инбредных браков среди малочисленных и более изолированных популяциях высокогорных районов, по сравнению с равнинными [130].

Популяции Волго-Уральского региона занимают промежуточное положение среди других регионов по коэффициенту инбридинга для всех трех групп гомозиготных регионов. Это можно объяснить как его территориальным положением на месте пересечения путей различных миграционных потоков, так и с давней традицией заключения браков лишь с родственниками не ближе пятого-шестого колена, что характерно для башкир, казанских татар, чувашей, коми, марийцев и мордвы. Внутри Волго-Уральского региона для популяций с низкой эффективной численностью и проживающих на периферии наблюдаются более высокие значения уровня гомозиготности генома. Максимальное значение обнаружено у бесермян (0,0199), обособленных территориально и имеющих численность около 2 тыс. человек. Минимальное значение у казанских татар и русских (0,0061). Наиболее близкие к этим значениям среди сибирских популяций калмыки (0,0058) и томские татары (0,0071).

Популяции Кавказа (за исключением Дагестана) и Средней Азии обладают минимальными значениями коэффициентов инбридинга, оцененных по R_{OH} , - от 0,0030 до 0,0116 для $F_{R_{OH}>1,5}$. Среди популяций Средней Азии максимальные значения наблюдаются у дунган (0,0116). Это, вероятно, объясняется недавней миграцией дунган из Китая с сохранением их языкового, культурного и демографического единства на фоне практически полного отсутствия межнациональных браков. Ранее было показано, что дунгане Узбекистана генетически намного ближе к популяциям Китая и северо-восточной Индии, чем к соседним с ними узбекским популяциям, и не имеют с ними недавнего обмена генами [131].

3.4 «Горячие точки» ROH

При анализе распределения частот ROH для всех популяций было найдено 56 регионов, содержащих «горячие точки» (таблица А.2). Среди 56 регионов 19 включают гены, совпадающие с данными из предыдущих исследований «горячих точек» ROH: 20 регионов для Дагестанских популяций [128], 3 для европейских [124] и 1 регион для мировых популяций [8]. Так, во всех трех исследованиях были пересечения с регионом на chr6:27904181-28413491, содержащим гены, кодирующие белки цинковых пальцев. Для трех генов (*ZKSCAN4*, *PGBD1*, *NKAPL*) из данного региона были отмечены сигналы направленного отбора по тесту nSL для сибирских популяций [10]. Для кросс-популяционного теста (XP-EHH) совпадение было только по одному гену *ABCA1*, кодирующего белок АТФ-связывающей кассеты (ATP-binding cassette). Этот переносчик является основным регулятором клеточного холестерина и гомеостаза фосфолипидов.

Для популяций Сибири был определен 61 регион горячих точек ROH (таблица 3), включающих 91 ген. Из них 4 гена (*ERC1*, *CCDC138*, *CDKAL1*, *LIMS1*) имеют сигналы естественного отбора по тесту nSL, 8 (*CCDC146*, *ERC1*, *LIMS1*, *LINC00261*, *CDKAL1*, *SLC2A9*, *CCDC138*, *ZBTB38*) по тесту iHS и 7 (*DOCK7*, *ANGPTL3*, *LEPR*, *GCC2*, *RHBDD1*, *CD55*, *RANBP2*) по кросс-популяционному тесту xp-ehh.

Также было выявлено, что некоторые из этих генов несут сигналы направленного отбора в коренных популяциях Сибири [12] 8 генов по тесту PBS (*SLAIN2*, *SELP*, *F5*, *CCDC146*, *FGB*, *PLRG1*, *HLA-DPB1*, *HLA-DPA1*) и 3 гена по тесту iHS (*GCC2*, *LIMS1*, *SNTG1*).

Таблица 3 - Горячие точки ROH, определённые в популяциях коренного населения Сибири

Хромосома	Гены	p1	p2	кб	Минимум	Медиана	Среднее	Максимум
2	-	623212	651105	27,894	0,455	0,675	0,639	0,675
8	<i>LPL</i>	19816888	19877732	60,845	0,426	0,618	0,598	0,669
15	<i>ADAMTSL3</i>	84566063	84587958	21,896	0,421	0,591	0,570	0,595
1	<i>F5, SELP</i>	169531135	169566827	35,693	0,415	0,589	0,542	0,589
6	<i>CDKAL1</i>	20646912	20683164	36,253	0,413	0,556	0,532	0,564
18	-	57736270	57909799	173,53	0,421	0,550	0,526	0,614
6	<i>MIR588</i>	126758540	127036174	277,635	0,415	0,543	0,503	0,566
1	<i>TMEM50A, RHCE, MACO1</i>	25679393	25859024	179,632	0,421	0,541	0,530	0,558
20	<i>LINC00261, FOXA2, LNCNEF</i>	22518523	22575792	57,27	0,413	0,532	0,534	0,581
4	<i>SLAIN2, SLC10A4, ZAR1, FRYL</i>	48408599	48591150	182,552	0,415	0,518	0,506	0,558
7	<i>CCDC146</i>	76910774	76927488	16,715	0,413	0,518	0,512	0,572
20	<i>UQCC1, GDF5OS, GDF5</i>	33896262	34025983	129,722	0,413	0,518	0,508	0,539
20	<i>CST9, CST3</i>	23578206	23619075	40,87	0,442	0,508	0,502	0,516
6	<i>HLA-DPA1, HLA-DPB1</i>	33032975	33056962	23,988	0,413	0,507	0,522	0,614
1	<i>LEPR</i>	65938332	66159877	221,546	0,413	0,505	0,489	0,549
4	<i>CWH43</i>	48978241	49063872	85,632	0,430	0,505	0,494	0,507
3	-	176035399	176078460	43,062	0,413	0,501	0,490	0,572
6	<i>LIN28B</i>	105360586	105462876	102,291	0,415	0,488	0,479	0,491
6	<i>MIR3135B, HLA-DQB2</i>	32675645	32761690	86,046	0,413	0,486	0,479	0,539
2	-	17250972	17340388	89,417	0,413	0,480	0,463	0,484
9	<i>CDKN2B-AS1</i>	22023805	22054690	30,886	0,415	0,474	0,469	0,476
22	<i>LINC02558,SLC5A1</i>	32381973	32445218	63,246	0,415	0,472	0,457	0,476
10	<i>MARCHF8</i>	45967768	46009818	42,051	0,413	0,470	0,459	0,484
4	-	156477400	156508196	30,797	0,413	0,468	0,465	0,474
5	-	4006099	4068808	62,71	0,421	0,465	0,459	0,478

Продолжение таблицы 3

Хромосома	Гены	p1	p2	кб	Минимум	Медиана	Среднее	Максимум
15	<i>LARP6, LRRC49, THAP10</i>	71107081	71184902	77,822	0,423	0,464	0,486	0,541
4	-	82144760	82195346	50,587	0,413	0,463	0,460	0,501
16	<i>PSKH1, CTRL, LCAT, SLC12A4, DPEP2;DPEP3 ;SLC12A4, DPEP2NB, DDX28, DUS2, NFATC3</i>	67928042	68190239	262,198	0,419	0,461	0,472	0,520
1	<i>DOCK7, ANGPTL3</i>	62960289	63130431	170,143	0,419	0,459	0,452	0,468
10	-	94899090	95055548	156,459	0,423	0,455	0,454	0,472
10	-	44745925	44773984	28,06	0,415	0,454	0,449	0,455
2	-	217863961	217917167	53,207	0,426	0,453	0,455	0,472
3	-	86868274	86927165	58,892	0,417	0,453	0,459	0,503
3	-	147684551	147942624	258,074	0,413	0,451	0,448	0,476
1	<i>LRRC7</i>	70305259	70380794	75,536	0,423	0,449	0,446	0,459
3	<i>ZBTB38</i>	141096736	141140968	44,233	0,415	0,449	0,448	0,459
3	<i>LINC01322</i>	165004966	165041791	36,826	0,415	0,444	0,441	0,484
17	<i>PSMD3, CSF3, MED24</i>	38137364	38197914	60,551	0,413	0,443	0,441	0,449
4	<i>PLRG1, FGB</i>	155452868	155488821	35,954	0,417	0,440	0,439	0,444
1	<i>CCDC17, GPBP1L1, TMEM69, IPP, MAST2</i>	46085050	46498375	413,326	0,413	0,438	0,439	0,465
3	-	164877855	164898170	20,316	0,421	0,438	0,439	0,453
2	<i>GCC2, LIMS1, CCDC138, RANBP2</i>	109031721	109412959	381,239	0,415	0,436	0,435	0,451

Продолжение таблицы 3

Хромосома	Гены	p1	p2	кб	Минимум	Медиана	Среднее	Максимум
13	<i>LINC00402</i>	74746219	74816423	70,205	0,413	0,435	0,432	0,451
6	<i>HCG22</i>	31024601	31058398	33,798	0,417	0,434	0,438	0,482
6	<i>RSPO3</i>	127476985	127495245	18,261	0,413	0,432	0,428	0,432
3	-	175673276	175806373	133,098	0,413	0,430	0,434	0,465
9	<i>ABCA1</i>	107646224	107656122	9,899	0,419	0,430	0,428	0,432
20	<i>TRPC4AP, EDEM2, MMP24-AS1- EDEM2, PROCR</i>	33623936	33788983	165,048	0,413	0,430	0,431	0,457
1	<i>CD55</i>	207441975	207517696	75,722	0,413	0,428	0,428	0,432
5	<i>UIMC1, ZNF346</i>	176335231	176454075	118,845	0,413	0,428	0,427	0,434
7	-	106324612	106364227	39,616	0,413	0,428	0,426	0,432
4	<i>SLC2A9</i>	10024413	10056082	31,67	0,413	0,426	0,425	0,438
4	<i>LINC02232</i>	65798434	65893504	95,071	0,413	0,426	0,425	0,430
1	<i>LYPLAL1-AS1</i>	219707578	219722104	14,527	0,413	0,417	0,417	0,419
2	-	21288321	21301641	13,321	0,413	0,417	0,417	0,421
9	<i>UBQLN1</i>	86274333	86309931	35,599	0,415	0,417	0,418	0,419
12	<i>ERC1</i>	1377408	1486249	108,842	0,413	0,417	0,418	0,426
3	-	39618940	39678712	59,773	0,413	0,415	0,415	0,417
6	<i>CEP85L</i>	118886188	118973137	86,95	0,413	0,415	0,415	0,419
2	<i>IRS1, RHBDD1</i>	227647783	227751141	103,359	0,413	0,413	0,421	0,440
8	<i>SNTG1</i>	51249185	51296270	47,086	0,413	0,413	0,414	0,415

Примечание: регионы отсортированы по уменьшению медианного значения

3.5 Блоки идентичные по происхождению

Чтобы лучше изучить степень родства сибирских популяций и выяснить, в какой степени наблюдаемая структура может быть объяснена недавними миграциями, мы проанализировали фрагменты ДНК, которые были унаследованы без рекомбинации каждой парой индивидов от их недавнего общего предка, то есть сегменты (блоки), идентичные по происхождению (IBD).

Получены суммы средних длин сегментов IBD между парами индивидов. Построена тепловая карта с дендрограммой на основании логарифма от суммы средних длин сегментов IBD между парами индивидов. Количество общих сегментов хорошо отображает географию проживания популяций, так как проживающие рядом народы имеют более высокую вероятность метисации (рисунок 16).

Анализ тепловых карт демонстрирует кластеризацию популяций Кавказской, Дагестанской, Сибирской и Волго-Уральской групп, связывающую народности по месту происхождения, и небольшие кластеры внутри выявленных групп, демонстрирующие общее происхождение и большую степень родства.

Для популяций, населяющих территорию Сибири, характерна уникальная генетико-демографическая история, которая отражается в распределении блоков IBD как внутри популяции, так и между ними. Среди сибирских популяций чукчи, коряки и нивхи формируют отдельный от основной группы сибирских популяций кластер, причём чукчи и коряки являются более близкородственными. Формируются отдельные субкластеры эвенков и якутов, кетов и чулымцев внутри Сибирского кластера.

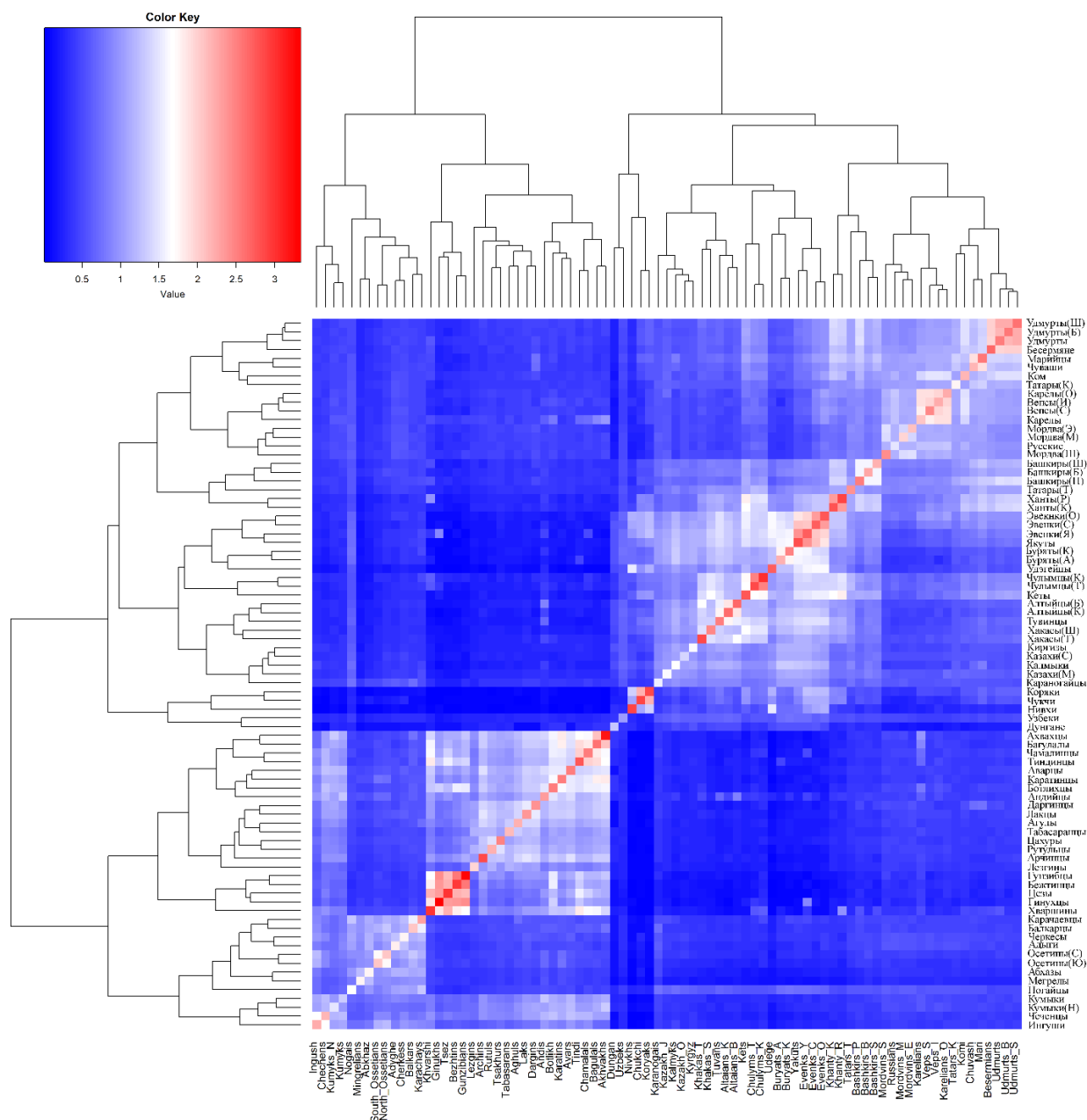


Рисунок 16 - Тепловая карта с дендрограммой на основании логорифма от суммы средних длин сегментов IBD (>1.5 cM) между парами индивидов.

При изменении анализируемых длин сегментов IBD в целом тенденция сохраняется, но появляются некоторые отличия, более точно характеризующие недавние взаимоотношения народов (рисунок А.1, А2). При увеличении длин блоков IBD кластеры больше соответствуют нынешнему географическому расположению популяций, отражая недавний обмен общими регионами. На

диаграмме с размерами IBD от 4 до 10 сМ (рисунок А.2) популяции лучше разбиваются на близко географически расположенные пары: коряки-чукчи, удэгейцы-нивхи. Кеты практически в равной степени имеют блоки IBD с чулымцами (8,7 сМ), хакасами и хантами, при этом среди хантов большее значение наблюдается у русскинских хантов (7,8 сМ), а среди хакасов у качинцев (8,2 сМ), что соответствует их более близкому географическому расположению по сравнению с хакасами Таштыпского района (4,2) и хантами Белоярского района (4,6).

За некоторыми исключениями, общие сегменты IBD между сибирскими популяциями лучше объясняются географической близостью популяций, а не их языковой принадлежностью. Так, например, якутские эвенки, проживающие на территории Якутии, имеют больше общих IBD с якутами (249 сМ), чем с забайкальскими эвенками (114,1 сМ). При этом сумма средних длин сегментов IBD между парами индивидов между двумя популяциями эвенков (114,4) сопоставима с суммой между забайкальскими эвенками и якутами (119,1 сМ). Большая сумма средних длин сегментов IBD между парами индивидов IBD для диапазона длин от 1 до 4 сМ по сравнению с диапазонами длин 4-10 сМ показывает более интенсивный обмен общими регионами в прошлом между популяциями эвенков, при этом обмен с якутами не менее интенсивен для забайкальских эвенков и более интенсивен для якутских эвенков, что соответствует их географическому положению. Далее для диапазона 4-10 сМ мы видим уменьшение обмена общими регионами между популяциями эвенков и между забайкальскими эвенками с якутами при сохранении значений для якутских эвенков с якутами. Можно отметить более интенсивное взаимодействие с другими популяциями в прошлом. Для эвенков отмечаются общие IBD (1-4 сМ) с ближайшими соседними популяциями, такими как тувинцы, удэгейцы, кеты и буряты. Чукчи активно взаимодействовали с коряками (115 сМ), что объяснимо их граничащими территориями и изолированностью от остальных популяций, помимо принадлежности к одной чукотско-камчатской языковой семье. Они также отделяются на графиках t-SNE от остальных Сибирских популяций

(Рисунок 4) и имеют уникальную предковую компоненту на относительно небольшом числе компонент ($K=6$).

С точки зрения анализа IBD внутри (а не между) популяциями, в целом индивиды из популяций, проживающих в настоящее время на Крайнем Севере или Дальнем Востоке (коряки, чукчи, нивхи), имеют больше общих IBD с образцами из той же группы, чем люди из популяций с более центральным расположением, таких как алтайцы и тувинцы. При этом у чукчей, коряков и нивхов наибольший вклад вносят короткие IBD фрагменты (55%–57%–59%), что может говорить о бутылочном горлышке в прошлом во время миграций на север и северо-восток и/или сильной изоляции от остальных популяций, населяющих территорию Сибири. При этом у чулымцев и хантов (Р), населяющих центральную часть Сибири и имеющих наибольшую сумму средних длин сегментов IBD между парами индивидов, наибольший вклад вносят IBD длиной более 10 cM (46–42%), что говорит о недавнем сильном инбридинге внутри популяции.

Популяции, характеризующиеся как наиболее гетерогенные среди сибирских популяций и имеющие минимальные значения по F_{ROH} , также имеют минимальные значения по сумме средних длин сегментов IBD. Наиболее выражено это у популяций калмыков и томских татар (Рисунок 17).

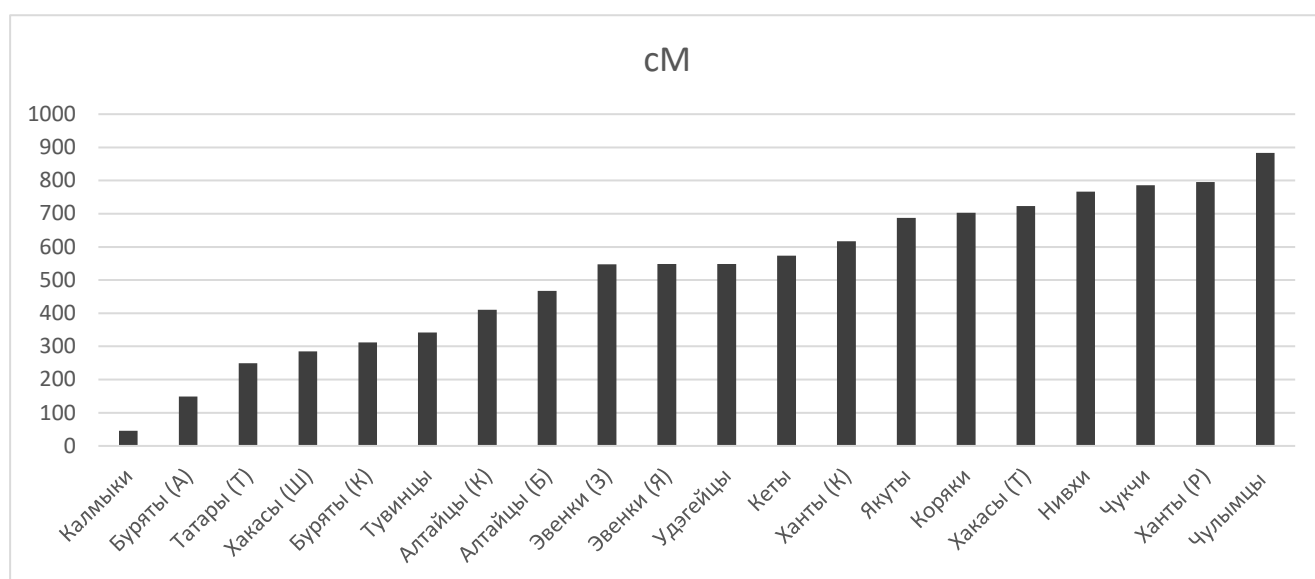


Рисунок 17 - Диаграмма суммы средних длин сегментов IBD между парами индивидов в исследуемых популяциях (> 1.5 cM).

В популяциях бурятов, имеющих схожее распределение предковых компонент в анализе Admixture, также нет различий по средней суммарной длине между парами индивидов с другими популяциями, но есть существенная разница в распределении IBD внутри популяций. Так у бурятов(К) значительно больше средняя суммарная длина между парами индивидов внутри популяции (312 сМ), по сравнению с бурятами(А) (149 сМ) во многом за счет средних и длинных IBD. У агинских бурят гораздо больше доля IBD размером от 1 до 4 сМ (56%), чем у курумканских (33%), с постепенным уменьшением для 4-10 сМ (35%) и IBD больше 10 сМ (19%). Несмотря на то что для бурятов, имеющих большую общую численность, и отмечено практически удвоение численности с 237 тыс. в 1926 году до 461 тыс. в 2010 году, данное различие между популяциями можно объяснить резким ростом численности населения поселка Агинское с 451 человека в 1908 году до 4556 чел в 1939 году и 15 тыс. в 2010 году, и стабильностью населения поселка Курумкан, с численностью 5617 человек в 1979 году и 5465 в 2010 году, находящегося в одном из отдаленных районов Бурятии. Похожая динамика наблюдается у популяций чукчей (55%-26%-19%), коряков (57%-21%-22%), нивхов (59%-24%-21%) и калмыков (69%-25%-6%) (Рисунок 18). Для чукчей, нивхов и коряков, имеющих небольшую численность (до 13 тыс.), характерно отсутствие резких колебаний численности популяций за последние сто лет с ростом на 29%-9%-6%, соответственно, что также могло способствовать сокращению общих длинных IBD сегментов при условии незначительного количества близкородственных браков внутри популяции.

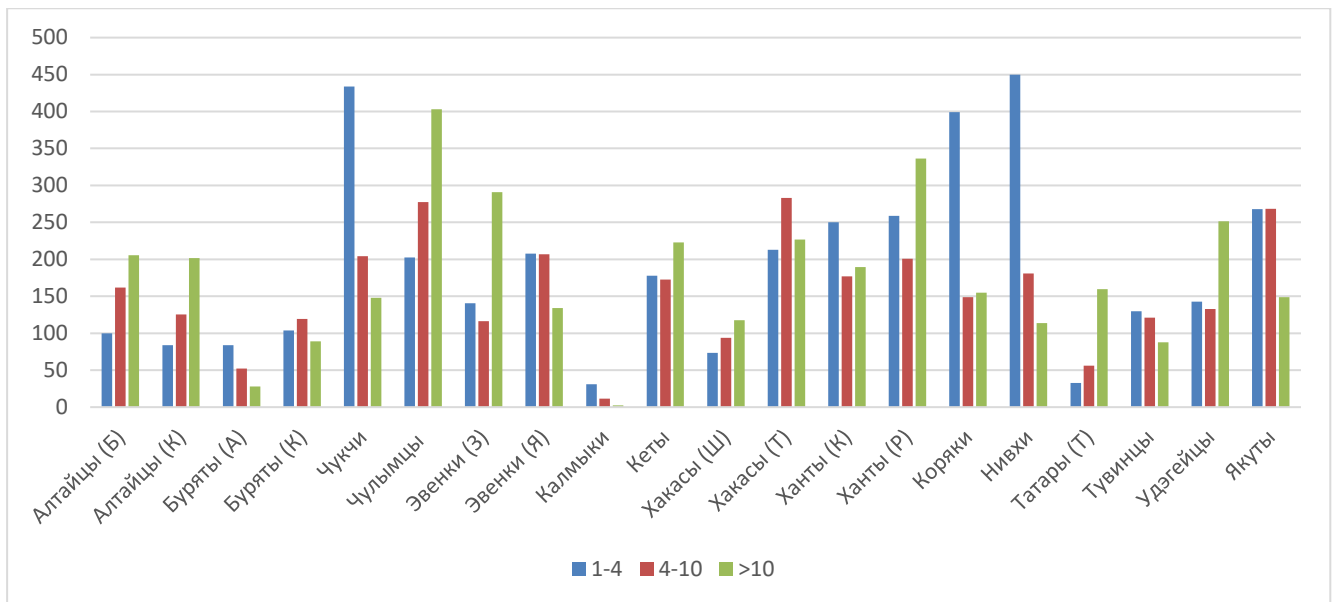


Рисунок 18 - Диаграмма суммы средних длин сегментов IBD между парами индивидов в исследуемых популяциях для трех диапазонов длин (сМ).

Для других исследуемых регионов также наблюдаются свои закономерности. В кластере Волго-Уральского региона формируются отдельные субкластеры, включающие карел и вепсов, бесермян и удмуртов. Данное разделение согласуется с близким географическим расположением и одной языковой принадлежностью данных популяций. Так, карелы и вепсы проживают в Республике Карелия и входят в одну прибалтийско-финскую языковую подгруппу, а бесермяне и удмурты занимают территорию Удмуртской республики и входят в пермскую языковую подгруппу. Также это подтверждается исследованием, где показана положительная корреляция между лексическим расстоянием уральских языков с генетическими расстояниями, рассчитанными по данным распределения аутосомных маркеров для уралоязычных популяций [132]. Помимо этого, при числе предполагаемых предковых популяций, равном 12 в Admixture, у бесермян и удмуртов выделяется собственный генетический компонент (Рисунок 7).

В дагестанском кластере все представители дидойской группы формируют отдельный кластер, включающий гинухцев, хваршин, цезов, бежтинцев и гунзибцев, при этом хваршины хоть и попадают в эту группу имеют меньше всего общих блоков с другими популяциями. Такое распределение хорошо коррелирует

с результатами PCA, так как хваршины фактически занимают промежуточное положение между остальными дагестанскими популяциями и представителями дидойской группы, хотя и сами в неё попадают.

Также стоит отметить более высокие значения между гинухцами и цезами, и бежтинцами, и гунзибцами внутри дидойской группы по сумме средних длин сегментов IBD между парами индивидов, что соответствует подразделению на восточную (бежтинцы, гунзибцы) и западную (гинухцы, цезы) языковые группы. По результатам Admixture разделение внутри дидойской группы на западноцезскую и восточноцезскую группы наблюдается на 10 предковых компонентах, при этом хваршины, попадая в западноцезскую группу, имеют всего 26% и 5% от восточноцезских популяций (Рисунок 7). Разделение протоцезской группы на протозападноцезскую и протовосточную по лексикостатистическими подсчетам датируется V в. до н. э. Разделение протозападноцезской группы на гинухский, протохваршинский и цезский датируется III в. н. э., а разделение протовосточноцезской на бежтинцев и гунзибцев IX в. н. э., в это же время протохваршинский разделяется на хваршинский и инхокваринский [133].

Популяции степных тюркских народов, проживающие на территории Кавказа и Дагестана, попадают в Сибирский кластер (кароногайцы) и в кластер с популяциями из Средней Азии (ногайцы), при этом, несмотря на географическую удаленность друг от друга, по данным Y-маркеров они генетически близки: в обеих группах выявлены восточно-евразийские гаплогруппы N1-LLY22 и C-M130 [134]. Это соотносится с антропологическими данными: по антропологическому типу их относят к южносибирской малой расе, переходной между большими монголоидной и европеоидной расами. Несмотря на то, что кароногайцы географически наиболее близки к кумыкам, они также попадают в разные кластеры. И наоборот, большее генетическое сходство балкарцев и карачаевцев согласуется с их географической близостью.

Анализ блоков IBD значительно дополняет картину в изучении образования и взаимодействия этносов. За некоторыми исключениями, общие сегменты IBD в

популяциях коренного населения Сибири лучше объясняются географической близостью популяций, а не их языковой принадлежностью.

3.6 Поиск сигналов направленного отбора с помощью тестов на протяженную гомозиготность гаплотипов

Оценка роли естественного отбора в формировании генотипической и фенотипической структуры популяций – одна из ключевых проблем эволюционной биологии, привлекающая внимание исследователей со времен Чарльза Дарвина до наших дней. Применительно к микроэволюционным процессам в популяциях человека разрабатывались и разрабатываются различные теоретические концепции и гипотезы, раскрывающие возможные механизмы действия отбора в отношении структуры генофонда и фенофонда популяций человека, в том числе моногенных и хронических распространенных болезней.

Результаты поиска сигналов, сформированных под действием направленного отбора в изучаемых популяциях коренного населения Сибири, с помощью теста nSL приведены в таблице 4 (для остальных изучаемых групп в таблице А.6). Всего было найдено 375 SNP ($\text{fdr} \geq 0,05$), попадающих в функциональные участки 284 генов, и 363 SNP, расположенных в межгенных участках. Для каждой из популяций обнаружено от 1 до 83 генов с сигналами направленного отбора. Наибольшее число значимых сигналов естественного отбора выявлено в популяциях хантов, коряков и чукчей. Пять генов, имеющих сигналы направленного отбора, повторяются не более чем у 3 популяций, 28 генов у двух, а остальные являются уникальными для отдельных популяций.

Обогащенность KEGG-путей генами под отбором была оценена с помощью ресурса webgestalt (WEB-based Gene SeT AnaLysis Toolkit). Из списка 287 генов, которые находятся в регионах генома, выявленных как мишени для направленного отбора по результатам тестов nSL, 84 по данным webgestalt оказались аннотированными по базе KEGG. 10 KEGG-путей, с учетом поправки Бенджамини–Хохберга (Таблица 5), были статистически значимы. Основными молекулярными путями были пути, отвечающие за синаптическую пластичность:

долговременная потенцияция (LTP - Long-term Potentiation) и долговременная депрессия (LTD - Long-term depression) – отвечающие за усиление или ослабление синаптической передачи между двумя нейронами, сохраняющейся на протяжении длительного времени после воздействия на синаптический проводящий путь, что обеспечивает нервную систему живого организма возможностью адаптироваться к изменяющимся условиям внешней среды. Также были отмечены пути дофаминергического и глутаматергического синапсов, передача сигналов ретроградного эндоканнабиноида, циркадный захват (отвечающий за синхронизацию эндокринных и поведенческих ритмов организма с сигналами окружающей среды), межклеточный щелевой контакт, обеспечивающий прямой перенос ионов и небольших молекул между соседними клетками, секреции слюны, регуляции липолиза в адипоцитах, сокращение гладкомышечных клеток сосудов.

При анализе объединённых данных тестов nSL и iHS значимыми остаются только пути долговременной депрессии (LTP) и циркадного захвата, но добавляется KEGG-путь морфиновой зависимости (hsa05032).

Основными биологическими процессами, подверженные направленному отбору, явились процессы регуляции, метаболические процессы, ответ на стимулы и др. Основными молекулярными функциями являлись связывание с белками, связывание ионов, нуклеиновых кислот, активность гидролазы и др. (Рисунок 19).

Таблица 4 - Результаты поиска сигналов направленного отбора с помощью теста nSL

Популяции	Гены с сигналами направленного отбора
Алтайцы (Б)	<i>ADCY1, DHX37, DYNC1H1, FOXP1, LINC00682, LINC02299, LOC101926960, MGLL, PDE10A, PPP2R2B, PTPRG, STK32B, SUCLG2-AS1</i>
Алтайцы (К)	<i>ADGRB3, ARL15, C5orf66, CELF2, LINC01378, LINC01622, RAPGEF2, RBFOX1, RBMS3, SOX5</i>
Буряты (А)	<i>STAG1</i>
Буряты (К)	<i>GLO1, PCDH15, PLSCR1, PRR16, ST8SIA6, TNIK</i>
Чукчи	<i>ACTA2, ADAMTS18, ANKRD60, CACNA2D3, CDCA7L, CDK5R2, CNTN5, FAM135B, FHIT, GABRG3, GMFG, HS1BP3-IT1, HYDIN, ITSN2, LOC100507283, MOBP, NALCN-AS1, NCALD, NLRP5, NRXN3, OPCML, PAPP, PCLO, PDLIM5, PGR, PRKACB, SAR1A, SCUBE2, TDRP, TENM2, TMEM65, TNIK, TSBP1-AS1, VAV2, VSTM2A</i>
Чулымцы	<i>ANKRD33B, CFAP299, DPP6, GRIA2, HLA-DQA1, LAMA1, LINC00473, LINC01435, LOC100128164, MEIS1, MIR646HG, MYH11, PREP, PTPRN2, RDH10-AS1, RGMB-AS1, SAMD7, SPATS2, STARD4-AS1, TMEM244, TMSB10, TMTC2, TSBP1-AS1, USH2A, XKR4, ZNF423</i>
Эвенки (Я)	<i>ADGRB3, AUTS2, DOCK8, DPP6, GABRB3, GAP43, LINC02294, LINC02743, LOC101928855, MYO15B, NR5A2, NTF3, PTPRN2, TEC, TENT5A</i>
Эвенки (З)	<i>ADH5, ANO3, FAM228B</i>
Калмыки	<i>CUEDC1, ERC1, MGST3, ZNF407</i>
Кеты	<i>LOC101928304, MCF2L2</i>
Хакасы (Ш)	<i>ANO3, CCDC138, CDH13, CUEDC1, EDAR, LINC00974, LINC02568, PRKG2, STK32B</i>
Хакасы (Т)	<i>AATK, CYP1B1-AS1, FOXP2, GRIK4, HLA-B, KIAA1324, LINC00448, LINC01630, LRP1B, MCPH1, MRC1, TMEM236, NCKAP5L, NCOA7-AS1, NRCAM, NTRK2, PALLD, PLD5, PPM1L, RARB, SDK2, ST6GALNAC1, TRPC6, UBB</i>
Ханты (К)	<i>ARFGEF3, ATP2C1, BTBD9, CCDC33, CD226, CDH13, CSMD2, DGKH, DNAH11, EGFLAM, FAM135B, GAP43, GDI2, GRIN2B, HCG9, ITPR3, KCNAB1, KDM1B, LARGE1, LINC00473, LINC01006, LINC01339, LINC02147, LOC100507283, MAPRE2, PDZD2, PRIMA1, ROR1, RSPO4, SAA3P, SLC6A15, SMARCA2, SMOC1,</i>

Продолжение таблицы 4

Популяции	Гены с сигналами направленного отбора
Ханты (P)	ADGRA3, AFF3, ALX1, ANK1, ANKRD33B, APBA1, ATP4B, ATXN1, CCND2, CDH13, CDKAL1, CNIH3, CNTN3, CSMD2, CXADR, DCANP1, TIFAB, DDX55, DSCAM, EPHA6, ERC1, FARS2, FAT2, GAREM1, GATAD2A, GDPD5, GLT1D1, GNAL, KIF6, KSR2, LINC00929, LINC01088, LINC01339, LINC02006, LINC02147, LINC02618, LOC101927630, LOC101929595, LOC105370586, LOC105374428, LOC339862, LRFN2, LYRM4-AS1, MTCH1, MYOM1, NAV2, OR5D16, OSBPL6, PBX4, PCDH15, PGR, PIP4K2A, PLEC, PLXNA2, PPP1R16B, PPP2R5C, RASGRP3, RDH13, RFC1, RMND1, RNF7, ROBO2, RYR1, SAA3P, SDK1, SLC14A2, SLC35G2, SLC46A2, SLX4IP, SNAP25, SOX1-OT, ST18, STAG1, SUGP1, TAS1R2, TCERG1L, TCF7L1, TCF7L2, TM6SF2, TMEM163, TSHZ2, TSNAX-DISC1, WWOX
Коряки	AGBL1, ATP8B1, CACNA2D3, CACNA2D3-AS1, CCDC169-SOHLH2, SOHLH2, CD96, CHI3L1, CLCN1, DOK7, FLJ22447, GACAT2, HDAC9, ITPR1, KCNAB1, LIMS1, LINC02542, LOC100507053, LOC105369911, LRRN2, MMP16, MREG, NLRP11, NRG1, OSTM1-AS1, - PCDH15, - PRKG1, SLC10A7, SLC16A12, SPIRE2, TMPRSS13, TNIP1, ZBED9
Нивхи	HCG9, LRRC52-AS1, TMEM237
Татары (Т)	ELMO1, LOC101927822
Тувинцы	ADH5, ANO3, BCAR3, CUEDC1, LIMS1, LOC101927822, LYRM4-AS1, RARB, RBFOX1, SULT1C4
Удэгейцы	HPSE, LOC101928764, RBMS3, ST8SIA1
Якуты	ADGRB3, CRTAC1, GLI3, HECW2, LCORL, NTRK3, PRDM6, RHOJ, SMOC1

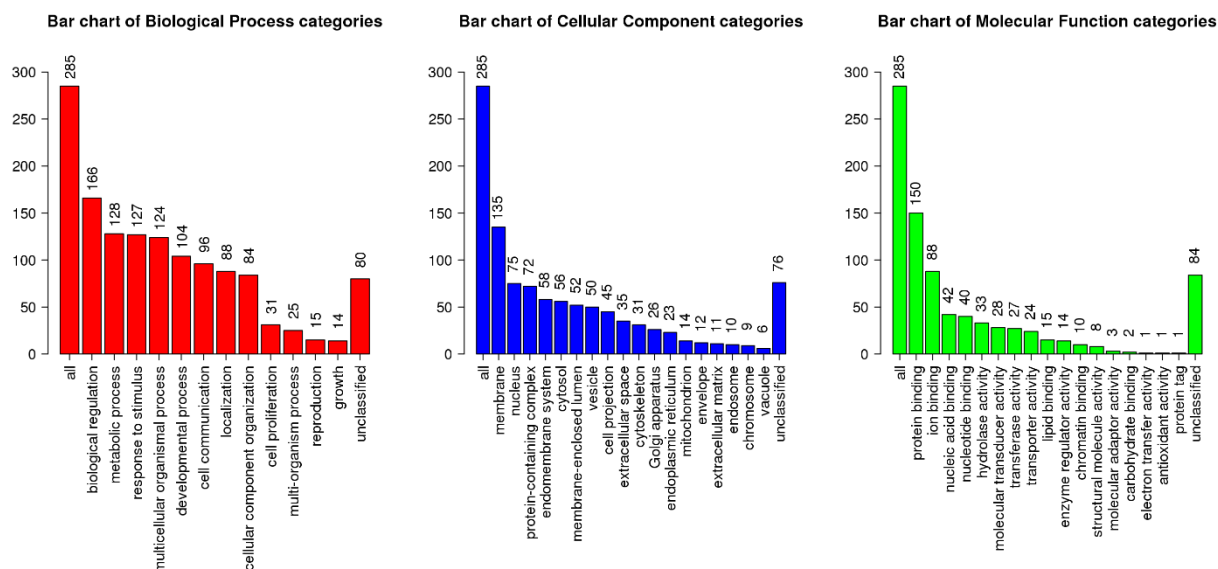


Рисунок 19 - Каждая категория биологического процесса, клеточного компонента и молекулярной функции представлена, красной, синей и зеленой полосой соответственно.

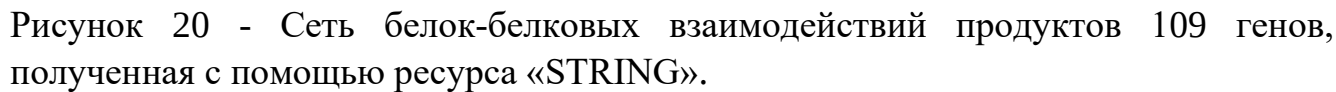
Таблица 5 - KEGG- пути

Набор генов	Описание пути	pValue	FDR	гены
hsa04713	циркадный режим	1,09E-06	3,6E-04	<i>ADCY1, GRIA2, GRIN2B, ITPR1, ITPR3, PRKACB, PRKG1, PRKG2, RYR1</i>
hsa04730	долговременная депрессия	5,21E-05	8,4E-03	<i>GRIA2, ITPR1, ITPR3, PRKG1, PRKG2, RYR1</i>
hsa04720	долговременная потенция	9,75E-05	8,4E-03	<i>ADCY1, GRIA2, GRIN2B, ITPR1, ITPR3, PRKACB</i>
hsa04728	дофаминергический синапс	1,03E-04	8,4E-03	<i>GNAL, GRIA2, GRIN2B, ITPR1, ITPR3, PPP2R2B, PPP2R5C, PRKACB</i>
hsa04723	ретроградный эндоканнабиноидный сигнал	2,40E-04	1,5E-02	<i>ADCY1, GABRB3, GABRG3, GRIA2, ITPR1, ITPR3, MGLL, PRKACB</i>
hsa04724	глутаматергический синапс	2,76E-04	1,5E-02	<i>ADCY1, GRIA2, GRIK4, GRIN2B, ITPR1, ITPR3, PRKACB</i>
hsa04923	регуляция липолиза в адипоцитах	3,27E-04	1,5E-02	<i>ADCY1, MGLL, PRKACB, PRKG1, PRKG2</i>

Продолжение таблицы 5

Набор генов	Описание пути	pValue	FDR	гены
hsa04270	сокращение гладкомышечных клеток сосудов	3,97E-04	1,6E-02	<i>ACTA2, ADCY1, ITPR1, ITPR3, MYH11, PRKACB, PRKG1</i>
hsa04540	щелевидное соединение	4,39E-04	1,6E-02	<i>ADCY1, ITPR1, ITPR3, PRKACB, PRKG1, PRKG2</i>
hsa04970	секреция слюны	4,95E-04	1,6E-02	<i>ADCY1, ITPR1, ITPR3, PRKACB, PRKG1, PRKG2</i>

Белок-белковые взаимодействия продуктов генов под отбором были визуализированы и проанализированы с помощью ресурса STRING (Search Tool for the Retrieval of Interacting Genes/Proteins). Анализ был проведен с учетом доказательств функциональных связей между белками, полученных экспериментально на основе их ко-экспрессии и совместной представленности в биологических процессах. Соупоминание в текстах публикаций было исключено из критериев взаимодействия. По результатам тестов nSL и iHS (таблица 7) выявляется 492 гена, расположенных в участках генома, несущих сигналы направленного отбора. При максимальном уровне доказательности связей (highest confidence, 0,9) из 399 белков, аннотированных в STRING, 109 оказались вовлечены в общие сети белок-белковых взаимодействий (уровень значимости обогащенности сети $P = 0,0435$). Полученная сеть представлена на рисунке 20.



Всего было найдено 86 категорий онтологии генов в биологических процессах. (Таблица А.4). 31 из них включают 101 ген относящихся к развитию и регуляции центральной нервной системы (*ATXN1*, *SNAP25*, *DOCK10*, *LRP6*, *RIMS2*, *CNTN3*, *RAPGEF2*, *APBA1*, *SMARCA2*, *BRINP1*, *MEIS1*, *ARHGAP26*, *NTRK2*, *GRID2*, *ADCY1*, *CSMD3*, *ZNF804A*, *CDK5R2*, *UBB*, *USH2A*, *GABRB3*, *RBFOX1*, *MOBP*, *ALX1*, *MYT1*, *DENND5A*, *EFNA5*, *OPCML*, *GAS6*, *RARB*, *PCLO*, *GDPD5*, *TRPC6*, *MCPH1*, *AUTS2*, *ARID1B*, *APOD*, *DLG1*, *LRFN2*, *ITSN2*, *SYNE2*, *NTRK3*, *PARK2*, *PLXNA2*, *LRRN2*, *TNR*, *RTN4IP1*, *PAX2*, *CRTAC1*, *BAI3*, *PRKACB*,

ROR1, KIF20B, MTCH1, PCDH15, PRKG1, NRCAM, RUFY3, SEMA5A, XRCC5, SDK2, HYDIN, GAP43, AK4, GLI3, NAV2, EPHB1, PCDHGA8, DSCAM, NRG1, PRDM6, SDK1, FOXP2, RDH13, NTF3, SOX5, TNIK, MYO16, HDAC9, ROBO2, PTPRG, KCNAB1, DCANP1, TENM2, TIFAB, CYFIP2, ATP8B1, KCNC2, NRXN3, ZNF423, DNAH11, GRIN2B, PDLIM5, LINGO2, CCND2, ITPR3, PRKAR1B, BTBD9, TMPRSS11E, MRAP2, GPHN).

Десять биологических процессов, включают 31 ген (*LRP6, MYOM2, MEIS1, NTRK2, ADAMTS18, USH2A, GABRB3, RHOJ, RARB, DLG1, MYOM1, NTRK3, SMOC1, BCAR3, PAX2, ROR1, PCDH15, SDK2, GLI3, EPHB1, DSCAM, SDK1, FOXP2, RDH13, DCANP1, TIFAB, ATP8B1, NAV2, EFNA5, NRCAM, ROBO2*) относящихся к процессам сенсорного восприятия, включающих формирование и развитие зрения, слуха. Также интересны гены, вовлеченные в биологический процесс развития костей и скелетной системы (*MEIS1, MMP16, BBX, FAM20C, RARB, MCPH1, RYR1, AKAP13, PAPP2, SLC38A10, PIP4K2A, GLI3, DCANP1, TIFAB, FOXP1, CHI3L1, LRP6, ALX1, DLG1, VDR, WWOX*) и системы кровообращения (*HDAC7, LRP6, RAPGEF2, CDH13, MEIS1, NTRK2, CXADR, TMEM65, RHOJ, RARB, APOD, RYR1, NTRK3, AKAP13, TCF7L2, VAV2, NRCAM, SEMA5A, NOX5, GLI3, MYH11, EPHB1, HPSE, NRG1, XIRP2, ACTA2, HDAC9, ROBO2, KCNAB1, NRXN3, DNAH11, PDLIM5*).

Таблица 6 - Результаты поиска сигналов направленного отбора с помощью теста iHS

Популяции	Гены с сигналами направленного отбора
Алтайцы (Б)	<i>FAM20C, FSTL1, LINC00565, LINC02299, MRC1, MRC1, TMEM236, RNASEH1, RNASEH1-AS1, RPS7, SUCLG2-AS1, TMEM255B, TTLL1</i>
Алтайцы (К)	<i>ATP4B, CCDC146, ERC1, FSIP2, GABBR1, GAS6, ISX-AS1, LIMS1, NUDCD1, PKIA, RNASEH1-AS1, SOX5, TMEM255B</i>
Буряты (А)	<i>ANKS1B, ATG10, BCAR3, ERC1, FAM184B, ISX-AS1, LINC01029, LOC101928937, MED28, MPPED1, RIMS2, STAG1, TMEM255B, TTLL1, TWIST2, YTHDC1</i>
Буряты (К)	<i>ADI1, COLEC11, ERC1, FAM20C, GAS6-AS1, GLO1, LINC00452, RNASEH1, RNASEH1-AS1, VWA3A</i>

Продолжение таблицы 6

Популяции	Гены с сигналами направленного отбора
Чукчи	<i>AKAP13, ANO2, CDCA7L, ENOX1, GALC, GMFG, HS1BP3-IT1, ISX-AS1, ITSN2, LINC02538, LOC100507283, MOBP, MRC1, TMEM236, NRXN3, OPCML, PRKACB, PRKAR1B, SCUBE2, TMEM65, TSBP1-AS1</i>
Чулымцы	<i>LAMA1, LINC00454, LOC643542, MIR646HG, PLD6, RALYL, SPATS2, TSBP1-AS1</i>
Эвенки (Я)	<i>ADGRB3, ARFGAP3, DOCK8, DPP6, DSCAM, ERC1, FAM20C, GAS6-AS1, HDAC7, LOC101928324, MYT1, SCHLAP1</i>
Эвенки (З)	<i>ADH5, FAM20C, TMPRSS11E</i>
Калмыки	<i>BBX, C4orf47, COL22A1, CUEDC1, EGFLAM, ERC1, FAM20C, GAS6, KCNC2, KDM4B, LINC00261, MPPED1, MRC1, ORC2, PAPP2, PAX2, RHEX, SPATS2, SUCLG2-AS1, TNF, UFSP2, USP20, ZKSCAN7, ZNF407, ZNF445, ZNF546</i>
Кеты	<i>ARID1B, ATE1, CDKAL1, CYP4X1, DMBT1, EFNA5, ERI2, FAM20C, GAS6-AS1, GNL3, GRID2, HCG17, HCG17, HCG18, HLA-L, LAMC1, LINC01500, LINC02234, LINC02487, LINGO2, LOC100506679, LOC101928304, MRC1, MYO16, PDE3A, PDZRN4, RABGAP1L, SLC2A9, STK38, TOM1L1, TTLL1, TWIST2, WWOX, ZDHHC14</i>
Хакасы (Ш)	<i>ADI1, ANO3, CCDC138, COLEC11, DENND5A, EDAR, EGFLAM, FAM20C, GABBR2, GAS6-AS1, MRC1, PPEF2, PRKAR1B, RNASEH1, TMEM255B, ZNRD1, ZNRD1ASP</i>
Хакасы (Т)	<i>CSMD1, DIAPH3, FAM186B, GAS6, GAS6-AS1, GNL1, GPHN, ISX-AS1, LOC440084, MCPH1, MRC1, MRC1, TMEM236, PPP1R10, TMEM255B</i>
Ханты (К)	<i>ATP2C1, CDH13, COLEC11, CSMD2, DNAH11, FAM20C, GAS6, GAS6-AS1, GDI2, LINC00452, LOC100506679, MAPRE2, MRC1, RNASEH1, RNASEH1-AS1, RPS7, TBCK, TSBP1, TSBP1-AS1</i>
Ханты (Р)	<i>ADGRA3, AFF3, ATP4B, C9orf129, CCND2, DDX55, DSCAM, EEFF1AKMT2, FAM53B, GALNT9, GAREM1, LINC02147, LINC02538, LRRIQ1, PAPP2, PBX4, PGR, PIP4K2A, PLEC, PRKAR1B, RDH13, RNASEH1, RNASEH1-AS1, RNF7, SNAP25, STAG1, TCERG1L, TMPRSS11E, ZNF236</i>
Коряки	<i>ADH5, AFG1L, ANO3, BRINP1, CABP5, CACNA2D3, CACNA2D3-AS1, CCDC169-SOHLH2, CCDC85A, CLCA4-AS1, CORO1B, CYFIP2, CYP2C19, CYTH4, DLG1, FAM228B, FLJ22447, FSTL1, GAS6, INPP5D, KIF20B, LGR6, LIMS1, LINC02227, LINC02532, LINC02538, MARCHF4, METTL5, NLRP11, PCDH15, PRKG1, PRKN, RABL3, RNASEH1, RNF217, RPN1, RPS6KB2, RTN4IP1, SUGCT, SYNE2, TMEM169, TRAPPC10, TRIM10, XRCC5, ZBED9, ZC3H14, ZFAND3, ZNF236, ZNF333, ZNF804B</i>

Продолжение таблицы 6

Популяции	Гены с сигналами направленного отбора
Нивхи	<i>ABCB11, ADGRB3, APOL1, BZW1, CDH13, ENDOD1, EPHB1, G3BP2, GABBR1, KCNIP4, KPNA4, LCORL, LINC00457, LINC02343, LOC284930, MAP3K2, MIR4500HG, MIR4713HG, MRC1, MRC1, TMEM236, MYOM2, NELFCD, NSUN7, OR4E2, ROBO2, SEMA5A, SLC35F3, SORCS2, TANC2, TCF7L2, TMEM236, VARS2, ZBTB38, ZNF398, ZNF568, ZNF804A, ZNRD1ASP</i>
Татары (Т)	<i>ADH5, AK4, ALMS1, APOD, ARHGAP26, BDNF-AS, BEND3P3, COLEC11, DOCK8, EFCAB6, ELMO1, GAS6, GPATCH1, HSD17B12, LOC101927822, LPIN1, LRP6, LRRC36, MPPED1, MRC1, MRC1, TMEM236, NLRP1, NOL8, NOX5, PAPP2, RNF121, RUFY3, SEMA5A, SENP7, SLC25A12, SLC38A10, SLC44A1, TMEM255B, TMPRSS11E, TMTC2, TOX2, XIRP2, ZBTB38, ZNF804A</i>
Тувинцы	<i>ADH5, CUEDC1, FAM20C, GAS6-AS1, LIMS1, MOK, RNASEH1-AS1, SLC2A9, SULT1C4, ZBTB38</i>
Удэгейцы	<i>ACSL1, ANKS1B, CSMD3, DOCK10, LEKR1, LOC374443, MAP3K7, MPPED1, MRAP2, MST1R, OR11A1, PCDHGA8, PPP1R15A, PRKAR1B, RBMS3, SLC25A26, TMEM255B, VDR, ZNF398, ZNF804A</i>
Якуты	<i>LCORL, MRC1, SACM1L</i>

Определенные функциональные полиморфизмы в ферментах *ADH*, которые вызывают различную эффективность превращения этанола в ацетальдегид, могут быть защитными. Голдман и Енох [135] предположили, что генетические вариации в генах *ALDH* и *ADH* избирательно поддерживаются, и предложили две вероятные селективные силы: микотоксины и инфекционные заболевания. *ADH5* участвует в механизме защиты от эндогенного формальдегида, детоксифицируя формальдегид, тем самым защищая клетку от этого генотоксина. В наших данных четыре популяции характеризуются высокими значениями тестов nSL (эвенки (3), тувинцы) и iHS (коряки, татары (Т), тувинцы, эвенки (3)) для *kgp7824500* располагающегося в UTR3 гена *ADH5*, что также может быть связано с защитными функциями. Регион, в котором находится *ADH5*, ранее был отмечен как регион, находящийся под отбором у восточно-азиатских популяций [136].

В восточноазиатском регионе одним из самых популярных кандидатом-участком является *EDAR*. Он кодирует белок, отвечающий за толщину волос и морфологию резцов, и неоднократно идентифицировался как ген-«кандидат» на отбор у жителей Восточной Азии [104,136,137]. В нашем исследовании сигнал

направленного отбора в этом гене, а также в близкорасположенном гене *CCDC138* встречается в популяции хакасов-качинцев (Ш). Помимо этого, рядом с *EDAR* располагается ген *LIMS1*, попадающий в регион, находящийся под отбором у восточно-азиатских популяций [136]. Для популяций коряков, тувинцев (nSL, iHS) и алтайцев (K) (iHS) наблюдаются высокие значения по гену *LIMS1*, белок которого участвует в регуляции выживаемости, пролиферации и дифференцировки клеток (Таблица 7).

Ранее в популяциях коренного населения Сибири был обнаружен целый ряд генов с сигналами естественного отбора, предположительно вовлеченных в процесс адаптации к холоду (*THADA*, *ITPR3*, *GNGT1*, *PRKG1*, *RELN*, *CPT1A* и *LRP5*) [12]. Среди них в наших популяциях сигналы направленного отбора встречаются у чукчей в гене *LRP5* кодирующим белок, связанный с рецептором липопротеинов низкой плотности, что играет важную роль в процессах роста костей, метаболизма холестерина, систолического кровяного давления, хантов (P) – в гене *ITPR3*, играющем ключевую роль в экзокринной секреции, лежащей в основе энергетического метаболизма и роста, коряков – в гене *PRKG1*, отвечающем за сокращение гладкомышечных клеток сосудов. Вероятно, гены и маркеры, находящиеся под направленным отбором в исследуемых популяциях, могут играть существенную роль в адаптации к неблагоприятным или экстремальным климато-географическим условиям.

Также для калмыков, хантов (P) и татар наблюдаются высокие значения по двум тестам в гене *PAPPA2*, регулирующем высвобождение ключевого фактора роста, называемого инсулиноподобным фактором роста 1 (IGF-1) и связанного с ростом человека [138].

Интересно, что у 9 из 20 популяций по тесту iHS (Таблица 6) в гене *FAM20C* есть значимые сигналы направленного отбора. Казеинкиназа Гольджи, кодируемая геном *FAM20C*, фосфорилирует белки секреторного пути внутри мотивов SxE, фосфорилирует казеины и несколько секретируемых белков, участвующих в биоминерализации, включая малый интегрин-связывающий лиганд, N-связанные гликопротеины, что играет важную роль в формировании

костей и зубов[139]. Мутации в этом гене ответственны за развитие синдрома Рейна (OMIM: 259775, неонатальная остеосклеротическая дисплазия костей).

Таблица 7 - Гены под направленным отбором, представленные в более чем 3 популяциях теста nSL или iHS в более чем 2 для nSL и iHS

Ген	количество популяций в nSL/iHS	Функция продукта
<i>ERC1</i>	2/5	Белок, кодируемый этим геном, является членом семейства RIM-связывающих белков. RIM — это белки активной зоны, которые регулируют высвобождение нейромедиаторов.
<i>ADH5</i>	2/4	Этот ген кодирует глутатион-зависимую формальдегиддегидрогеназу или хи-субъединицу алкогольдегидрогеназы класса III, которая является членом семейства алкогольдегидрогеназ. Играет важную роль в метаболизме формальдегида в организме человека.
<i>LIMS1</i>	2/3	Белок, кодируемый этим геном, представляет собой адаптерный белок, который содержит пять доменов LIM, или двойные цинковые пальцы. Белок, вероятно, участвует в передаче сигналов интегрин через его LIM-домен-опосредованное взаимодействие с интегрин-связанной киназой, обнаруживаемой в очаговых адгезионных бляшках. Также считается, что он действует как мостик, связывающий интегрин-связанную киназу с адаптерным белком NCK 2, который участвует в сигнальных путях киназы рецептора фактора роста.
<i>TSBP1-AS1</i>	2/3	<i>TSBP1-AS1</i> (<i>TSBP1</i> и <i>BTNL2</i> антисмысловая РНК 1) представляет собой ген РНК и относится к классу lncRNA. Заболевания, связанные с <i>TSBP1-AS1</i> , включают саркоидоз 2.
<i>ADGRB3</i>	3/2	Этот ген-мишень P53 кодирует специфический для мозга ингибитор ангиогенеза, трансмембранный белок с семью участками, и считается, что он является членом семейства рецепторов секретина.

Продолжение таблицы 7

Ген	количество популяций в nSL/iHS	Функция продукта
<i>ANO3</i>	3/2	Кодирует белок аноктамин 3. Он принадлежит к семейству генов (<i>ANO1</i> – <i>ANO10</i>), которые, по-видимому, кодируют активируемые кальцием хлоридные каналы. Мутации в этом гене были связаны с некоторыми случаями аутосомно-доминантной краниоцервикальной дистонии. Клетки от людей с мутацией в этом гене проявляли аномалии в передаче сигналов кальция, зависимой от эндоплазматического ретикулула.
<i>CDH13</i>	3/2	Ген <i>CDH13</i> кодирует белок кадгерин 13, участвующий во многих процессах в мозге: от миграции клеток у эмбриона до развития аксонов и генерации синапсов между нейронами у взрослых. Разные варианты этого гена связаны с расстройствами аутистического спектра, синдромом гиперактивности и дефицита внимания, депрессией, ухудшением памяти и другими нарушениями
<i>CUEDC1</i>	3/2	<i>CUEDC1</i> (домен CUE, содержащий 1) является геном, кодирующим белок. Заболевания, связанные с <i>CUEDC1</i> , включают проксимальный симпалангизм и синдром Перро. <i>CUEDC1</i> связывается с K33 и K63 убиквитина в различных типах клеток человека, предполагая, что он может участвовать в путях транспортировки белков, передачи сигналов и деградации.
<i>STAG1</i>	2/2	Этот ген является членом семейства <i>SCC3</i> и экспрессируется в ядре. Он кодирует компонент когезина, многосубъединичного белкового комплекса, который обеспечивает сцепление сестринских хроматид по длине хромосомы за счет репликации ДНК через профазу и прометафазу, после чего он диссоциирует для подготовки к сегрегации во время анафазы.
<i>PCDH15</i>	3/0	Этот ген является членом суперсемейства кадгеринов. Члены семейства кодируют интегральные мембранные белки, которые обеспечивают кальций-зависимую межклеточную адгезию. Играет важную роль в поддержании нормальной функции сетчатки и улитки. Мутации в

Продолжение таблицы 7

Ген	количество популяций в nSL/iHS	Функция продукта
		этом гене были связаны с потерей слуха, также было обнаружено, что вариации внутри него связаны с нормальными различиями у людей внешнего вида лица.
<i>FAM20C</i>	0/9	Серин / треониновая протеинкиназа Гольджи, которая фосфорилирует белки секреторного пути в мотивах Ser-x-Glu / P Ser и играет ключевую роль в биоминерализации костей и зубов. Помимо своей роли в биоминерализации, он также играет роль в гомеостазе липидов, заживлении ран, миграции и адгезии клеток
<i>MRC1</i>	0/9	Белок, кодируемый этим геном, представляет собой мембранный рецептор типа I, который опосредует эндоцитоз гликопротеинов макрофагами. Было показано, что белок связывает высокоманнозные структуры на поверхности потенциально патогенных вирусов, бактерий и грибов, так что они могут быть нейтрализованы фагоцитарным поглощением.
<i>GAS6-AS1</i>	0/7	Представляет собой ген РНК и относится к классу днРНК. Заболевания, связанные с <i>GAS6-AS1</i> , включают рак легких и почечно-клеточную карциному.
<i>TMEM255B</i>	0/7	<i>TMEM255B</i> представляет собой ген, кодирующий трансмембранный белок 255В. Заболевания, связанные с <i>TMEM255B</i> , включают гематурию, доброкачественную семейную.
<i>GAS6</i>	0/6	Gas6 представляет собой белок, содержащий домен гамма-карбоксиглутаминовой кислоты (Gla), который, как считается, участвует в стимуляции пролиферации клеток.
<i>RNASEH1</i>	0/6	Ген кодирует фермент RNase H1 являющийся неспецифической эндонуклеазой и катализирующей расщепление РНК по гидролитическому механизму.
<i>RNASEH1-AS1</i>	0/6	<i>RNASEH1-AS1</i> (антисмысловая РНК 1 RNASEH1) представляет собой ген РНК и относится к классу днРНК.

Продолжение таблицы 7

Ген	количество популяций в nSL/iHS	Функция продукта
<i>MRC1</i> , <i>TMEM236</i>	0/5	<i>MRC1</i> принадлежит к семейству рецепторов маннозы (MR), <i>TMEM236</i> (трансмембранный белок 236).
<i>COLEC11</i>	0/4	Этот ген кодирует члена семейства лектинов С-типа коллинов, которые обладают коллагеноподобными последовательностями и доменами распознавания углеводов. Коллектины — это секретируемые белки, которые играют важную роль в системе врожденного иммунитета, связываясь с углеводными антигенами микроорганизмов, облегчая их распознавание и удаление. Кодированный белок связывается с множеством сахаров с предпочтением фукозы и маннозы. Мутации в этом гене являются причиной синдрома ЗМС-2.
<i>ISX-AS1</i>	0/4	<i>LINC02885</i> (длинная межгенная небелковая кодирующая РНК 2885) представляет собой ген РНК и относится к классу днРНК.
<i>MPPED1</i>	0/4	<i>MPPED1</i> - представляет собой ген, кодирующий белок белок 1, содержащий домен металлофосфоэстеразы. Заболевания, связанные с <i>MPPED1</i> , включают опухоль Вильмса, аниридию, мочеполовые аномалии, синдром умственной отсталости и аниридию 1. Аннотации Gene Ontology (GO), относящиеся к этому гену, включают гидролазную активность.
<i>PRKAR1B</i>	0/4	Белок, кодируемый этим геном, представляет собой регуляторную субъединицу циклической АМР-зависимой протеинкиназы А (РКА), которая участвует в сигнальном пути второго мессенджера цАМФ. Холофермент участвует во многих клеточных событиях, включая перенос ионов, метаболизм и транскрипцию. Протеинкиназа А (РКА, также известная как цАМФ-зависимая протеинкиназа) участвует в регуляции метаболизма липидов и глюкозы и является компонентом механизма передачи сигнала некоторых GPCR.
<i>LINC02538</i>	0/3	<i>LINC02538</i> (длинная межгенная небелковая кодирующая РНК 2538) представляет собой ген РНК и относится к классу днРНК.

Продолжение таблицы 7

Ген	количество популяций в nSL/iHS	Функция продукта
<i>PAPPA2</i>	0/3	<i>PAPPA2</i> - представляет собой ген, кодирующий белок паппализин 2. Среди связанных с ним путей - метаболизм белков и регуляция транспорта и поглощения инсулиноподобного фактора роста (IGF) белками, связывающими инсулиноподобный фактор роста (IGFBP). Аннотации Gene Ontology (GO), относящиеся к этому гену, включают активность металлопептидазы.
<i>TMPRSS11E</i>	0/3	<i>TMPRSS11E</i> - представляет собой ген, кодирующий белок трансмембранной сериновой протеазы 11E. Генная Онтология (ГО) аннотации, связанные с этим геномом, включают эндопептидазную активность серинового типа и пептидазу активность серинового типа.
<i>TTLL1</i>	0/3	<i>TTLL1</i> - представляет собой ген, кодирующий белок тубулинполиглютамилазу TTLL1. Заболевания, связанные с <i>TTLL1</i> , включают астму с кашлевым вариантом и ларингит. Аннотации Gene Ontology (GO), относящиеся к этому гену, включают активность лигазы тубулина и глутаминовой кислоты.
<i>ZBTB38</i>	0/3	Белок, кодируемый этим геном, является активатором транскрипции «цинковый палец», который связывает метилированную ДНК. Кодируемый белок может образовывать гомодимеры или гетеродимеры через домены «цинковые пальцы». У мышей ингибирование этого белка было связано с апоптозом некоторых типов клеток. Аннотации онтологии гена (GO), относящиеся к этому гену, включают активность связывания нуклеиновых кислот и гомодимеризации белков.
<i>ZNF804A</i>	0/3	Кодирует белок цинковых пальцев 804A. У людей <i>ZNF804A</i> широко экспрессируется в головном мозге, особенно в развивающемся гиппокампе и коре головного мозга, а также в мозжечке взрослых. Предполагается, что <i>ZNF804A</i> будет связывать ДНК и, таким образом, регулировать экспрессию генов, как и другие белки «цинковые пальцы».

Продолжение таблицы 7

Ген	количество популяций в nSL/iHS	Функция продукта
		Полногеномный поиск ассоциаций (GWAS) определил <i>ZNF804A</i> как ген-кандидат на шизофрению. <i>ZNF804A</i> предположительно может быть связан с риском обморока и коллапса.

3.6.1 Сигналы направленного отбора, выявленные с помощью теста ХР-ЕНН

Для поиска сигналов отбора в популяциях Сибири, проживающих в условиях холодного климата, для сравнения были выбраны наиболее отличные от Сибирских – евразийские популяции, не имеющие общих генетических компонент при анализе Admixture, проживающие в умеренном климате (узбеки, русские, дунгане, вепсы, карелы, южные и северные осетины, даргинцы, аварцы) (Таблица А.5).

Всего было найдено 41755 SNP ($FDR \geq 0,05$), попадающих в функциональные участки 1129 генов и 16707 SNP в межгенных участках. Из них 245 SNP несут несинонимичные замены и находятся в 62 генах, 12 из которых входят в группу генов, отвечающих за трансдукцию сенсорного сигнала (*LC24A1*, *SLC45A2*, *OR51Q1*, *OR5AK2*, *OR51B2*, *OR51M1*, *OR13C8*, *SLC24A5*, *OR52B6*, *OPA1*, *OR13C5*, *OR51B6*) ($FDR = 6.97e-05$), 8 – обоняния (*SLC24A1*, *SLC45A2*, *OR51Q1*, *OR5AK2*, *OR51B2*, *OR51M1*, *OR13C8*, *SLC24A5*, *OR52B6*, *OPA1*, *OR13C5*, *OR51B6*) ($FDR = 0,0087$), и 59 генов несущие различные полиморфизмы (*DH17*, *SETX*, *TEX14*, *FLT3*, *IGHMBP2*, *EXOC6*, *SLCO1B3*, *SLC24A1*, *EFCAB6*, *MECR*, *COL6A5*, *ZKSCAN7*, *MCAT*, *TPCN2*, *SLC45A2*, *OR51Q1*, *VPS37C*, *RNASE3*, *ADAT1*, *APOL1*, *GEN1*, *CTNNAL1*, *OR5AK2*, *CHMP4A*, *OR51B2*, *BRWD1*, *OAF*, *OR51M1*, *OR13C8*, *PRPF4B*, *PPL*, *UGT2B15*, *SLC24A5*, *OR52B6*, *TNFSF13*, *SERTAD1*, *ANXA10*, *OPA1*, *ARHGEF11*, *SRSF4*, *OR13C5*, *OR51B6*, *FANCA*, *VWA3A*, *TRMT44*, *RAD51AP2*, *SEH1L*, *MCC*, *DNAH10*, *TLR6*, *HERC1*, *MFSD1*, *CAMTA2*, *ZNF805*,

GPR156, MUC4, MRC1, WWOX, ZNF501) ($FDR = 1.09e-07$). При этом три гена (*OR51B2, OR51M1, OR51B6*) попадают в локальный сетевой кластер (local network cluster, STRING) из семи генов ($FDR = 0,0161$), отвечающих за обонятельные рецепторы, и восемь генов (*OR51Q1, OR5AK2, OR51B2, OR51M1, OR13C8, OR52B6, OR13C5, OR51B6*) по базе данных основных молекулярных путей (KEGG-путь) входят в кластер, отвечающий за пути обонятельной трансдукции ($FDR = 0,0194$).

Для выявления наиболее значимых генов с сигналами направленного отбора для популяций, проживающих на территории Сибири, мы взяли гены, в которых выявлены сигналы отбора как минимум в четырех сибирских популяциях и в трех популяциях из группы сравнения. В итоге было выявлено 26 генов (*APOL1, ATP4B, CHD2, EZR, FBN1, GCC2, GCC2-AS1, GRK1, LIMS1, LINC00376, LINC00565, LINC01250, LINC01578, LOC100506403, LOC112694756, MIR1246, MPPED1, MRC1, MYT1, PCDH15, RNASEH1, RNASEH1-AS1, RPS7, SLC24A5, SLC45A2, TMEM255B*) из них 14 генов (*CHD2, EZR, FBN1, GCC2, GCC2-AS1, GRK1, LINC00376, LINC01250, LINC01578, LOC100506403, LOC112694756, MIR1246, SLC24A5, SLC45A2*) не обнаружены по тестам nsI и iHS в популяциях Сибири. Два гена связаны с синтезом меланина (*SLC45A2*) и регуляции концентрации кальция в меланоцитах (*SLC24A5*), что определяет интенсивность пигментации человека и является адаптацией к уровню УФ-излучения [140,141]. Кроме того, в различных европейских популяциях было показано участие генетических вариантов *SLC45A2* в восприимчивости к меланоме [142]. Ген *GRK1* кодирует белок, фосфорилирующий родопсин и участвующий в восстановлении и адаптации клеток сетчатки глаза. Это играет важную роль в обеспечении ночного зрения и для быстрой адаптации глаза к быстрой смене интенсивности света [143]. Положительный отбор по *SLC45A2* особенно интересен, поскольку сцепленные вредные SNP могут быть связаны с заболеваниями человека [144]. На примере северных европейцев из Юты показано (CEU - Northern Europeans from Utah), что положительный отбор по этому гену мог влиять на частоту «вредных» SNP в *FBN1*, располагающийся на 265 т.п.о. ниже *SLC24A5*. *FBN1* имеет пять

«вредных» SNP по данным HarMap CEU (частота 0,5–1,4%), но данные SNP отсутствуют в африканских и в азиатских образцах HarMap. Три из этих «вредных» SNP вызывают синдром Марфана [145,146], а один был обнаружен у пациентов с синдромом Марфана или родственными фенотипами [147].

Популяции Северной Евразии, в частности, в Сибири и Северо-Восточной Азии, представляют собой «крайние» ветви как с точки зрения расселения современного человека, так и с позиций экстремальности климато-географических условий их среды обитания. Очевидно, что генетические особенности этих популяций сложились под сильным действием естественного отбора, и их генофонд прошел наиболее долгий путь к деканализации сложившихся в предковых популяциях человека отношений геном-феном и к формированию собственных канализованных экстремальными средовыми условиями характеристик.

ЗАКЛЮЧЕНИЕ

Выполненное научное исследование находится в рамках фундаментальных направлений популяционной и эволюционной генетики человека – выявления этноспецифичных особенностей структуры геномов и популяционных генофондов современных популяций человека, реконструкции древних миграций и экспансии численности населения, а также эволюции генофондов. В результате проведения исследования на основании генотипирования большого массива маркеров и комплексного анализ различных составляющих генофонда получена информация о структуре и составе генофондов, генетических взаимоотношениях, миграциях различных коренных народов Сибири, что дополняет сложную картину истории генетико-демографических процессов.

Проведено масштабное популяционное исследование генофонда коренных народов Сибири и сопредельных регионов на уровне анализа протяженных аутосомных гаплотипов, маркирующих различные генетические пласты генофонда популяций. Полученные результаты свидетельствуют о высоком уровне генетической дифференциации населения Северной Евразии. Популяции Сибири выделяются на фоне других евразийских этносов значительно более высоким уровнем межэтнической и межрегиональной генетической дифференциации. Подробно охарактеризован состав генетических компонент на уровне популяций и отдельных образцов, который демонстрирует достоверную взаимосвязь с антропологической и языковой принадлежностью исследованных выборок, а также с информацией о их этногенезе.

Анализ с помощью метода главных компонент (РСА) позволил выявить интегральные характеристики суммарного генофонда. Наблюдается значимая корреляция картины распределения образцов с их географической локализацией. Распределение отдельных популяций на графике хорошо согласуется с их национальной принадлежностью, территориальной и этнической близостью исследованных выборок.

Генетические расстояния между популяциями Сибири и Восточной Европы значительно больше по сравнению с популяциями Кавказа и, в частности, Дагестана. Наиболее гетерогенными среди популяций Сибири являются томские татары, максимально близкие европейским башкирам и также имеющие минимальное значение F_{RON} . Близкое расположение киргизов с популяциями алтайцев и хакасов подтверждает Алтае-Саянские истоки их генофонда. Дальневосточные образцы разделены в полном соответствии с географической локализацией на северную группу – чукчей и коряков, и южную, включающую нивхов и удэгейцев.

Популяции Сибири демонстрируют значительные различия в распределении суммы длин и количества по всем классам RON как между исследуемыми географическими регионами, этносами, так и внутри локальных популяций. В популяциях Сибири, говорящих на чукотко-камчатских и нивхском языках, выявлен один из самых высоких уровней геномного инбридинга, по сравнению с остальными северо-евразийскими популяционными выборками. Сибирские популяции наравне с дагестанскими обладают самыми длинными геномными участками гомозиготности для всех категорий длины RON и более имеют более высокую долю людей с длинными областями аутозиготности. Эта тенденция похожа на наблюдаемое сокращение разнообразия гаплотипов с увеличением расстояния от Африки.

В популяциях коренного населения Сибири количество и суммарная длина RON выше, чем в остальных исследуемых популяциях Средней Азии, Восточной Европы и Дагестана. Количество и общий суммарный размер коротких и средних RON показывают более высокие значения для исследуемых популяций Сибири. Однако у коренного населения Дагестана, несмотря на меньшее количество коротких RON , существенный вклад вносят более протяженные регионы высокой гомозиготности, что свидетельствует о сильном инбридинге. Среди коренного населения Сибири максимальные количество и длина RON наблюдается у чукчей, коряков, нивхов и удэгейцев. Минимальным значением как среди сибирских, так

и среди остальных исследуемых популяций, за исключением узбеков, обладают томские татары, что отражает их гетерогенное происхождение.

При характеристике региональных особенностей показано, что коренное население различных районов Сибири значительно различается по степени гомозиготности их индивидуальных геномов: наименьшие значения характерны для этносов Южной Сибири, а наибольшие для популяций Дальнего Востока. Миграционные процессы в этом регионе протекали гораздо менее интенсивно, и отсутствие метисации, обусловленное территориальной разобщенностью, привело к образованию специфичных особенностей распределения регионов высокой гомозиготности. Можно заключить, что совокупность данных по структуре генофонда популяций Северной Евразии свидетельствует о наличии сильных региональных различий, отражающих особенности формирования населения в условиях изоляции, отсутствия интенсивного потока генов и объединении разнородных по происхождению генетических компонент в составе генофондов отдельных этносов.

В результате исследования выявлена более детальная структура предковых компонент у коренных народов Сибири. Большинство этносов являются своеобразным пазлом из различных предковых компонентов, которые различаются по времени появления, направлениям миграций их носителей, их антропологической и языковой принадлежности. Можно выделить наиболее выраженные в популяциях Сибири генетические компоненты: монгольский, енисейский, тюркский и дальневосточный. Оценка относительного вклада различных предковых групп населения в состав разных этносов у коренных народов позволяет проанализировать структуру генофонда населения разных территорий на протяжении различных периодов его формирования.

Анализ блоков IBD демонстрирует кластеризацию популяций Кавказской, Дагестанской, Сибирской, Волго-Уральской группы и Восточной Европы, связывающую народности по месту происхождения и много мелких кластеров внутри выявленных групп, демонстрирующих общее происхождение и большую степень родства. Для популяций, населяющих территорию Сибири, характерна

уникальная генетико-демографическая история, которая отражается в распределении блоков IBD как внутри популяции, так и между ними. Анализ блоков IBD значительно дополняет картину в изучении образования и взаимодействия этносов, хотя не дает однозначных ответов в случае сложной популяционной истории. За некоторыми исключениями, общие IBD в Сибири лучше объясняются географической близостью популяций, а не их языковой принадлежностью.

Структура коренного населения Сибири и Крайнего Севера сложилась под существенным влиянием естественного отбора: в наиболее неблагоприятных условиях среды обитания в арктическом климате Крайнего Севера и резко континентальном климате Сибири и Дальнего Востока. При проведении гаплотипических тестов на сигналы естественного отбора в разных популяциях коренного населения Сибири были получены свидетельства для нескольких десятков генов. Выявлены значительные межпопуляционные различия по численности и составу генов с выявленными сигналами направленного отбора в разных популяциях.

По результатам тестов nSL, his выявляется 492 гена, расположенных в участках генома, несущих сигналы направленного отбора. Наибольшее число значимых сигналов естественного отбора выявлено в популяциях хантов, коряков и чукчей. Пять генов, имеющих сигналы направленного отбора, повторяются не более чем у 3 популяций, 28 генов – у двух, а остальные являются уникальными для отдельных популяций. По кросс-популяционному тесту chr-ehh выявлено 1129 генов, среди сибирских популяций наиболее часто встречаются сигналы отбора в генах, отвечающих за сенсорное восприятие, обоняние и выработку меланина. Это показывает, что анализ регионов, находящихся под отбором, можно использовать в качестве полезной информации для прогнозирования фенотипических и биомедицинских особенностей людей. Отмечено участие этих генов в метаболических путях длительной депрессии синаптической передачи в мозжечке (hsa04730), циркадного захвата (hsa04713) и пути морфиновой зависимости (hsa05032).

Биологические процессы, в которые включены эти гены, включают в себя процессы регуляции, метаболические процессы, ответ на стимулы. Основными молекулярными функциями являлись связывание с белками, связывание ионов, нуклеиновых кислот, активность гидролазы. Полученные результаты, видимо, свидетельствуют об адаптивных изменениях частот аллелей и генотипов в популяциях человека, вызванных направленным действием средовых факторов, связанных с климато-географическими условиями среды.

Таким образом, подробно охарактеризованный состав генетических субстратных составляющих, на различных уровнях организации – от индивидуального до этнического и регионального уровня – характеристика родственных взаимоотношений между различными этносами, субэтносами и родами в пределах одного этноса, выявленные сходства и различия в структуре состава компонент у разных народов, межэтнические взаимосвязи и межпопуляционные различия по компонентному составу, различным блокам сцепления, уровню инбридинга, позволяет по-новому взглянуть на историю формирования генофонда народов Северной Евразии.

Выполненное исследование значительно расширяет наши представления о генофонде населения Северной Евразии и вносит весомый вклад в понимание вопросов исторического развития и эволюции человечества. Кроме фундаментального значения, данное исследование имеет перспективу практического использования. Материалы работы могут быть использованы в научно-образовательном процессе при создании курсов лекций для студентов биологических, медицинских, исторических специальностей, а также для издания научно-популярной литературы. Полученные результаты исследования могут быть применены в разработке систем и стандартов генетических маркеров ДНК-идентификации, разработки референтных баз данных для ДНК-тестирования и геномной регистрации. Применение возможно в области криминалистики, судебной медицины, ДНК-тестирования, разработки услуг геномной медицины.

ВЫВОДЫ

1. Для популяций Сибири характерно большее значение суммарной длины коротких и средних RON на индивида по сравнению с другими географическими регионами. Практически для всех популяций Сибири сумма длинных RON более изменчива и значительно больше, чем для популяций Кавказа, Средней Азии, Европы и Волго-Уральского региона, за исключением Дагестана.

2. Наибольшим коэффициентом геномного инбридинга для длин $RON > 1,5$ м.п.о. обладают популяции нивхов ($F_{RON} = 0,0389$), чукчей ($F_{RON} = 0,0431$) и коряков ($F_{RON} = 0,0446$), что свидетельствует о высоком уровне кровного родства.

3. Распределение блоков IBD разной длины на межиндивидуальном и популяционном уровнях показывает различия в генетико-демографической истории отдельных региональных и этнических групп, что хорошо согласуется с анализом компонентного состава, динамикой изменения численности и коэффициентом геномного инбридинга.

4. В популяциях коренного населения Сибири выявлено значительное число генов, находящихся под направленным естественным отбором. Биоинформационный анализ показал участие большинства генов выявленных с помощью геномных тестов на протяженную гомозиготность гаплотипов в биологических процессах развития и регуляции центральной нервной системы, синтеза меланина, процессах сенсорного восприятия, включающих формирование и развитие зрения, слуха, а также генов, вовлеченных в биологические процессы развития костей и скелетной системы.

СПИСОК СОКРАЩЕНИЙ

ДНК – дезоксирибонуклеиновая кислота

мтДНК – митохондриальная ДНК

Кб – килобаза (1000 пар нуклеотидов)

м.п.о. – миллион пар оснований (1000000 пар нуклеотидов)

SNP - (single nucleotide Polymorphism) однонуклеотидный полиморфный вариант

ЕНН (Extended Haplotype Homozygosity) – тест на протяженность гомозиготных гаплотипов

PC1 (First Principal component) – первая главная компонента

PC2 (Second Principal component) – вторая главная компонента

IBD (Identical by descent) – блоки идентичные по происхождению

ROH (Runs of Homozygosity) – регионы высокой гомозиготности

PCA (Principal components analysis) – метод главных компонент

LD (Linkage disequilibrium) – неравновесное сцепление генов

ППП – предполагаемые предковые популяции

F Ped (pedigree based inbreeding coefficient) – коэффициент инбридинга на основе родословной

F_{ROH} – коэффициент геномного инбридинга

MAF (minor allele frequency) – частота минорного аллеля

t-SNE (t-distributed Stochastic Neighbor Embedding) – метод стохастического вложения соседей с t-распределением

LTP (Long-term Potentiation) – долговременная потенция

LTD (Long-term depression) – долговременная депрессия

СПИСОК ЛИТЕРАТУРЫ

1. Prahlow J.A., Cameron T., Arendt A. et al. DNA testing in homicide investigations // *Medicine, Science and the Law*. – 2017. – Vol. 57. – № 4. – P. 179–191.
2. Palamara P.F., Lencz T., Darvasi A., Pe'er I. Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History // *American Journal of Human Genetics*. – 2012. – Vol. 91. – № 5. – P. 809.
3. Browning B.L., Browning S.R. A fast, powerful method for detecting identity by descent // *American Journal of Human Genetics*. – 2011. – Vol. 88. – № 2. – P. 173–182.
4. Albrechtsen A., Moltke I., Nielsen R. Natural selection and the distribution of identity-by-descent in the human genome // *Genetics*. – 2010. – Vol. 186. – № 1. – P. 295–308.
5. Han L., Abney M. Identity by descent estimation with dense genome-wide genotype data // *Genetic Epidemiology*. – 2011. – Vol. 35. – № 6. – P. 557–567.
6. Gusev A., Palamara P.F., Aponte G., Zhuang Z., Darvasi A., Gregersen P., Pe'er I. The architecture of long-range haplotypes shared within and across populations // *Molecular Biology and Evolution*. – 2012. – Vol. 29. – № 2. – P. 473–486.
7. Hochreiter S. HapFABIA: Identification of very short segments of identity by descent characterized by rare variants in large sequencing data // *Nucleic Acids Research*. – 2013. – Vol. 41. – № 22. – P. e202.
8. Pemberton T.J., Absher D., Feldman M.W., Myers R.M., Rosenberg N.A., Li J.Z. Genomic patterns of homozygosity in worldwide human populations // *American Journal of Human Genetics*. – 2012. – Vol. 91. – № 2. – P. 275–292.
9. Ropers H.H. New perspectives for the elucidation of genetic disorders // *American Journal of Human Genetics*. – 2007. – Vol. 81. – № 2. – P. 199–207.

10. Степанов В.А., Харьков В.Н., Вагайцева К.В. et al. Сигналы направленного отбора в популяциях народов Сибири и европейской части России // Генетика. – 2019. – Т. 55. – № 10. – С. 1198–1207.
11. Hancock A.M., Alkorta-Aranburu G., Witonsky D.B., Di Rienzo A. Adaptations to new environments in humans: The role of subtle allele frequency shifts // Philosophical Transactions of the Royal Society B: Biological Sciences. – 2010. – Vol. 365. – № 1552. – P. 2459–2468.
12. Cardona A., Pagani L., Antao T. et al. Genome-wide analysis of cold adaptation in indigenous Siberian populations // PLoS ONE. – 2014. – Vol. 9. – № 5. – P. 98076.
13. Чередниченко А.А., Трифонова Е.А., Вагайцева К.В., Бочарова А.В., Варзарь А.М., Раджабов М.О., Степанов В.А. Распространенность аллелей полиморфных вариантов генов, ассоциированных с иммунозависимыми заболеваниями, в популяциях Северной Евразии // Молекулярная биология. – 2015. – Т. 49. – № 6. – С. 984–992.
14. Харьков В.Н., Степанов В.А., Медведева О.Ф., Спиридонова М.Г., Максимова Н.Р., Ноговицына А.Н., Пузырев В.П. Происхождение якутов: анализ гаплотипов Y-хромосомы // Молекулярная биология. – 2008. – Т. 42. – № 2. – С. 226–237.
15. Харьков В.Н., Хамина К.В., Медведева О.Ф., Симонова К.В., Хитринская И.Ю., Степанов В.А. Структура генофонда тувинцев по маркерам Y-хромосомы // Генетика. – 2013. – Т. 49. – № 12. – С. 1416–1425.
16. Duggan A.T., Whitten M., Wiebe V. et al. Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers // PLoS One. – 2013. – Vol. 8. – № 12.
17. Malyarchuk B.A., Derenko M., Denisova G. On the Y-chromosome haplogroup C3c classification // J Hum Genet. – 2012. – Vol. 57. – № 10. – P. 685–686.

18. Derenko M., Malyarchuk B., Denisova G., Wozniak M., Grzybowski T., Dambueva I., Zakharov I. Y-chromosome haplogroup N dispersals from south Siberia to Europe // *Journal of Human Genetics*. – 2007. – Vol. 52. – № 9. – P. 763–770.
19. Mathieson I., Lazaridis I., Rohland N. et al. Genome-wide patterns of selection in 230 ancient Eurasians // *Nature*. – 2015. – Vol. 528. – № 7583. – P. 499–503.
20. Starikovskaya E.B., Sukernik R.I., Derbeneva O.A. et al. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups // *Annals of Human Genetics*. – 2005. – Vol. 69. – № 1. – P. 67–89.
21. Pakendorf B., Novgorodov I.N., Osakovskij V.L., Danilova A.P., Protod'jakonov A.P., Stoneking M. Investigating the effects of prehistoric migrations in Siberia: Genetic variation and the origins of Yakuts // *Human Genetics*. – 2006. – Vol. 120. – № 3. – P. 334–353.
22. Tamm E., Kivisild T., Reidla M. et al. Beringian standstill and spread of native American founders // *PLoS ONE*. – 2007. – Vol. 2. – № 9.
23. Karmin M., Saag L., Vicente M. et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture // *Genome Research*. – 2015. – Vol. 25. – № 4. – P. 459–466.
24. 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 Biology*. – 2019. – Vol. 17. – № 1. – P. 3.
25. Triska P., Chekanov N., Stepanov V. et al. Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe // *BMC Genetics*. – 2017. – Vol. 18. – № S1. – P. 110.
26. Flegontov P., Changmai P., Zidkova A. et al. Genomic study of the Ket: A Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry // *Scientific Reports*. – 2016. – Vol. 6. – № 1. – P. 1–12.

27. Sikora M., Pitulko V. V., Sousa V.C. et al. The population history of northeastern Siberia since the Pleistocene // *Nature*. – 2019. – Vol. 570. – № 7760. – P. 182–188.
28. Рычков Ю.Г. Генофонд и геногеография народонаселения. СПб.: Наук. – 2000. 611 с.
29. Харьков В.Н., Зарубин А.А., Хитринская И.Ю., Раджабов М.О., Литвинов С.С., Джаубермезов М.А., Екомасова Н.В., Хуснутдинова Э.К., Степанов В.А. Тюркский генетический компонент в генофонде коренных народов Дагестана // *Медицинская генетика*. – 2020. – Т. 7. – № 216. – С. 8–10.
30. Хитринская И.Ю., Харьков В.Н., Зарубин А.А., Литвинов С.С., Екомасова Н.В., Хуснутдинова Э.К., Степанов В.А. Монгольский генетический компонент в генофонде коренных народов Сибири, Средней Азии и Восточной Европы // *Медицинская генетика*. – 2020. – Т. 7. – № 216. – С. 11–12.
31. Степанов В.А. Этногеномика населения Северной Евразии. Печатная м. – 2002. 244 с.
32. Хуснутдинова Э.К. Этногеномика // *Вавиловский журнал генетики и селекции*. – 2013. – Т. 17. – № 4–2. – С. 943–956.
33. Степанов В.А. Популяционная геномика народов России. // *Медицинская генетика*. – 2020. – Т. 19. – № 7. – С. 6–7.
34. Biagini S.A., Solé-Morata N., Matisoo-Smith E., Zalloua P., Comas D., Calafell F. People from Ibiza: an unexpected isolate in the Western Mediterranean // *European Journal of Human Genetics* 2019 27:6. – 2019. – Vol. 27. – № 6. – P. 941–951.
35. Gopalan S., Berl R.E.W., Myrick J.W. et al. Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa // *Curr Biol*. – 2022. – Vol. 32. – № 8. – P. 1852–1860.
36. Спицын В.. А. Биохимический полиморфизм человека. МГУ. – 1985. 215 с.

37. Алтухов Ю.П., Корочкин Л.И., Рычков Ю.Г. Наследственное биохимическое разнообразие в процессах эволюции и индивидуального развития // Генетика. – 1996. – Т. 32. – № 11. – С. 1450–1473.
38. Сулимова Г.Е. ДНК-маркеры в генетических исследованиях: типы маркеров, их свойства и области применения // Успехи современной биологии. – 2004. – Т. 124. – № 3. – С. 260–271.
39. Алтухов Ю.П. Генетические процессы в популяциях. ИКЦ «Акаде. – 2003. 431 с.
40. Кучер А.Н. Динамика генетико-демографической структуры и генетическое разнообразие коренных и пришлых популяций Сибирского региона. Дис. ... д. – 2001. 335 с.
41. Cavalli-Sforza L.L. The DNA revolution in population genetics // Trends Genet. – 1998. – Vol. 14. – № 2. – P. 60–65.
42. Underhill P.A., Kivisild T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations // Annu Rev Genet. – 2007. – Vol. 41. – P. 539–564.
43. Маркелов М.Л., Шипулин Г.А., Покровский В.И. Технологии биочипов – новые возможности в диагностике болезней человека // Терапевтический архив. – 2008. – Т. 4. – С. 79–85.
44. Heller M.J. DNA microarray technology: devices, systems, and applications // Annu Rev Biomed Eng. – 2002. – Vol. 4. – P. 129–153.
45. Baye T.M. Inter-chromosomal variation in the pattern of human population genetic structure // Human Genomics. – 2011. – Vol. 5. – № 4. – P. 220.
46. Auton A., Bryc K., Boyko A.R. et al. Global distribution of genomic diversity underscores rich complex history of continental human populations // Genome Research. – 2009. – Vol. 19. – № 5. – P. 795.

47. Brookes A.J. The essence of SNPs // *Gene*. – 1999. – Vol. 234. – № 2. – P. 177–186.
48. Lai E. Application of SNP technologies in medicine: lessons learned and future challenges // *Genome Res*. – 2001. – Vol. 11. – № 6. – P. 927–929.
49. Crow J.F. Spontaneous mutation as a risk factor // *Exp Clin Immunogenet*. – 1995. – Vol. 12. – № 3. – P. 121–128.
50. Menozzi P., Piazza A., Cavalli-Sforza L. Synthetic maps of human gene frequencies in Europeans // *Science* (1979). – 1978. – Vol. 201. – № 4358. – P. 786–792.
51. Patterson N., Price A.L., Reich D. Population structure and eigenanalysis // *PLoS Genetics*. – 2006. – Vol. 2. – № 12. – P. 2074–2093.
52. Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals // *Genome Research*. – 2009. – Vol. 19. – № 9. – P. 1655–1664.
53. Cavalli-Sforza L.L., Feldman M.W. The application of molecular genetic approaches to the study of human evolution // *Nature Genetics*. – 2003. – Vol. 33. – № 3S. – P. 266–275.
54. Li W., Cerise J.E., Yang Y., Han H. Application of t-SNE to human genetic data // *Journal of Bioinformatics and Computational Biology*. – 2017. – Vol. 15. – № 4.
55. Van Der Maaten L., Hinton G. Visualizing Data using t-SNE // *Journal of Machine Learning Research*. – 2008. – Vol. 9. – № 86. 2579–2605 p.
56. Li W., Cerise J.E., Yang Y., Han H. Application of t-SNE to human genetic data // *Journal of Bioinformatics and Computational Biology*. – 2017. – Vol. 15. – № 4.
57. McQuillan R., Leutenegger A.L., Abdel-Rahman R. et al. Runs of Homozygosity in European Populations // *American Journal of Human Genetics*. – 2008. – Vol. 83. – № 3. – P. 359–372.

58. Szpiech Z.A., Blant A., Pemberton T.J. GARLIC: Genomic Autozygosity Regions Likelihood-based Inference and Classification // *Bioinformatics*. – 2017. – Vol. 33. – № 13. – P. 2059–2062.
59. Lemes R.B., Nunes K., Carnavalli J.E.P., Kimura L., Mingroni-Netto R.C., Meyer D., Otto P.A. Inbreeding estimates in human populations: Applying new approaches to an admixed Brazilian isolate // *PLoS ONE*. – 2018. – Vol. 13. – № 4.
60. Ropers H.H. New perspectives for the elucidation of genetic disorders // *American Journal of Human Genetics*. – 2007. – Vol. 81. – № 2. – P. 199–207.
61. Frazer K.A., Ballinger D.G., Cox D.R. et al. A second generation human haplotype map of over 3.1 million SNPs // *Nature*. – 2007. – Vol. 449. – № 7164. – P. 851–861.
62. Kirin M., McQuillan R., Franklin C.S., Campbell H., Mckeigue P.M., Wilson J.F. Genomic runs of homozygosity record population history and consanguinity // *PLoS ONE*. – 2010. – Vol. 5. – № 11.
63. Ceballos F.C., Joshi P.K., Clark D.W., Ramsay M., Wilson J.F. Runs of homozygosity: windows into population history and trait architecture. // *Nat Rev Genet*. – 2018. – Vol. 19. – № 4. – P. 220–234.
64. Ceballos F.C., Joshi P.K., Clark D.W., Ramsay M., Wilson J.F. Runs of homozygosity: Windows into population history and trait architecture // *Nature Reviews Genetics*. – 2018. – Vol. 19. – № 4. – P. 220–234.
65. Purcell S., Neale B., Todd-Brown K. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses // *American Journal of Human Genetics*. – 2007. – Vol. 81. – № 3. – P. 559–575.
66. Meyermans R., Gorssen W., Buys N., Janssens S. How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species // *BMC Genomics*. – 2020. – Vol. 21. – № 1. – P. 1–14.

67. Ceballos F.C., Hazelhurst S., Ramsay M. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. // *BMC Genomics*. – 2018. – Vol. 19. – № 1. – P. 106.
68. Joshi P.K., Esko T., Mattsson H. et al. Directional dominance on stature and cognition in diverse human populations // *Nature*. – 2015. – Vol. 523. – № 7561. – P. 459–462.
69. Browning S.R., Browning B.L. High-Resolution Detection of Identity by Descent in Unrelated Individuals // *American Journal of Human Genetics*. – 2010. – Vol. 86. – № 4. – P. 526–539.
70. Karafet T.M., Bulayeva K.B., Bulayev O.A., Gurganova F., Omarova J., Yepiskoposyan L., Savina O. V., Veeramah K.R., Hammer M.F. Extensive genome-wide autozygosity in the population isolates of Daghestan // *European Journal of Human Genetics*. – 2015. – Vol. 23. – № 10. – P. 1405–1412.
71. 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. 1231–1236.
72. Lemes R.B., Nunes K., Meyer D., Mingroni-Netto R.C., Otto P.A. Estimation of inbreeding and substructure levels in African-derived Brazilian quilombo populations // *Human Biology*. – 2015. – Vol. 86. – № 4. – P. 276–288.
73. Wright S. Coefficients of Inbreeding and Relationship // *The American Naturalist*. – 1922. – Vol. 56. – № 645. – P. 330–338.
74. Ceballos F.C., Álvarez G. Royal dynasties as human inbreeding laboratories: The Habsburgs // *Heredity (Edinb)*. – 2013. – Vol. 111. – № 2. – P. 114–121.
75. Leutenegger A.L., Sahbatou M., Gazal S., Cann H., Génin E. Consanguinity around the world: What do the genomic data of the HGDP-CEPH diversity panel tell us? // *European Journal of Human Genetics*. – 2011. – Vol. 19. – № 5. – P. 583–587.

76. Pemberton T.J., Absher D., Feldman M.W., Myers R.M., Rosenberg N.A., Li J.Z. Genomic patterns of homozygosity in worldwide human populations // *American Journal of Human Genetics*. – 2012. – Vol. 91. – № 2. – P. 275–292.
77. Woods C.G., Cox J., Springell K. et al. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease // *American Journal of Human Genetics*. – 2006. – Vol. 78. – № 5. – P. 889–896.
78. Pitulko V. V., Nikolsky P.A., Girya E.Y., Basilyan A.E., Tumskoy V.E., Koulakov S.A., Astakhov S.N., Pavlova E.Y., Anisimov M.A. The Yana RHS Site: Humans in the Arctic before the Last Glacial Maximum // *Science* (1979). – 2004. – Vol. 303. – № 5654. – P. 52–56.
79. Aldenderfer M. Peopling the tibetan plateau: Insights from archaeology // *High Altitude Medicine and Biology*. – 2011. – Vol. 12. – № 2. – P. 141–147.
80. Adler C.J., Dobney K., Weyrich L.S. et al. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions // *Nature Genetics*. – 2013. – Vol. 45. – № 4. – P. 450–455.
81. Степанов В.А., Пузырев В.П. Популяционно-генетический анализ четырех микросателлитных локусов у населения Тувы. // *Бюлл. эксперимент. биологии и медицины*. – 2000. – Т. 129. – № 1. – С. 56–59.
82. Stepanov V.A., Lemza S. v. PvuII restriction fragment length polymorphism at the lipoprotein lipase gene in Russians from West Siberia // *Human Heredity*. – 1993. – Vol. 43. – № 6. – P. 388–390.
83. Nielsen R., Bustamante C., Clark A.G. et al. A scan for positively selected genes in the genomes of humans and chimpanzees // *PLoS Biology*. – 2005. – Vol. 3. – № 6. – P. 0976–0985.
84. Sabeti P.C., Varilly P., Fry B. et al. Genome-wide detection and characterization of positive selection in human populations // *Nature*. – 2007. – Vol. 449. – № 7164. – P. 913–918.

85. Pickrell J.K., Coop G., Novembre J. et al. Signals of recent positive selection in a worldwide sample of human populations // *Genome Research*. – 2009. – Vol. 19. – № 5. – P. 826–837.
86. Fagny M., Patin E., Enard D., Barreiro L.B., Quintana-Murci L., Laval G. Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets // *Molecular Biology and Evolution*. – 2014. – Vol. 31. – № 7. – P. 1850–1868.
87. Enard D., Messer P.W., Petrov D.A. Genome-wide signals of positive selection in human evolution // *Genome Research*. – 2014. – Vol. 24. – № 6. – P. 885–895.
88. Manel S., Perrier C., Pratlong M., Abi-Rached L., Paganini J., Pontarotti P., Aurelle D. Genomic resources and their influence on the detection of the signal of positive selection in genome scans // *Molecular Ecology*. – 2016. – Vol. 25. – № 1. – P. 170–184.
89. Jones B.L., Raga T.O., Liebert A. et al. Diversity of lactase persistence alleles in ethiopia: Signature of a soft selective sweep // *American Journal of Human Genetics*. – 2013. – Vol. 93. – № 3. – P. 538–544.
90. Enattah N.S., Jensen T.G.K., Nielsen M. et al. Independent Introduction of Two Lactase-Persistence Alleles into Human Populations Reflects Different History of Adaptation to Milk Culture // *American Journal of Human Genetics*. – 2008. – Vol. 82. – № 1. – P. 57–72.
91. Ingram C.J.E., Elamin M.F., Mulcare C.A. et al. A novel polymorphism associated with lactose tolerance in Africa: Multiple causes for lactase persistence? // *Human Genetics*. – 2007. – Vol. 120. – № 6. – P. 779–788.
92. Tishkoff S.A., Reed F.A., Ranciaro A. et al. Convergent adaptation of human lactase persistence in Africa and Europe // *Nature Genetics*. – 2007. – Vol. 39. – № 1. – P. 31–40.

93. Gerbault P., Moret C., Currat M., Sanchez-Mazas A. Impact of selection and demography on the diffusion of lactase persistence // PLoS ONE. – 2009. – Vol. 4. – № 7.
94. Morgan A.P. argyle: An R Package for Analysis of Illumina Genotyping Arrays. // G3 (Bethesda). – 2015. – Vol. 6. – № 2. – P. 281–286.
95. Wilde S., Timpson A., Kirsanow K. et al. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y // Proc Natl Acad Sci U S A. – 2014. – Vol. 111. – № 13. – P. 4832–4837.
96. Harding R.M., Healy E., Ray A.J. et al. Evidence for variable selective pressures at MC1R // American Journal of Human Genetics. – 2000. – Vol. 66. – № 4. – P. 1351–1361.
97. Myles S., Somel M., Tang K., Kelso J., Stoneking M. Identifying genes underlying skin pigmentation differences among human populations // Human Genetics. – 2007. – Vol. 120. – № 5. – P. 613–621.
98. Rocha J. The Evolutionary History of Human Skin Pigmentation // Journal of Molecular Evolution. – 2020. – Vol. 88. – № 1. – P. 77–87.
99. Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. A map of recent positive selection in the human genome // PLoS Biology. – 2006. – Vol. 4. – № 3. – P. 0446–0458.
100. Hancock A.M., Witonsky D.B., Alkorta-Aranburu G. et al. Adaptations to climate-mediated selective pressures in humans // PLoS Genetics. – 2011. – Vol. 7. – № 4. – P. 1001375.
101. Lappalainen T., Salmela E., Andersen P.M., Dahlman-Wright K., Sistonen P., Savontaus M.L., Schreiber S., Lahermo P., Kere J. Genomic landscape of positive natural selection in Northern European populations // European Journal of Human Genetics. – 2010. – Vol. 18. – № 4. – P. 471–478.

102. Cherednichenko A.A., Trifonova E.A., Vagaitseva K. V., Bocharova A. V., Stepanov V.A. Association of the genetic polymorphism of cytokines and their receptors with climate and geographic factors in human populations // *Russian Journal of Genetics*. – 2014. – Vol. 50. – № 10. – P. 1112–1116.
103. Zhang C., Bailey D.K., Awad T. et al. A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations // *Bioinformatics*. – 2006. – Vol. 22. – № 17. – P. 2122–2128.
104. Sabeti P.C., Varilly P., Fry B. et al. Genome-wide detection and characterization of positive selection in human populations // *Nature*. – 2007. – Vol. 449. – № 7164. – P. 913–918.
105. Sabeti P.C., Reich D.E., Higgins J.M. et al. Detecting recent positive selection in the human genome from haplotype structure // *Nature*. – 2002. – Vol. 419. – № 6909. – P. 832–837.
106. Szpiech Z.A., Hernandez R.D. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection // *Molecular Biology and Evolution*. – 2014. – Vol. 31. – № 10. – P. 2824–2827.
107. Purcell S., Neale B., Todd-Brown K. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses // *American Journal of Human Genetics*. – 2007. – Vol. 81. – № 3. – P. 559–575.
108. Gusev A., Lowe J.K., Stoffel M., Daly M.J., Altshuler D., Breslow J.L., Friedman J.M., Pe’Er I. Whole population, genome-wide mapping of hidden relatedness // *Genome Research*. – 2009. – Vol. 19. – № 2. – P. 318–326.
109. Thompson E.A. The IBD process along four chromosomes // *Theoretical Population Biology*. – 2008. – Vol. 73. – № 3. – P. 369–373.
110. Поспелов В.М. Географические названия мира. Топонимический словарь. 2nd ed. / ed. P.A. A. – 2002. 512 с.

111. ВПН-2010 [Electronic resource]. URL: https://gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm (accessed: 20.06.2022).
112. Основы этнографии: учебное пособие для СПО. Издательств / ed. Козьмин В.А., Бузин В.С. – 2019. 243 с.
113. Алексеев В.П. Палеоантропология Алтае-Саянского нагорья эпохи неолита и бронзы // Антропологический сборник. – 1961. – Т. 3. 107–206 с.
114. Алексеев В.П. География человеческих рас. Мысль. – 1974. 352 с.
115. Всероссийская перепись населения 2002 года [Electronic resource]. URL: <http://www.perepis2002.ru/index.html?id=17> (accessed: 20.06.2022).
116. Этнология (этнография) / ed. Козьмина В.А., Бузина В.С. – 2019. 438 с.
117. Guo Y., He J., Zhao S., Wu H., Zhong X., Sheng Q., Samuels D.C., Shyr Y., Long J. Illumina human exome genotyping array clustering and quality control // Nature Protocols 2014 9:11. – 2014. – Vol. 9. – № 11. – P. 2643–2662.
118. Browning S.R., Browning B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering // American Journal of Human Genetics. – 2007. – Vol. 81. – № 5. – P. 1084–1097.
119. Maaten L. van der. Visualizing data using t-SNE // Journal of Machine Learning Research. – 2008. – Vol. 9. – № 2605. – P. 2579–2605.
120. Browning B.L., Browning S.R. Improving the accuracy and efficiency of identity-by-descent detection in population data // Genetics. – 2013. – Vol. 194. – № 2. – P. 459–471.
121. Монгуш Б.Б. Чики: происхождение и роль в этногенезе тувинцев (по сведениям восточных письменных источников) // Омский научный вестник. – 2007. – Т. 5. – С. 15–19.

122. Rendón J.A.G., Adelaar W. Loanwords in Imbabura Quechua // *Loanwords in the world's language: A Comparative Handbook*. – 2009. – P. 944–967.
123. Харьков В.Н., Хамина К.В., Медведева О.Ф., Симонова К.В. Генофонд бурят: клинальная изменчивость и территориальная подразделенность по маркерам Y-хромосомы // *Генетика*. – 2014. – Т. 50. – № 2. – С. 203–213.
124. Nothnagel M., Lu T.T., Kayser M., Krawczak M. Genomic and geographic distribution of snpdefined runs of homozygosity in Europeans // *Human Molecular Genetics*. – 2010. – Vol. 19. – № 15. – P. 2927–2935.
125. Karafet T.M., Bulayeva K.B., Bulayev O.A., Gurganova F., Omarova J., Yepiskoposyan L., Savina O. v., Veeramah K.R., Hammer M.F. Extensive genome-wide autozygosity in the population isolates of Daghestan // *European Journal of Human Genetics*. – 2015. – Vol. 23. – № 10. – P. 1405–1412.
126. Деренко М.В., Малярчук Б.А., Возняк М., Дамбуева И.К., Доржу Ч.М., Лузина Ф.А., Ли Х.К., Мишчицка-Шливка Д., Захаров И.А. Разнообразие линий Y-хромосомы у коренного населения Южной Сибири // *Доклады РАН*. – 2006. – Т. 411. – № 2. – С. 1–5.
127. Lemes R.B., Nunes K., Carnavalli J.E.P., Kimura L., Mingroni-Netto R.C., Meyer D., Otto P.A. Inbreeding estimates in human populations: Applying new approaches to an admixed Brazilian isolate // *PLoS ONE*. – 2018. – Vol. 13. – № 4.
128. Karafet T.M., Bulayeva K.B., Bulayev O.A., Gurganova F., Omarova J., Yepiskoposyan L., Savina O. V., Veeramah K.R., Hammer M.F. Extensive genome-wide autozygosity in the population isolates of Daghestan // *European Journal of Human Genetics*. – 2015. – Vol. 23. – № 10. – P. 1405–1412.
129. Глазунова Е.О., Харьков В.Н., Раджабов М.О., Хитринская И.Ю., Степанов В.А. Генофонд коренных народов Дагестана цезской группы по маркерам Y-хромосомы // *Медицинская генетика*. – 2016. – Т. 15. – № 4. – С. 29–31.
130. Гаджиев А.Г. Антропология малых популяций Дагестана. – 1971. 368 с.

131. Zhabagin M., Balanovska E., Sabitov Z. et al. The Connection of the Genetic, Cultural and Geographic Landscapes of Transoxiana // *Scientific Reports*. – 2017. – Vol. 7. – № 1.
132. Tambets K., Yunusbayev B., Hudjashov G. et al. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations // *Genome Biology*. – 2018. – Vol. 19. – № 1. – P. 1–20.
133. Языки мира. Кавказские языки / ed. Алексеев М.Е. et al. – 1998. 480 с.
134. Схаляхо Р.А., Почешкова Э.А., Теучеж И.Э. et al. Тюрки Кавказа: сравнительный анализ генофондов по данным о Y-хромосоме // *Вестник Московского университета. Серия XXIII. Антропология*. – 2013. – Т. 2. – С. 34–48.
135. Goldman D., Enoch M.A. Genetic epidemiology of ethanol metabolic enzymes: a role for selection. // *World review of nutrition and dietetics*. – 1990. – Vol. 63. – P. 143–160.
136. Racimo F. Testing for ancient selection using cross-population allele frequency differentiation // *Genetics*. – 2016. – Vol. 202. – № 2. – P. 733–750.
137. Grossman S.R., Shylakhter I., Karlsson E.K. et al. A composite of multiple signals distinguishes causal variants in regions of positive selection // *Science* (1979). – 2010. – Vol. 327. – № 5967. – P. 883–886.
138. Dauber A., Muñoz-Calvo M.T., Barrios V. et al. Mutations in pregnancy-associated plasma protein A2 cause short stature due to low IGF-I availability // *EMBO Molecular Medicine*. – 2016. – Vol. 8. – № 4. – P. 363.
139. Tagliabracci V.S., Engel J.L., Wen J., Wiley S.E., Worby C.A., Kinch L.N., Xiao J., Grishin N. V., Dixon J.E. Secreted kinase phosphorylates extracellular proteins that regulate biomineralization // *Science* (1979). – 2012. – Vol. 336. – № 6085. – P. 1150–1153.

140. Jablonski N.G. The evolution of human skin colouration and its relevance to health in the modern world // *J R Coll Physicians Edinb.* – 2012. – Vol. 42. – № 1. – P. 58–63.
141. Chaplin G., Jablonski N.G. Vitamin D and the evolution of human depigmentation // *Am J Phys Anthropol.* – 2009. – Vol. 139. – № 4. – P. 451–461.
142. Ibarrola-Villava M., Hu H.H., Guedj M. et al. MC1R, SLC45A2 and TYR genetic variants involved in melanoma susceptibility in southern European populations: results from a meta-analysis // *Eur J Cancer.* – 2012. – Vol. 48. – № 14. – P. 2183–2191.
143. Sakurai K., Young J.E., Kefalov V.J., Khani S.C. Variation in Rhodopsin Kinase Expression Alters the Dim Flash Response Shut Off and the Light Adaptation in Rod Photoreceptors // *Investigative Ophthalmology & Visual Science.* – 2011. – Vol. 52. – № 9. – P. 6793.
144. Chun S., Fay J.C. Evidence for Hitchhiking of Deleterious Mutations within the Human Genome // *PLoS Genetics.* – 2011. – Vol. 7. – № 8. – P. 1002240.
145. Nijbroek G., Sood S., McIntosh I., Francomano C.A., Bull E., Pereira L., Ramirez F., Pyeritz R.E., Dietz H.C. Fifteen novel FBN1 mutations causing Marfan syndrome detected by heteroduplex analysis of genomic amplicons. // *American Journal of Human Genetics.* – 1995. – Vol. 57. – № 1. – P. 8.
146. Schrijver I., Liu W., Brenn T., Furthmayr H., Francke U. Cysteine Substitutions in Epidermal Growth Factor–Like Domains of Fibrillin-1: Distinct Effects on Biochemical and Clinical Phenotypes // *American Journal of Human Genetics.* – 1999. – Vol. 65. – № 4. – P. 1007.
147. Liu W.O., Oefner P.J., Qian C., Odom R.S., Francke U. Denaturing HPLC-identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders // *Genet Test.* – 1997. – Vol. 1. – № 4. – P. 237–242.

ПРИЛОЖЕНИЕ А

Таблица А.1 – Значения коэффициентов $F_{\text{РОН}}$ для трех диапазонов длин РОН , среднее суммарное количество (А, В, С) и средняя суммарная длин (а, b, с) для трех классов длин исследуемых популяций.

Популяция	Регион	Группа	«Plink»						« GARLIC»							
			$F_{\text{РОН}}$			$F_{\text{РОН}}$			NROH			SROH			Σ	
			0.5	1.5	5	0.5	1.5	5	А	В	С	а	b	с	NROH	SROH
Башкиры(Б)	В	1	0,0438	0,0122	0,0043	0,0388	0,0094	0,0035	544	189	14	95	95	31	746	221
Башкиры(П)	В	1				0,0450	0,0147	0,0065	561	188	17	95	95	45	765	235
Башкиры(С)	В	1				0,0368	0,0083	0,0034	439	157	10	69	74	24	606	167
Татары(К)	В	1				0,0334	0,0061	0,0023	575	192	9	100	97	20	776	216
Чуваши	В	1				0,0406	0,0103	0,0032	443	136	15	79	70	31	594	180
Бесермяне	В	2	0,0498	0,0169	0,0063	0,0509	0,0199	0,0105	678	202	16	118	110	55	896	283
Коми	В	2				0,0468	0,0164	0,0084	453	111	13	84	63	43	577	191
Удмурты	В	2				0,0504	0,0158	0,0047	476	143	19	89	81	46	637	216
Удмурты(В)	В	2				0,0494	0,0155	0,0037	576	197	24	107	107	52	796	266
Удмурты(Ш)	В	2				0,0514	0,0171	0,0039	574	205	28	98	105	57	806	259
Вепсы	В	3	0,0445	0,0107	0,0027	0,0458	0,0126	0,0031	432	145	18	76	75	38	594	188
Карелы	В	3				0,0432	0,0089	0,0024	453	153	16	80	78	31	622	189
Марийцы	В	4	0,0437	0,0117	0,0040	0,0496	0,0162	0,0044	450	159	23	81	83	51	632	215
Мордва(М)	В	4				0,0396	0,0082	0,0026	595	179	10	109	97	24	784	231
Мордва(Ш)	В	4				0,0460	0,0141	0,0067	635	189	13	109	98	39	837	246
Мордва(Э)	В	4				0,0396	0,0083	0,0026	596	204	16	100	99	30	816	229
Русские	В	5	0,0377	0,0061	0,0019	0,0377	0,0061	0,0019	607	189	8	111	102	19	804	231
Агулы	Д	6	0,0490	0,0193	0,0119	0,0464	0,0171	0,0104	617	163	11	114	93	46	790	253
Арчинцы	Д	6				0,0782	0,0459	0,0315	547	161	25	106	94	123	733	323
Лезгины	Д	6				0,0412	0,0124	0,0076	568	156	8	105	88	33	733	226
Рутульцы	Д	6				0,0433	0,0131	0,0072	445	131	10	78	70	34	586	182
Табасараны	Д	6				0,0350	0,0066	0,0028	573	196	9	101	99	20	778	220

Продолжение таблицы А.1

Популяция	Регион	Группа	«Plink»						« GARLIC»							
			F _{ROH}			F _{ROH}			N _{ROH}			S _{ROH}			Σ	
			0.5	1.5	5	0.5	1.5	5	A	B	C	a	b	c	N _{ROH}	S _{ROH}
Цахуры	Д	6				0,0499	0,0206	0,0120	463	132	14	81	70	54	609	206
Андийцы	Д	7	0,0696	0,0378	0,0240	0,0578	0,0279	0,0185	577	167	17	101	90	76	761	266
Ахвахцы	Д	7				0,1020	0,0678	0,0430	424	114	35	90	79	174	572	343
Багулалы	Д	7				0,0721	0,0400	0,0253	621	171	22	119	103	105	814	328
Ботлихцы	Д	7				0,0512	0,0196	0,0102	673	230	20	114	117	59	924	290
Каратинцы	Д	7				0,0775	0,0460	0,0321	547	164	24	101	93	122	735	316
Тиндинцы	Д	7				0,0569	0,0253	0,0139	536	164	16	95	91	66	715	253
Чамалалы	Д	7				0,0697	0,0377	0,0246	501	151	23	91	82	102	675	276
Бежтинцы	Д	8	0,1126	0,0727	0,0460	0,1105	0,0689	0,0417	423	113	28	101	97	158	563	356
Гинухцы	Д	8				0,1437	0,1039	0,0733	259	71	27	80	85	226	357	391
Гунзибцы	Д	8				0,1305	0,0887	0,0549	318	98	33	85	95	204	449	383
Хваршины	Д	8				0,0773	0,0448	0,0309	594	183	24	108	102	121	801	331
Цезы	Д	8				0,1011	0,0571	0,0293	492	141	31	109	108	137	664	354
Караногайцы	Д	9	0,0318	0,0037	0,0010	0,0324	0,0039	0,0011	582	198	13	99	94	20	793	213
Кумыки	Д	9				0,0312	0,0036	0,0009	548	183	9	100	95	16	739	210
Даргинцы	Д	10	0,0696	0,0369	0,0243	0,0696	0,0369	0,0243	594	143	19	114	87	97	756	297
Лакцы	Д	11	0,0666	0,0347	0,0230	0,0666	0,0347	0,0230	587	179	20	107	100	93	786	300
Аварцы	Д	12	0,0761	0,0442	0,0320	0,0761	0,0442	0,0320	589	178	24	109	100	119	791	328
Абхазы	К	13	0,0335	0,0044	0,0006	0,0350	0,0047	0,0004	519	224	22	86	101	28	764	215
Адыгейцы	К	13				0,0339	0,0048	0,0004	552	193	14	91	91	21	759	204
Черкесы	К	13				0,0317	0,0037	0,0010	589	173	7	106	92	13	769	211
Балкарцы	К	14	0,0345	0,0069	0,0028	0,0357	0,0071	0,0027	514	151	8	100	85	21	674	205
Карачаевцы	К	14				0,0382	0,0104	0,0055	589	172	10	106	92	30	771	228
Ногайцы	К	14				0,0295	0,0030	0,0003	528	239	21	83	102	24	788	209

Продолжение таблицы А.1

Популяция	Регион	Группа	«Plink»						« GARLIC»							
			F _{ROH}			F _{ROH}			N _{ROH}			S _{ROH}			Σ	
			0.5	1.5	5	0.5	1.5	5	A	B	C	a	b	c	N _{ROH}	S _{ROH}
Ингуши	К	15	0,0396	0,0099	0,0030	0,0392	0,0094	0,0026	521	196	16	91	98	33	733	222
Чеченцы	К	15				0,0399	0,0105	0,0034	550	177	15	99	92	33	741	223
Осетины(С)	К	16	0,0363	0,0066	0,0017	0,0359	0,0063	0,0015	515	220	20	86	101	31	755	217
Осетины(Ю)	К	16				0,0367	0,0069	0,0018	570	212	20	92	97	31	801	220
Мегрелы	К	17	0,0369	0,0049	0,0006	0,0369	0,0049	0,0006	529	202	16	93	100	23	747	216
Алтайцы(Б)	С	18	0,0533	0,0171	0,0089	0,0519	0,0173	0,0086	517	163	14	93	86	46	695	224
Алтайцы(К)	С	18				0,0552	0,0168	0,0100	632	177	13	121	101	47	821	269
Татары(Т)	С	18				0,0338	0,0071	0,0037	483	127	5	86	68	17	615	171
Тувинцы	С	18				0,0591	0,0151	0,0061	731	242	19	139	136	46	992	322
Хакасы(Т)	С	18				0,0664	0,0318	0,0183	585	173	23	111	97	87	781	295
Хакасы(Ш)	С	18				0,0534	0,0146	0,0070	652	199	12	122	112	38	862	273
Буряты(А)	С	19а	0,0656	0,0181	0,0076	0,0599	0,0112	0,0027	691	317	31	116	152	50	1040	318
Буряты(К)	С	19а				0,0615	0,0166	0,0081	733	235	16	135	130	47	983	312
Калмыки	С	19а				0,0465	0,0058	0,0008	688	260	23	121	126	33	970	280
Эвенки(З)	С	19б				0,0632	0,0195	0,0080	608	195	19	111	110	55	822	276
Эвенки(Я)	С	19б				0,0786	0,0256	0,0105	705	221	23	138	133	69	949	339
Удэгейцы	С	19б				0,0774	0,0246	0,0134	764	219	14	152	136	59	997	347
Якуты	С	19в				0,0723	0,0233	0,0097	725	240	25	137	137	68	990	343
Кеты	С	20а	0,0707	0,0274	0,0127	0,0720	0,0259	0,0121	753	221	21	139	132	70	995	341
Ханты(К)	С	20б				0,0716	0,0280	0,0119	620	201	24	123	125	75	844	323
Ханты(Р)	С	20б				0,0716	0,0266	0,0102	633	212	24	120	126	70	869	316
Чулымцы	С	20в				0,0676	0,0292	0,0167	541	165	22	99	92	79	727	270
Нивхи	С	21а	0,1139	0,0398	0,0133	0,1073	0,0363	0,0119	699	239	28	147	161	84	965	393
Коряки	С	21б				0,1209	0,0422	0,0142	806	301	30	172	210	98	1136	480

Продолжение таблицы А.1

Популяция	Регион	Группа	«Plink»						« GARLIC»							
			F _{ROH}			F _{ROH}			N _{ROH}			S _{ROH}			Σ	
			0.5	1.5	5	0.5	1.5	5	A	B	C	a	b	c	N _{ROH}	S _{ROH}
Чукчи	С	216				0,1136	0,0409	0,0137	703	257	32	149	174	100	992	424
Казахи(М)	С. А.	22а	0,0367	0,0055	0,0018	0,0360	0,0040	0,0006	575	198	13	100	98	19	786	217
Казахи(С)	С. А.	22а				0,0358	0,0039	0,0002	511	232	25	83	101	28	768	212
Киргизы	С. А.	22а				0,0414	0,0067	0,0022	589	194	8	106	103	19	791	227
Узбеки	С. А.	22а				0,0334	0,0073	0,0043	549	170	6	97	88	20	725	205
Дунгане	С. А.	22б	0,0562	0,0116	0,0070	0,0562	0,0116	0,0070	808	218	6	151	126	27	1032	304
	В		0,0438	0,0122	0,0043				535	173	16	94	90	37	723	221
	Д		0,0704	0,0377	0,0239				525	156	20	100	93	96	701	289
	К		0,0357	0,0065	0,0018				543	196	15	94	96	26	755	215
	С		0,0702	0,0229	0,0099				663	218	21	127	127	62	902	316
	С. А.		0,0406	0,0067	0,0028				606	202	12	107	103	23	820	233

Примечание1: В – Волго-уральский регион и Европа, К – Кавказ, Д – Дагестан, С – Сибирь, С.А. – Средняя Азия.

Примечание2: 1 – алтайская/тюркская; 2 – пермская подгруппа; 3 – прибалтийско-финская подгруппа; 4 – волжско-финская подгруппа; 5 – восточнославянская группа; 6 – нахско-дагестанская/лезгинская ветвь; 7 – нахско-дагестанская/андийская группа; 8 – нахско-дагестанская/дидойская группа; 9 – алтайская/тюркская; 10 – нахско-дагестанская/даргинская ветвь; 11 – нахско-дагестанская/лакская ветвь; 12 – нахско-дагестанская/аваро–андийская группа; 13 – абхазо-адыгские языки; 14 – алтайская/тюркская; 15 – нахско-дагестанская/нахская ветвь; 16 – индоевропейская/иранская группа; 17 – картвельская семья; 18 – алтайская/тюркская; 19 – а) алтайская/монгольская, б) алтайская/тунгусо-маньчжурские языки, в) алтайская/тюркская; 20 – а) енисейская, б) уральская/обско-угорская, в) алтайская/тюркская; 21 – а) палеоазиатский язык, в) чукотско-камчатские языки; 22 – а) алтайская/тюркская, б) сино-тибетские языки/китайские языки.

Таблица А.2 – «Горячие точки» RОН определённые в исследуемых популяциях.

№	хромосома	гены	p1(kb)	p2(kb)	kb	минимум	1st Qu.	медиана	среднее	3rd Qu.	максимум	В предыдущих исследованиях
1	2	-	623,216	651,38	28,165	0,385	0,607	0,682	0,635	0,687	0,688	
2	15	<i>ADAMTSL3</i>	84557,68	84589,61	31,938	0,389	0,543	0,564	0,538	0,565	0,565	2015
3	1	<i>F5</i>	169532,6	169545,6	12,955	0,386	0,509	0,535	0,511	0,537	0,537	
4	8	<i>LPL</i>	19820,43	19871,11	50,681	0,386	0,414	0,525	0,508	0,580	0,583	2015
5	1	<i>TMEM50A, RHCE, MACO1</i>	25679,39	25859,02	179,632	0,391	0,487	0,520	0,506	0,540	0,540	
6	1	<i>ANGPTL3, DOCK7</i>	62960,25	63194,01	233,76	0,401	0,485	0,516	0,504	0,536	0,542	2015
7	22	<i>LINC02558, SLC5A1</i>	32381,27	32445,95	64,675	0,404	0,468	0,496	0,483	0,503	0,508	
8	20	<i>CST9, CST3</i>	23578,21	23622,55	44,345	0,400	0,487	0,495	0,491	0,506	0,507	
9	18	-	57736,27	57905,22	168,947	0,391	0,418	0,493	0,465	0,506	0,514	
10	3	-	86760,1	86946,46	186,356	0,391	0,457	0,490	0,480	0,508	0,539	
11	6	-	126752,9	127036,2	283,291	0,388	0,444	0,489	0,477	0,510	0,520	
12	6	<i>HLA-DPA1, HLA-DPB1</i>	33031,68	33056,96	25,286	0,408	0,454	0,487	0,498	0,545	0,586	2015
13	6	<i>CDKAL1</i>	20649,25	20682,62	33,369	0,385	0,437	0,480	0,463	0,485	0,486	2015
14	10	-	44745,93	44774,09	28,162	0,388	0,466	0,478	0,464	0,480	0,480	
15	3	<i>LINC01322</i>	165000,9	165048,2	47,306	0,395	0,446	0,476	0,465	0,487	0,497	
16	11	<i>OR10AG1</i>	55666,55	55737,07	70,521	0,388	0,414	0,466	0,467	0,520	0,524	
17	4	<i>LCORL</i>	17948,67	18028,99	80,32	0,388	0,438	0,466	0,453	0,469	0,475	2015
18	20	<i>UQCC1, GDF5OS, GDF5</i>	33903,35	34022,67	119,319	0,386	0,449	0,464	0,456	0,473	0,475	2015

Продолжение таблицы А.2

№	хромосома	гены	p1(kb)	p2(kb)	kb	минимум	1st Qu.	медиана	среднее	3rd Qu.	максимум	В предыдущих исследованиях
19	3	-	175673,3	175806,4	133,16	0,386	0,439	0,460	0,461	0,485	0,533	
20	3	-	147684,6	147704	19,488	0,386	0,432	0,459	0,446	0,464	0,465	
21	15	<i>LARP6, LRRC49, THAP10</i>	71110,31	71187,78	77,48	0,387	0,426	0,458	0,471	0,522	0,532	
22	1	<i>LRRC7</i>	70303,1	70404,16	101,067	0,386	0,435	0,455	0,447	0,464	0,467	
23	4	-	111654,1	111714,4	60,296	0,385	0,395	0,455	0,441	0,483	0,486	
24	4	<i>CWH43</i>	48950,59	49051,77	101,177	0,388	0,439	0,451	0,440	0,452	0,459	
25	3	-	176057,9	176077,5	19,635	0,386	0,415	0,447	0,440	0,467	0,478	
26	4	-	156477,4	156509,6	32,196	0,391	0,422	0,431	0,427	0,434	0,435	
27	1	<i>LEPR</i>	65998,47	66159,88	161,411	0,385	0,423	0,431	0,426	0,436	0,440	2015
28	9	<i>CDKN2B-AS1</i>	22023,81	22053,71	29,905	0,385	0,411	0,424	0,417	0,425	0,425	2015
29	4	<i>SLC10A4, ZAR1, FRYL</i>	48444,37	48618,43	174,066	0,385	0,399	0,423	0,440	0,489	0,502	
30	6	<i>OR2B6, OR1F12, ZNF165, ZSCAN12P1, ZSCAN16-AS1, ZSCAN16, ZKSCAN8, ZNF192P1, ZSCAN9, ZKSCAN4, NKAPL, ZSCAN26, - PGBD1, ZKSCAN3, ZSCAN31, ZSCAN12, ZSCAN23</i>	27904,18	28413,49	509,311	0,385	0,413	0,422	0,421	0,426	0,461	2010, 2012, 2015, 2019*

Продолжение таблицы А.2

№	хромосома	гены	p1(kb)	p2(kb)	kb	минимум	1st Qu.	медиана	среднее	3rd Qu.	максимум	В предыдущих исследованиях
31	3	<i>LINC02046</i>	147822,1	148090,6	268,524	0,385	0,401	0,413	0,413	0,424	0,450	
32	4	-	82144,93	82279,67	134,741	0,386	0,392	0,412	0,414	0,435	0,446	
33	6	<i>ADGRG6</i>	142694	142780,3	86,203	0,387	0,401	0,412	0,411	0,422	0,428	2015
34	17	<i>PSMD3, CSF3, MED24</i>	38137,62	38183,09	45,468	0,385	0,406	0,412	0,408	0,414	0,417	
35	6	<i>HLA-DQB2</i>	32686,2	32761,83	75,631	0,385	0,404	0,411	0,425	0,412	0,530	2015
36	1	<i>FPGT-TNNI3K, TNNI3K</i>	74991,6	75006,08	14,481	0,389	0,404	0,411	0,408	0,414	0,415	
37	1	<i>NOS1AP</i>	162031	162109,1	78,13	0,385	0,403	0,408	0,413	0,424	0,440	
38	16	<i>FTO</i>	53814,36	53827,96	13,6	0,387	0,400	0,408	0,409	0,421	0,422	
39	6	<i>HCG17, HLA-L</i>	30230,09	30240,07	9,99	0,385	0,405	0,407	0,405	0,408	0,410	
40	7	<i>CCDC146</i>	76914,47	76920,56	6,087	0,385	0,393	0,405	0,402	0,410	0,416	
41	13	<i>LINC00402</i>	74786,64	74816,16	29,517	0,385	0,393	0,404	0,402	0,410	0,413	
42	1	<i>LGALS8, HEATR1</i>	236702,6	236758,9	56,293	0,385	0,398	0,404	0,401	0,405	0,406	2015
43	20	<i>LINC00261</i>	22527,66	22548,29	20,64	0,385	0,395	0,401	0,400	0,405	0,411	
44	2	<i>CARF, WDR12</i>	203758,5	203830,2	71,708	0,388	0,397	0,400	0,400	0,407	0,407	2015
45	2	<i>CCDC85A</i>	56542,47	56564,3	21,835	0,385	0,387	0,398	0,398	0,401	0,430	
46	7	-	15051,64	15064,26	12,613	0,387	0,395	0,398	0,397	0,399	0,400	
47	6	<i>RSPO3</i>	127455,1	127495,2	40,108	0,385	0,388	0,396	0,395	0,404	0,404	2015
48	9	<i>ABCA1</i>	107646,2	107661,1	14,906	0,385	0,392	0,395	0,395	0,398	0,400	2019*
49	2	-	21286,06	21313,58	27,521	0,385	0,393	0,393	0,394	0,397	0,401	

Продолжение таблицы А.2

№	хромосома	гены	p1(kb)	p2(kb)	kb	минимум	1st Qu.	медиана	среднее	3rd Qu.	максимум	В предыдущих исследованиях
50	6	<i>OR2J2</i>	29135,84	29159,01	23,172	0,386	0,389	0,392	0,390	0,392	0,393	2010, 2015
51	6	<i>HCG14, TRIM27, LINC01556</i>	28854,27	28920,58	66,307	0,385	0,389	0,390	0,390	0,393	0,394	2010, 2015
52	16	<i>DUS2</i>	68074,42	68099,26	24,844	0,386	0,388	0,389	0,390	0,391	0,395	
53	6	<i>TSBP1-AS1, TSBP1</i>	32309,32	32316,6	7,282	0,385	0,388	0,389	0,388	0,389	0,389	2015
54	6	<i>STXBP5</i>	147620,3	147649	28,717	0,385	0,387	0,387	0,387	0,387	0,387	
55	2	<i>IRS1</i>	227649,6	227652	2,415	0,385	0,386	0,386	0,387	0,387	0,389	2015
56	2	<i>RHBDD1</i>	227765,5	227773,9	8,437	0,385	0,386	0,386	0,386	0,387	0,387	2015

Примечание: регионы отсортированы по уменьшению медианного значения.

2010 - Nothnagel M, Lu TT, Kayser M, Krawczak M: Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum Mol Genet 2010; 19: 2927-2935.

2015 - Karafet TM, Bulayeva K B, Bulayev O A, et al. Extensive genome-wide autozygosity in the - Population isolates of Daghestan. Eur J Hum Genet. 2015 Oct;23(10):1405-12.

2012 - - Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ: Genomic - Patterns of homozygosity in worldwide human - Populations. Am J Hum Genet 2012; 91: 275-292.

2019 - Степанов В. А., Харьков В. Н., Вагайцева К. В. и др., Сигналы направленного отбора в популяциях народов Сибири и европейской части России. January 2019Genetika 55(10):1198-1207

Таблица А.3 – Доля предковых компонент (K=12) в популяциях Северной Евразии.

Популяции	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
Вепсы	0,002	0,001	0,000	0,001	0,003	0,000	0,002	0,953	0,001	0,000	0,000	0,037
Карелы	0,002	0,004	0,004	0,000	0,001	0,000	0,001	0,956	0,001	0,000	0,000	0,031
Мордва (М)	0,001	0,012	0,054	0,001	0,008	0,003	0,000	0,907	0,001	0,003	0,005	0,005
Мордва (Э)	0,001	0,012	0,065	0,001	0,004	0,005	0,001	0,895	0,003	0,007	0,001	0,005
Мордва (Ш)	0,001	0,009	0,038	0,001	0,003	0,002	0,000	0,931	0,001	0,005	0,001	0,007
Русские	0,000	0,022	0,018	0,000	0,001	0,000	0,000	0,947	0,002	0,002	0,003	0,003
Коми	0,002	0,212	0,077	0,001	0,005	0,002	0,007	0,601	0,005	0,004	0,000	0,084
Чуваши	0,002	0,016	0,626	0,001	0,014	0,004	0,003	0,316	0,004	0,004	0,001	0,008
Марийцы	0,000	0,000	0,998	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,001
Татары (К)	0,002	0,070	0,153	0,001	0,005	0,011	0,003	0,631	0,004	0,008	0,099	0,011
Татары (Т)	0,006	0,061	0,096	0,006	0,007	0,042	0,018	0,403	0,075	0,014	0,243	0,030
Башкиры (П)	0,003	0,202	0,193	0,000	0,005	0,021	0,015	0,305	0,020	0,016	0,101	0,119
Башкиры (С)	0,006	0,099	0,172	0,002	0,000	0,021	0,023	0,307	0,026	0,028	0,212	0,103
Башкиры (Б)	0,005	0,109	0,128	0,003	0,002	0,029	0,020	0,297	0,021	0,022	0,244	0,121
Удмурты (Ш)	0,000	0,986	0,004	0,000	0,000	0,000	0,000	0,006	0,000	0,000	0,000	0,003
Удмурты (Б)	0,000	0,937	0,008	0,000	0,001	0,000	0,000	0,048	0,000	0,001	0,000	0,005
Удмурты	0,000	0,924	0,013	0,000	0,001	0,000	0,000	0,054	0,000	0,000	0,000	0,006
Бесермяне	0,002	0,670	0,069	0,000	0,001	0,004	0,001	0,246	0,000	0,001	0,000	0,004
Ханты (Р)	0,002	0,001	0,002	0,000	0,000	0,001	0,019	0,000	0,013	0,000	0,000	0,962
Ханты (К)	0,001	0,027	0,010	0,000	0,001	0,001	0,012	0,001	0,009	0,002	0,000	0,937
Узбеки	0,002	0,077	0,027	0,000	0,003	0,023	0,000	0,398	0,003	0,030	0,435	0,002
Казахи (М)	0,003	0,033	0,035	0,038	0,003	0,056	0,081	0,236	0,015	0,026	0,467	0,007
Казахи(С)	0,004	0,021	0,030	0,036	0,003	0,061	0,091	0,239	0,013	0,023	0,471	0,007
Киргизы	0,008	0,031	0,018	0,044	0,004	0,098	0,092	0,155	0,015	0,024	0,499	0,011
Алтайцы (Б)	0,001	0,008	0,006	0,003	0,001	0,901	0,004	0,016	0,002	0,049	0,006	0,004
Алтайцы (К)	0,004	0,003	0,003	0,031	0,002	0,859	0,043	0,005	0,004	0,003	0,042	0,001
Хакасы (Ш)	0,009	0,015	0,014	0,012	0,003	0,076	0,119	0,026	0,198	0,091	0,413	0,024

Продолжение таблицы А.2

Популяции	К=1	К=2	К=3	К=4	К=5	К=6	К=7	К=8	К=9	К=10	К=11	К=12
Хакасы (Т)	0,002	0,006	0,002	0,006	0,000	0,009	0,014	0,023	0,015	0,918	0,004	0,001
Чулымцы	0,001	0,002	0,000	0,001	0,001	0,003	0,000	0,046	0,941	0,002	0,000	0,003
Кеты	0,014	0,000	0,000	0,009	0,014	0,000	0,101	0,027	0,654	0,015	0,000	0,166
Тувинцы	0,005	0,001	0,000	0,049	0,008	0,529	0,211	0,000	0,069	0,007	0,119	0,003
Калмыки	0,006	0,009	0,035	0,085	0,001	0,079	0,163	0,045	0,011	0,017	0,541	0,006
Буряты (А)	0,010	0,000	0,000	0,097	0,005	0,018	0,309	0,000	0,015	0,003	0,539	0,004
Буряты (К)	0,006	0,001	0,002	0,055	0,006	0,001	0,270	0,009	0,010	0,001	0,630	0,008
Дунгане	0,002	0,003	0,000	0,058	0,000	0,002	0,002	0,015	0,000	0,001	0,917	0,000
Нивхи	0,000	0,000	0,000	0,966	0,000	0,000	0,003	0,032	0,000	0,000	0,000	0,000
Удэгейцы	0,005	0,000	0,003	0,569	0,001	0,002	0,206	0,028	0,003	0,000	0,182	0,002
Эвенки (З)	0,021	0,004	0,011	0,117	0,007	0,002	0,620	0,182	0,010	0,002	0,006	0,018
Эвенки (Я)	0,024	0,000	0,002	0,016	0,005	0,002	0,932	0,011	0,001	0,001	0,003	0,001
Якуты	0,002	0,001	0,001	0,003	0,000	0,001	0,935	0,044	0,001	0,003	0,009	0,000
Чукчи	0,093	0,000	0,000	0,000	0,903	0,000	0,000	0,004	0,000	0,000	0,000	0,000
Коряки	0,907	0,003	0,000	0,012	0,032	0,000	0,009	0,031	0,002	0,001	0,000	0,002

Таблица А.4 – Краткая характеристика онтологии генов (GO) в биологических процессах.

#term ID	описание термина	наблюдаемое количество генов	количество генов в категории	strength	коэффициент ложного обнаружения
GO:0007399	развитие нервной системы	93	2206	0,32	6,72E-08
GO:0007275	развитие многоклеточного организма	147	4726	0,18	1,84E-05
GO:0022008	нейрогенез	66	1519	0,33	1,84E-05
GO:0048699	образование нейронов	63	1422	0,34	1,84E-05
GO:0048731	развитие системы	133	4144	0,2	1,84E-05
GO:0032502	процесс развития	161	5401	0,16	2,39E-05
GO:0048856	развитие анатомической структуры	154	5085	0,17	2,39E-05
GO:0120035	регуляция организации проекции клеток, связанных с плазматической мембраной	35	600	0,46	3,57E-05
GO:0048468	развитие клеток	61	1493	0,3	0,00012
GO:0045664	регуляция дифференцировки нейронов	33	595	0,43	0,00019
GO:0051960	регуляция развития нервной системы	40	817	0,38	0,00022
GO:0032501	многоклеточный организменный процесс	180	6507	0,13	0,00025
GO:0007611	обучение или память	19	237	0,59	0,00031
GO:0010975	регуляция развития проекции нейронов	27	443	0,48	0,00031
GO:0030182	дифференциация нейронов	43	940	0,35	0,00034
GO:0050890	познание	20	274	0,55	0,00054
GO:0060284	регуляция развития клеток	39	846	0,35	0,00082
GO:0060348	развитие костей	15	166	0,65	0,00085
GO:0048869	клеточный процесс развития	109	3533	0,18	0,00091
GO:0030154	дифференциация клеток	107	3457	0,18	0,00096
GO:0050767	регуляция нейрогенеза	35	730	0,37	0,00097
GO:0048666	развитие нейрона	35	758	0,35	0,002
GO:0007423	развитие органов чувств	27	515	0,41	0,0025
GO:0008038	распознавание нейронов	7	34	1	0,0026
GO:0051962	положительная регуляция развития нервной системы	26	488	0,42	0,0026

Продолжение таблицы А.4

#term ID	описание термина	наблюдаемое количество генов	количество генов в категории	strength	коэффициент ложного обнаружения
GO:0021650	формирование вестибулокохлеарного нерва	4	5	1,59	0,0029
GO:0001654	развитие глаз	20	339	0,46	0,0059
GO:0009653	морфогенез анатомической структуры	67	1992	0,22	0,0059
GO:0021602	морфогенез черепных нервов	6	27	1,04	0,0059
GO:0007420	развитие мозга	30	650	0,35	0,0061
GO:0010769	регуляция клеточного морфогенеза, участвующего в дифференцировке	17	263	0,5	0,0061
GO:0045595	регулирование дифференцировки клеток	59	1695	0,23	0,0061
GO:0071625	вокализационное поведение	5	16	1,19	0,0061
GO:0043010	развитие глаз по типу камеры	18	292	0,48	0,0064
GO:0019932	передача сигналов, опосредованная вторым мессенджером	17	266	0,5	0,0066
GO:0022604	регуляция морфогенеза клеток	23	442	0,41	0,0073
GO:0065007	биологическая регуляция	277	11740	0,06	0,0073
GO:0048592	морфогенез глаза	12	145	0,61	0,0076
GO:2000026	регуляция развития многоклеточного организма	63	1876	0,22	0,0076
GO:0007610	поведение	26	541	0,37	0,0077
GO:0021545	развитие черепных нервов	7	46	0,87	0,0078
GO:0031175	развитие проекции нейрона	28	616	0,35	0,0096
GO:0051094	положительная регуляция процесса развития	47	1286	0,25	0,0096
GO:0051963	регулирование сборки синапсов	9	86	0,71	0,0102
GO:0007417	развитие центральной нервной системы	35	861	0,3	0,0108
GO:0048593	морфогенез глаза камерного типа	10	109	0,65	0,0113
GO:0021554	развитие зрительного нерва	4	11	1,25	0,0142
GO:0007155	клеточная адгезия	34	843	0,3	0,0144
GO:0021675	развитие нервов	8	72	0,74	0,0149
GO:0008037	распознавание клеток	11	138	0,59	0,0156

Продолжение таблицы А.4

#term ID	описание термина	наблюдаемое количество генов	количество генов в категории	strength	коэффициент ложного обнаружения
GO:0050789	регулирование биологического процесса	262	11116	0,06	0,0156
GO:0098609	межклеточная адгезия	21	416	0,39	0,0156
GO:0050808	организация синапсов	13	189	0,53	0,0157
GO:0061387	регулирование степени роста клеток	9	96	0,66	0,0177
GO:0050769	положительная регуляция нейрогенеза	21	425	0,38	0,0186
GO:0050772	положительная регуляция аксоногенеза	8	76	0,71	0,0186
GO:0050793	регулирование процесса развития	74	2416	0,18	0,0186
GO:0051239	регуляция многоклеточных процессов в организме	83	2788	0,16	0,0186
GO:0010720	положительная регуляция развития клеток	23	491	0,36	0,0188
GO:0030029	актиновый филаментный процесс	23	493	0,36	0,0189
GO:0045666	положительная регуляция дифференцировки нейронов	18	337	0,42	0,0189
GO:0048513	развитие органов животных	86	2926	0,16	0,0189
GO:0048646	формирование анатомической структуры, участвующей в морфогенезе	33	831	0,29	0,0189
GO:0071695	созревание анатомической структуры	10	121	0,61	0,0189
GO:0050794	регуляция клеточного процесса	248	10484	0,06	0,0221
GO:0072359	развитие системы кровообращения	32	807	0,29	0,0221
GO:0030516	регуляция удлинения аксона	8	82	0,68	0,0251
GO:0048812	морфогенез проекции нейронов	21	448	0,36	0,0298
GO:0030036	организация актинового цитоскелета	20	418	0,37	0,0311
GO:0045860	положительная регуляция активности протеинкиназы	23	517	0,34	0,0319
GO:0051965	положительная регуляция сборки синапсов	7	65	0,72	0,0319

Продолжение таблицы А.4

#term ID	описание термина	наблюдаемое количество генов	количество генов в категории	strength	коэффициент ложного обнаружения
GO:0021631	морфогенез зрительного нерва	3	6	1,39	0,0321
GO:0008361	регулирование размера клеток	11	159	0,53	0,0326
GO:0033674	положительная регуляция киназной активности	24	553	0,33	0,0326
GO:0001501	развитие скелетной системы	21	457	0,35	0,0341
GO:0031290	ведение аксонов ганглиозных клеток сетчатки	4	17	1,06	0,035
GO:0048522	положительная регуляция клеточного процесса	129	4898	0,11	0,035
GO:0050807	регулирование организации синапсов	11	161	0,53	0,035
GO:0050770	регуляция аксоногенеза	11	162	0,52	0,0355
GO:0035556	передача внутриклеточного сигнала	50	1528	0,21	0,0377
GO:0019935	передача сигналов, опосредованная циклическими нуклеотидами	9	115	0,58	0,0398
GO:0051128	регуляция организации клеточного компонента	69	2306	0,17	0,0398
GO:0090596	морфогенез сенсорного органа	14	248	0,44	0,0398
GO:0097094	морфогенез черепно-лицевого шва	4	18	1,04	0,0398
GO:0120036	организация проекции клеток, ограниченной плазматической мембраной	37	1034	0,24	0,0398
GO:0061564	развитие аксонов	18	377	0,37	0,0466

Таблица А.5 – Результаты поиска сигналов направленного отбора с помощью теста *xp-ehh*

Популяции	Гены с сигналами направленного отбора
Аварцы	ALDH3B1, ALDOA, LOC112694756, AOPEP, APBA2, APOL1, ARHGAP18, ARHGAP8, - PRR5-ARHGAP8, ARID4B, ATP2C1, ATP4B, ATP8A2, BACE2, BACH2, BBX, BDH1, BLVRB, BNC2, BRWD1, C16orf92, TLCD3B, C19orf47, C5orf64, LOC101928651, C8orf44-SGK3, CHD2, CHKA, CNTN6, COLEC11, COPS3, CSMD2, CTSH, CTXN2, CUEDC1, DAB1, DCP2, DGLUCY, DISP1, DOC2A, DOK7, DPP6, DUT, EEA1, EFCAB6, EFCAB6-AS1, EIF3D, ELMO1, EPB41, EPB41L4A, EXOC6, EZR, FAM71D, FAM81A, FANCA, FBN1, FCER2, FSTL1, GALNT5, GCC2, GCC2-AS1, GPHN, GPM6A, GRIK2, GRK1, GRM5, HADH, HECW2, HERC1, HIRIP3, HMX1, HSD17B12, IGSF21, INO80E, ISX-AS1, ITFG1, KIF3C, KLK13, LEF1, LIMS1, LINC00565, LINC00583, LINC00649, LINC00942, LINC01088, LINC01435, LINC01539, LOC642484, LINC01905, LINC02012, LINC02147, LINC02453, LOC100506403, LOC101926963, LOC107984507, LOC401478, LOC613266, LRRK1, LTBP4, MC1R, MCC, MICU2, MIPEPP3, MPPED1, MRC1, MYEF2, NDUFS8, NLRP11, NPAS3, NPHP4, NTF3, OR56A1, - PAPP2, - PAX2, - PHKB, - PRDM11, - PREX1, - PRKAR1B, - PRKG1, - PRX, - PSMG1, - PTENP1, - PTENP1-AS, - PTK2, RBFOX3, RNASEH1-AS1, RPS7, SAMD12, SCARB1, SEC23B, SEPHS2, SERTAD1, SERTAD3, SHKBP1, SKA3, SLC12A1, SLC16A11, SLC16A13, SLC24A5, SLC25A21, SLC2A9, SLC44A5, SLC45A2, SLC6A13, SMOC1, SPIRE2, SPTBN4, SRSF4, SULT6B1, TAF4B, TAF7, TAOK2, TMEM219, TCF7L1, TCIRG1, TENM2, TMEM255B, TMPRSS11E, TNKS2, TWIST2, USP48, VWC2, VWF, ZDHHC20,
Даргинцы	ACOX3, ACSL1, ADAMTSL2, AFG1L, ALDH3B1, ALDOA, LOC112694756, AMBRA1, ANO2, APOL1, ARHGAP8, - PRR5-ARHGAP8, ARID4B, ART3, ASTE1, ATAD3A, ATP2C1, BIK, BPIFB3, BPIFB4, BRCA1, C11orf80, C16orf92, TLCD3B, CHD2, CHKA, COLEC11, CSNK1G2, CSNK1G2-AS1, CTNNAL1, CTSH, CTXN2, CUEDC1, D2HGDH, DAB1, DACH1, DCAF4L1, DENND1A, DGKZ, DOC2A, DPP6, DUT, EFNA5, ELF1, EPN1, EXOC6, EZR, FAM163B, FAM81A, FANCA, FBN1, FLRT1, MACROD1, FSTL1, GALNT9, GGPS1, GLYR1, HERC1, HIRIP3, INO80E, INSC, KCNS3, KLHDC4, LINC00452, LINC00539, MIPEPP3, LINC00565, LINC01229, LINC01305, LINC02453, LOC100506403, LOC101927100, LOC101928048, LOC101928764, LOC105374428, LOC401478, LOC728095, LRRK1, LTBP4, MACROD2, MAP1S, MCAT, MICU2, MIR1246, MIR4300HG, MPPED1, MYEF2, NAP1L4, NDST4, NDUFS8, NOX3, NPAS3, NPHP4, NTF3, NTM, OR56A1, OSMR-AS1, - PAX2, - PC, - PCDH15, - PCMTD2, - PCNX1, - PLD3, - PLEKHF2, - PPP6R3, - PRKG1, - PRSS1, - PRX, - PTK2, - PTPN2, RABEP1, RAD54L2, RAPGEF2, RASA3, RBFOX3, RNASEH1, RNASEH1-AS1, RPS7, SDAD1, SEC23B, SERTAD1, SLC12A1, SLC24A5, SLC2A9, SLC45A2, SLIT3, SMARCA2, SP9, SPATA5, SPIRE2, SPTBN4, TAOK2, TMEM219, TBCE, TCIRG1, TICAM2, TMED7-TICAM2, TMED7, TMEM245, TNKS2, TNKS2-AS1, TOPAZ1, TTLL1, TWIST2, USP20, USP3, VWC2, ZDHHC20, ZKSCAN7, ZKSCAN7-AS1, ZNF197, ZNF660-ZNF197, ZNF276, ZNF35, ZNF501, ZNF502, ZNF660, ZBPB,
Дунгане	ACAD10, ACADVL, ACOX3, ACSL1, ACTN3, ADAMTSL1, ADAT1, ADGRL3, ADGRL3-AS1, AGBL1-AS1, AHCTF1, AJAP1, ALDH2, ALDH3B1, ANGPTL3, DOCK7, ANK3, ANKRD11, ANKRD13D, ANKRD31, ANKS1B, ANO2, ANXA10, APOL1, APP, APTR, ARFGEF1, ARHGAP42, ARHGAP8, - PRR5-ARHGAP8, ARHGEF11, LRRC71, ARID4B, ATP4B, ATP6V0A2, ATP6V1H, ATXN1, AXIN1, B3GALNT2, TBCE, B4GAT1, BBS1, DPP3, BBX, BIK, BRMS1, BTAF1, C11orf24,

Продолжение таблицы А.5

Популяции	Гены с сигналами направленного отбора
	<i>C11orf49, C11orf80, C14orf39, C3orf14, C4orf50, C5orf64-AS1, CACNA1I, CACNA2D3, CADM2, CADPS, CAMSAP2, CARNS1, CCDC169, CCDC169-SOHLH2, CCDC87, CCS, CD68, CDH13, CDK15, CDYL, CEP192, CHKA, CHRNA1, CLCF1, CNOT1, SETD6, COG6, COL6A5, COPS5, CORO1B, CPEB3, CPT1A, CRAMP1, CREB5, CSGALNACT1, CSMD1, CSNK1G1, CSNK2B, CSPP1, CST7, CTBP2, CTSF, CTXN2, CUBN, CXCR2, DACT1, DAPK2, DDX47, DENND1A, DENND4A, DLG2, DLG4, DNAH10, DOCK2, DPF3, DPP8, DTD1, DUSP8, DVL2, DZANK1, ECT2, EFCAB6, EFCAB6-AS1, EIF2B1, ELL2, EPB41L3, ERG, EXOC6, FADS2, FADS3, FAM120A, FAM20C, FAM234A, FAM81A, FBLN1, FBXO3, FBXO3-DT, FGD6, FGF11, FXR2, GAL, GALK2, GALNT16, GALNT17, GALNT9, GAS6-AS1, GCC2, GGPS1, GIT2, GNAL, GRAMD1B, GRIA1, GRK1, GRM3, HACD3, HELZ, HERPUD2, HRAT92, ICOSLG, LOC102723996, IFT81, IGFLR1, IGHMBP2, IGSF21, IL16, INAVA, INPP4B, INPP5A, INTS14, IPCEF1, ISX-AS1, ITGB3, ITPR1, JPT2, JUN, LINC01135, KARS, KBTBD12, KCNMA1, KCNS3, KDM2A, KDM5B, KIF14, KIFC1, KLC3, KLF12, KLHDC4, KLHL26, KMT5B, KRTAP3-3, KRTAP5-3, KRTAP5-4, KRTAP5-5, KRTAP5-AS1, LCLAT1, LEPR, LINC00189, LINC00452, LINC00470, LINC00565, LINC00635, LINC00944, LINC01250, LINC01256, LINC01378, LINC01500, LINC01934, LINC02101, LINC02209, LINC02420, LINC02538, LINGO2, LOC100130987, LOC101926963, LOC101927359, LOC101928069, LOC101928135, LOC101928174, LOC101928937, LOC101928942, LOC101929294, LOC101929526, LOC102724064, LOC105370473, LOC105378127, LOC285766, LOC440084, LRCH1, LRIG1, LRP1B, LRP5, LY6G5B, LYRM4-AS1, MAPK8IP3, MAPKAPK5, MARCHF5, MAVS, MCTP1, MDH1B, MEP1B, MIEF1, MIPEP, MIR1268A, MIR3945HG, MIR4697HG, MIR99AHG, MME, MOB2, MOSMO, MPDU1, MPPED1, MRC1, TMEM236, MRGPRF, MROH3P, MRPL21, MUCL1, MYEF2, MYO9A, MYOF, MYRFL, MYRIP, MZF1, NAALADL2, NCAM1, NDUFA12, NDUFS8, NLRP1, NOD2, NR2C1, NRG3, NUP210L, NUPR1, OLFM5P, OLR1, OPA1, OPA1-AS1, OR10H4, OR51B5, OR52H1, OR5AK2, OSBP16, OTOA, - PACS1, - PARD3B, - PC, - PCBP1-AS1, - PCDH15, - PCLAF, - PCNX1, - PCSK5, - PDGFA, - PDZRN4, - PEAR1, - PHC3, - PHF11, SETDB2-PHF11, - PHYHIP, - PID1, - PIP4K2A, - PKHD1, - PLA2G4A, - PLEKHH2, - POLR2A, - POLR3F, - PPP1CA, RAD9A, - PPP1R12A, - PPP1R12A-AS1, - PPP6R3, - PRKAG2, - PRKAR1B, - PRKCH, - PRKG2, - PRSS1, - PSD3, - PSMG1, - PTK2, - PTPN2, - PTPRCAP, - PTPRD, - PTPRG, - PTPRG-AS1, - PTPRZ1, - PXYP1, QDPR, RAB11FIP3, RAB39A, RAB3IL1, RABGAP1, RASA3, RBBP9, RBM14-RBM4, RBM34, RBM4, RBM4B, RHOJ, RNASE2, RNASE3, RNASEH1, RNASEH1-AS1, RNF138, RNF152, RNF2, RPS6KB2, RSPRY1, SBSN, SDR42E2, SEC23B, SEH1L, SENP3, SENP3-EIF4A1, SERGEF, SERPINE3, SERTAD4, SGCZ, SGF29, SGPL1, SIPA1L3, SLA, TG, SLC12A1, SLC24A1, SLC24A5, SLC29A2, SLC35F2, SLC6A11, SLIT2, SMIM26, SNAP25-AS1, SORCS2, SOX15, SPAG9, SPART, SPART-AS1, SPTBN2, SSBP2, ST8SIA2, SULT1A1, SULT1A2, SYN3, SYT14, TBC1D10C, TCIRG1, TENM2, TENT4A, TESMIN, TMC01, TMEM178B, TMEM255B, TMEM52B, TMEM65, TMPRSS11E, TMTC1, TNFSF12, TNFSF12-TNFSF13, TNFSF13, TNKS2, TNKS2-AS1, TOMM20, TPCN2, TPH1, TPM3, TRIM6, TRIM6-TRIM34, TRIO, TRIP4, TRMT44, TSHZ1, TTLL1, UBAP2, UBE2E3, UBE2F-SCLY, UBQLNL, UGT2B15, UGT2B17, UNC93B1, VWA3A, WDFY3, WDR88, ZBTB4, ZDHHC13, ZDHHC24, ZFH4-AS1, ZFR2, ZFYVE9, ZNF407, ZNF609, ZNF618, ZNF670-ZNF695, ZNF706,</i>
Карелы	A4GALT, ACSL1, ADGRL2, AKT2, ALDH3B1, AMBRA1, APBA2, APOL1,

Продолжение таблицы А.5

Популяции	Гены с сигналами направленного отбора
	ARHGAP26, ARHGAP8, - PRR5-ARHGAP8, ARID4B, ATP4B, ATP6V1H, ATXN2, BDH1, BEND3P3, BNC2, BTAF1, C11orf24, C11orf80, C16orf92, C16orf92, TLCD3B, C19orf47, CAMTA2, CCDC86, CDH13, CHD2, CHKA, COL6A5, CPEB3, CSMD1, CTXN2, CUX2, CYP11A1, CYP4X1, CYP4Z1, DCP2, DDX10, DEF8, DENND1A, DHX37, DOC2A, DPP6, EFCAB6, EFCAB6-AS1, EPB41L4A, EPC2, EPS15, EXOC6, EZR, FAM114A1, FAM81A, FANCA, FEM1C, FGFBP3, FLRT1, FLRT1, MACROD1, FOXP1, FSTL1, GAL, GALNT9, GCC2, GCC2-AS1, GGPS1, GPATCH1, GPR39, GRIA1, GRIK2, GRK1, HERC1, HIRIP3, HIVEP1, HMX1, IGFLR1, INCA1, INO80E, INPP4B, INSC, ISX-AS1, JAKMIP2-AS1, KIAA0586, KIF1C, KMT2B, LEKR1, LIMS1, LINC00539, MIPEPP3, LINC00565, LINC01088, LINC01500, LINC01539, LOC642484, LINC01844, LINC02012, LINC02299, LINC02679, LOC100130357, LOC100291105, LOC100506403, LOC101927100, LOC101927143, LOC101927378, LOC101928764, LOC102724579, LOC105374428, LOC107984132, LOC112694756, LOC401478, LOC649133, LRP5, LRRK1, LTBP4, MACROD1, MACROD2, MAP1S, MBP, MED16, MICU2, MIPEPP3, MIR1246, MIR422A, MKS1, MORN3, MPPED1, MSH4, MSX2P1, OR4D1, MYEF2, MYT1, NDUFS8, NFIB, NPAS3, NPHP4, NTM, OAF, OR4D1, OR56A1, ORAI1, OSTCP1, - PARP11, - PAX2, - PC, - PCDH9, - PCMTD2, - PCNX1, - PHACTR1, TBC1D7-LOC100130357, - PLEKHF2, - POU2F3, - PPP5D1, - PPP6R3, - PROSER3, - PSMG1, - PTENP1, - PTENP1-AS, - PTGDR, - PTK2, - PTPRM, R3HCC1L, RABEP1, RAD9A, RAE1, RAPGEF2, RBFOX3, RBM34, RBM4, RBM4B, RGS12, RHBDD1, RHPN2, RIT2, RNASEH1, RNASEH1-AS1, ROBO3, ROBO4, RPS7, SAMD12, SCRIN3, SDK1, SFTPA1, SH2B3, SHKBP1, SLC12A1, SLC24A5, SLC2A9, SLC45A2, SLC52A1, SP9, SPIRE2, SPTBN2, SPTBN4, STX8, TAOX2, TAOX2, TMEM219, TBC1D22A, TBCE, TCF7L1, TCF7L2, TCIRG1, TENM2, TESMIN, TEX14, THEG, TICAM2, TMED7-TICAM2, TLCD3B, TLR10, TLR6, TMED7, TMED7-TICAM2, TMEM165, TMEM255B, TNKS2, TNKS2-AS1, TOMM20, TTLL5, USP3, VPS13B, VPS37C, VWC2, WDR88, WWOX, ZBTB32, ZDHHC20, ZKSCAN7, ZKSCAN7-AS1, ZNF197, ZNF660-ZNF197, ZNF276, ZNF35, ZNF664-RFLNA, ZNF91, ZPBP,
Северные осетины	ACOX3, ACSL3, ADGRL2, ALDH3B1, AMBRA1, ANXA10, AOPEP, APOL1, ARHGAP26, ARHGAP8, - PRR5-ARHGAP8, ARID4B, ATP2C1, ATP4B, BMP7, BNC2, BPIFB3, BRWD1, BTAF1, BTBD11, C11orf24, C16orf92, TLCD3B, C5orf64, LOC101928651, C5orf64-AS1, CARS, CARS-AS1, CBFA2T2, CBWD1, CDH13, CELSR1, CHD2, CHKA, CHMP4A, TSSK4, CHODL, CNTN5, COG6, COPS3, CPEB3, CREB5, CSNK1G1, CTNND2, CTXN2, CUEDC1, CYP4X1, CYP4Z1, DGKZ, DHRS1, DOC2A, DOK7, DPP6, DUT, EFNA5, ELMO1, EPN1, EPX, ERC1, EXOC6, EZR, FAM234A, FAM81A, FBLN1, FBN1, FEM1C, FGFBP3, FHIT, FLCN, FSTL1, GAL, GAS6-AS1, GBE1, GCC2, GCNT1, GGPS1, GLRA3, GRIK2, GRK1, HECW2, HERC1, HIPK2, HIRIP3, IGHMBP2, INO80E, IPO4, IRF9, ISX-AS1, JAK1, KDM2A, KLHDC4, KLHL26, KLK1, KLK13, LARS2, LARS2-AS1, LEPR, LGR5, LINC00376, LINC00565, LINC00583, LINC01037, LINC01305, LINC01310, LINC01416, LINC01500, LINC01539, LOC642484, LINC01543, LINC01844, LINC01869, LINC02453, LOC100506403, LOC100507065, LOC101926934, LOC101926963, LOC101927100, LOC101928048, LOC101928119, LOC101928446, LOC101928551, LOC102467217, LOC112694756, LOC284930, LOC401478, LRP5, LRPAP1, LRRIQ1, MACROD2, MAML3, MGST2, MARCHF5, MED13L, MFSD1, MIB1, MICU2, MIPEPP3, MIR1246, MKS1, MRPL21, MRPL57, MYEF2, MYT1, NAP1L4, NDST4, NDUFS8, NPAS3, NPHP4, NRG3, NTM, NTRK3, - PAPP2, - PAX2, - PC, - PCDH15, -

Продолжение таблицы А.5

Популяции	Гены с сигналами направленного отбора
	PCDH9, - PCMTD2, - PDE4D, - PISD, - PLEKHF2, - PLEKHH2, - PPP3CA, - PPP6R3, - PPT1, - PRKAR1B, - PRR16, - PRSS1, - PSMG1, - PTK2, - PTPRM, RABEP1, RAI14, RASA3, RASGRF2, RAVER2, RBFOX3, RBM34, RBM4, RBM4B, RGS12, RNASE2, RNASEH1, RNASEH1-AS1, RPS6KA2, RPS7, SDK1, SENP2, SGPL1, SKA3, SLC12A1, SLC24A5, SLC25A27, SLC44A5, SLC45A2, SLC7A5, SP9, SPACA1, SPATA5, TAF4B, TAOK2, TMEM219, TCF7L1, TCIRG1, TDRD6, TENM2, TENT2, TESMIN, TICAM2, TMED7-TICAM2, TM9SF1, TMC3-AS1, TMED7, TMEM255B, TNKS2, TNKS2-AS1, TOMM20, TRDN, TRMT44, UTRN, VAV2, VPS37C, VWC2, WIPF1, ZDHHC20, ZFYVE9, ZNF12, ZNF609, ZNF805, ZPBP,
Южные осетины	ADAMTSL2, AGBL4, ALDH3B1, ANHX, APOL1, APOL2, ARHGAP42, ARHGAP8, - PRR5-ARHGAP8, ARID4B, ARNT, ATP5F1B, ATP6AP1L, B3GALNT2, TBCE, BAZ1A, BAZ2A, BPIFB3, BTAF1, C11orf24, C16orf92, TLCD3B, CASZ1, CCDC146, CDH13, CDYL, CELSR1, CERS2, CHKA, COLEC11, CPEB3, CPEB4, CPXM1, CRYZ, CTXN2, CTXND2, CUEDC1, DCUN1D2, DHX37, DOC2A, DUT, EFCAB6, EFCAB6-AS1, EFNA5, ERC1, EXOC6, FAM163B, FBN1, FGFBP3, FRG1-DT, FSTL1, GAL, GAS6, GAS6-AS1, GBE1, GCC2, GCC2-AS1, GFER, GGPS1, GMEB1, GPR65, GRIK2, HERC1, HIPK2, HIRIP3, HSD17B12, INO80E, ISX-AS1, KCNS3, KDM2A, KIAA0586, KLHDC4, KLK1, LIMS1, LINC00376, LINC00649, LINC01831, LINC01869, LINC02453, LOC100506403, LOC101928048, LOC105374428, LOC112694756, LOC401478, LOXHD1, LRP5, MACROD2, MAJIN, MCTP1, MFSD1, MIR1246, MPPED1, MRC1, TMEM236, MUC1, MUC4, MYEF2, MYLIP, MYT1, NACA, NAP1L4, NDUFS8, NKAIN2, NPAS3, NPHP4, NTF3, NTM, OPA1, OSBPL6, - PAX2, - PCDH15, - PCMTD2, - PITRM1, - PITRM1-AS1, - PPM1F, - PPP1R12A, - PPP2R2A, - PPP2R5E, - PPP6R3, - PRIM1, - PRKAR1B, - PSMG1, - PTGES3, - PTK2, - PTPRD, RABEP1, RAI14, RARB, RAVER2, RBFOX1, RBM34, RIT2, RNASEH1, RNASEH1-AS1, RNF125, RPS7, SETDB1, SGPL1, SKA3, SLC12A1, SLC24A5, SLC41A3, SLC45A2, SNX2, SPIRE2, STRN3, TAOK2, TMEM219, TCIRG1, TENM2, TMC03, TMEM108, TNKS2, TNKS2-AS1, TOMM20, TOP3B, TRIM46, TRNAU1AP, TYW3, VAV2, VWA3A, WDFY3, WDR20, WWOX, ZBTB7C, ZKSCAN7, ZKSCAN7-AS1, ZNF660, ZNF805,
Русские	A4GALT, ABCG8, ACSL1, ACTN3, ACTR2, ADAMTSL2, ADGRB3, ALDH3B1, ALDOA, LOC112694756, ALKAL1, AMBRA1, APBA2, APOL1, ARHGAP26, ARHGAP26-AS1, ARHGAP8, - PRR5-ARHGAP8, ARID4B, ATP4B, BDH1, BNC2, BPIFB3, BPIFB4, BRWD1, BTAF1, C11orf24, C11orf80, C16orf92, TLCD3B, C5orf64, LOC101928651, C5orf64-AS1, CABP4, CARNS1, CARS, CCDC138, CCDC87, CCS, CDH13, CHD2, CHKA, CLCF1, COL6A5, CORO1B, CPEB3, CRTCL1, CTSF, CTXN2, CUEDC1, CYP4X1, CYP4Z1, CYTH3, DAB1, DCC, DCP2, DGKZ, DHDDS, DOC2A, DPM2, DPP6, DUT, EFCAB6, EFCAB6-AS1, EFNA5, EPB41, EPC2, EPS15, ERC1, ERH, ERP44, EXOC6, EXOC6B, EZR, FCHO1, FGF12, FGFBP3, FLT3, FOXP1, FSTL1, GAL, GALNT17, GALNT9, GCC2, GCC2-AS1, GPATCH1, GRIA2, GRK1, HERC1, HERC4, HIPK2, HIRIP3, HRAT92, IGF2BP3, MALSU1, INO80E, INSC, KIRREL3, KLHDC4, KLHL26, LIMS1, LINC00376, LINC00381, LINC00565, LINC00942, LINC00955, LINC01258, LINC01416, LINC01435, LINC01500, LINC01539, LOC642484, LINC01844, LINC02147, LINC02299, LINC02420, LOC100130987, LOC100506403, LOC102724579, LOC107984132, LOC401478, LOC649133, LRP5, LRRK1, LTBP4, MAP1S, MARCHF5, MAST4, MCC, MECP, MFSD1, MICU2, MIPEPP3, MIR1246, MIR422A, MIR4275, MIR4300HG, MIR600HG, MIR646HG, MKS1, MPPED1,

Продолжение таблицы А.5

Популяции	Гены с сигналами направленного отбора
	<i>MSX2P1, OR4D1, MYEF2, MYT1, NAP1L4, NDUFS8, NPAS3, NRG1, NTF3, NTRK3, OAF, OPA1, OPA1-AS1, OPRD1, OSTCP1, - PC, - PCDH15, - PCDH9, - PCMTD2, - PDE4D, - PLEKHH2, - PPP1CA, RAD9A, - PPP3CB, - PPP6R3, - PRKG1, - PRPF4B, - PRX, - PSMG1, - PTK2, - PTPRCAP, - PTPRS, RABEP1, RABGAP1, RAD50, RANBP2, RBFOX3, RBM14-RBM4, RBM34, RBM4, RBM4B, RC3H2, RGS12, RHPN2, RNASEH1, RNASEH1-AS1, RPS6KB2, RPS7, SERTAD1, SERTAD3, SGPL1, SHKBP1, SLC12A1, SLC24A5, SLC39A9, SLC45A2, SPIRE2, SPTBN2, SPTBN4, SRSF4, SSR1, STRBP, STX8, SYT9, TAOK2, TMEM219, TBC1D10C, TBC1D22A, TCAIM, TCF7L1, TCIRG1, TMEM255B, TNKS2, TNKS2-AS1, TRAPPC9, UMAD1, VPS37C, VSNL1, VWC2, WARS2, ZBTB26, ZBTB6, ZDHHC20, ZKSCAN7, ZKSCAN8, ZNF12, ZNF165, ZNF192P1, ZNF264, ZNF445, ZNF609, ZNF805, ZPBP, ZSCAN12P1, ZSCAN16-AS1,</i>
Узбеки	<i>A4GALT, ABCA10, ABCB1, RUNDC3B, ACSL1, ACTN3, ACTR3B, ADAMTS9-AS2, ADAMTSL2, ADGRL3-AS1, AGBL1-AS1, AGO2, AGTPBP1, AHCTF1, ALDH1L2, ALDOA, LOC112694756, ANKRD12, ANKRD46, ANXA10, APAF1, ARHGAP8, - PRR5-ARHGAP8, ARL6IP5, ASPH, ASTN2, ATP2C1, ATP4B, ATPSCKMT, B4GAT1, BACE2, BBS1, DPP3, BBX, BRMS1, BRWD1, BTAF1, C11orf80, C15orf61, C16orf92, TLCD3B, C2orf76, C5orf66, C5orf66-AS2, C7orf50, CAMSAP2, CASP3, CCDC146, CCDC87, CCNJ, CCS, CD44, CD55, CD96, CDH13, CDK8, CDX2, CDYL, CELSR1, CENPU, CEP192, CFAP53, CHD2, CLDN16, CLVS1, COLEC11, CPEB3, CREB5, CRTCL1, CSMD1, CSMD2, CT69, LINC01010, CTIF, CTSF, CTXN2, DAB1, DCC, DCP2, DDX59, DEF8, DENND5A, DHRS3, DMRT1, DMRT3, DNAJA4, DNMT3B, DOC2A, DOCK10, DOCK2, DOCK8, DOK7, DPP6, DTD1, EDNRB, EDNRB-AS1, EFCAB6, EFCAB6-AS1, EFCAB8, EGFLAM, EGFLAM-AS4, ELMO1, ENPP6, ENTPD1-AS1, ENTR1, EOGT, EPB41, EPN1, EPS15, ERC1, ETFA, EXD3, EZR, FAM13A, FAM163B, FAM20C, FAM234A, FAM81A, FANCA, FBLN1, FER, FGFBP3, FKBPL, FLT3, FOXP1, FSTL1, GAL, GALNT9, GAREM1, GAS6, GAS6-AS1, GBE1, GEN1, GPATCH1, GPR156, GRIK2, GRK1, GSG1L, GTF2IRD1, HERPUD2, HIRIP3, HMGCS2, HMX1, HRAT92, HSPA13, IATPR, IFT43, IGHMBP2, IGSF9B, IKBP, IL27, INO80C, INO80E, INPP5E, INSC, IQCH, IQCH-AS1, ISX-AS1, ITFG1, ITGAL, ITPR1, JAKMIP2-AS1, KAT14, - PET117, KCNS3, KCTD13, KDM2A, KIAA0586, KIAA1109, KIF14, KLHDC4, KLHL26, LEPR, LINC00452, LINC00565, LINC01234, LINC01578, LINC01677, LINC01934, LINC01990, LINC02005, LINC02104, LINC02361, LINC02365, LINC02453, LINC02511, LOC100506403, LOC101926963, LOC101927421, LOC101928069, LOC101929526, LOC102724064, LOC102724467, LOC105372441, LOC105375131, LOC105376360, LOC339166, LOC401478, LRP5, LRPAP1, LRRC58, LTBP4, MACROD2, MAP2K4, MAP2K5, MAPRE1, MC1R, MC3R, MCC, MDM4, MECR, MIR3945HG, MIR4713HG, MIR548AG1, MIR600HG, MPPED1, MRC1, TMEM236, MRGPRF, MRPL21, MTAP, MTSS1, MYEF2, MYH11, MYRIP, NAA60, NDUFS8, NDUFV1, NME7, NTF3, NUP37, OR13C5, OR13C8, OR13C9, OR51B2, OR51B5, OR51B6, OR51I1, OR51I2, OR51J1, OR51M1, OR51Q1, OR52B6, OR5AK2, OR5AK4P, OSBPL6, OSMR-AS1, - P2RY1, - PAFAH2, - PAPP2, - PC, - PCBP2, - PCDH15, - PCDH9, - PCNX1, - PDE11A, - PGBD2, - PHF21B, - PHKB, - PLEKHA5, - PLUT, - PMPCA, - PMS2CL, - PPM1E, - PPM1F, - PPM1L, - PPP5D1, - PPP6R3, - PPT2-EGFL8, - PRIMPOL, - PRKAR1B, - PRR16, - PRSS1, - PSMG1, - PTK2, - PTPN2, - PTPRN2, - PTPRU, - PTPRZ1, RAB11FIP3, RABGAP1, RAD51AP2, RAD51C, TEX14, RAG1, RALGAP2, RALY, RASA3, RBBP9, RBFOX3, RBM14-RBM4, RBM23, RBM4, RBM45, RBM4B, REM2, RGS12,</i>

Продолжение таблицы А.5

Популяции	Гены с сигналами направленного отбора
	<i>RHBDD1, RHPN2, RNASEH1, RNASEH1-AS1, RNF125, RNF138, RNF180, RNF39, RNF6, RNLS, ROBO4, RPL3, RPS7, SDK1, SEC16A, SEC23B, SEH1L, SENP2, SEPHS2, SEPTIN4-AS1, SERTAD4, SERTAD4-AS1, SGF29, SGPL1, SH3BP5L, SLC12A1, SLC16A12, SLC16A12-AS1, SLC16A9, SLC24A3, SLC24A5, SLC25A3, SLC25A40, SLC29A2, SLC8A3, SLCO1B3, SLCO1B3-SLCO1B7, SMC6, SMIM26, SP1, SPART, SPART-AS1, SPATA13, SPIRE2, SPOPL, SPTBN2, SRSF4, ST7L, STK24, STRBP, STX8, SULT1A1, SULT6B1, SYNGR1, SYT14, TAF4B, TAOK2, TMEM219, TBC1D22A, TCF25, TESMIN, TGM4, TMEM207, TMEM255B, TMEM50B, TMF1, TNFSF15, TNKS2, TNKS2-AS1, TRIM24, TRIM6, TRIM6-TRIM34, TRPC3, TSPAN10, TTLL1, TUBB3, TWIST2, UBA3, UBE2U, UPF1, VAV2, VSNL1, VWC2, WASHC3, WDR61, WFDC5, ZBTB26, ZBTB6, ZDHHC24, ZDHHC3, ZFPM2, ZFYVE9, ZNF12, ZNF248, ZNF25, ZNF276, ZNF316, ZNF33A, ZNF33BP1, ZNF460-AS1, ZNF609, ZNF805, ZPBP,</i>
Вепсы	<i>ABCG8, ACSL1, ADGRB3, ADGRL2, AGBL4, AGBL4-IT1, ALDH3B1, AMBRA1, APBA2, APOL1, ARHGAP26, ARHGAP8, - PRR5-ARHGAP8, ARID4B, ARL15, ATP4B, BDH1, BMP7, BNC2, C11orf24, C16orf92, TLCD3B, C19orf47, CCDC87, CDH13, CDH17, CFAP77, CHAMP1, CHD2, CHKA, CHRN1, COL19A1, CPEB3, CREB5, CSMD1, CTXN2, CUEDC1, CYP4X1, CYP4Z1, DCC, DCP2, DGKB, DHDDS, DIAPH3, DOC2A, EFCAB6, EFCAB6-AS1, ERC1, ERH, EXD3, EXOC6, EZR, FAM102A, FAM114A1, FAM163B, FAM81A, FANCA, FEM1C, FGF11, FOXP1, GAL, GAS6-AS1, GCC2, GCC2-AS1, GPATCH1, GRIK2, GRK1, HERC1, HIRIP3, HMX1, IGFLR1, IGHMBP2, INO80E, INSC, ISX-AS1, ITGAL, JAKMIP2-AS1, KCNH7, KDM2A, KLHDC4, KLK13, KMT2B, LAMA2, LDAH, LIMS1, LIN37, LINC00381, LINC00886, LINC01088, LINC01258, LINC01416, LINC01435, LINC01500, LINC01539, LOC642484, LINC01844, LINC02012, LINC02208, LINC02299, LMBRD1, LOC100506403, LOC101926963, LOC101927100, LOC101928048, LOC101928894, LOC101929721, LOC102723899, LOC102724579, LOC112694756, LOC284930, LOC401478, LOC649133, LRP5, LURAP1L-AS1, MARCHF1, MED13L, MICU2, MIPEPP3, MIR1246, MORN3, MPPE1, MRPL21, MSH4, MSX2P1, OR4D1, MYEF2, MYLIP, NDUFS8, NELL1, NFATC2, NKAIN2, NOS1AP, NOX3, NPAS3, NTF3, OAF, OR2G2, ORAI1, - PAX2, - PAX8, - PAX8-AS1, - PCDH15, - PCDH9, - PGBD2, - PLEKHF2, - PLEKHH2, - POLR2A, - POU2F3, - PPL, - PPP3CA, - PPP5D1, - PPP6R3, - PREX1, - PRKAR1B, - PROSER3, - PTENP1, - PTGDR, - PTK2, RABEP1, RAE1, RBM14-RBM4, RBM4, RBM4B, RHPN2, RNASEH1, RNASEH1-AS1, RPS7, SAMHD1, SEPHS2, SETX, SGPL1, SH2B3, SLC12A1, SLC14A2, SLC24A5, SLC39A9, SLC44A5, SLC45A2, SPIRE2, SPO11, SPTBN2, SPTBN4, STAG1, STX8, SYT9, TAOK2, TMEM219, TBC1D22A, TCAIM, TCF7L1, TCIRG1, TESMIN, TICAM2, TMED7-TICAM2, TLDC2, TLR6, TMED7, TMEM255B, TOPAZ1, TOX2, TTF1, U2AF1L4, U2AF2, UBN1, UPK1A, USP48, VAV2, VWC2, ZBTB32, ZBTB4, ZDHHC20, ZKSCAN7, ZKSCAN7-AS1, ZNF197, ZNF660-ZNF197, ZNF276, ZNF35, ZNF419, ZNF445, ZNF501, ZNF502, ZNF549, ZNF550, ZNF660, ZNF664-RFLNA, ZNF773, ZNF852, ZPBP, ZRANB3,</i>

Таблица А.6 Результаты поиска сигналов направленного отбора с помощью теста nSL

Популяции	Гены
Башкиры (Б)	<i>LINC00682,LINGO2,RBMS3,SDK1,SOS1</i>
Башкиры (П)	<i>DNAJC5B,GABRB2,GPX6,LINC00824,NCAN,PREP,RHOJ,STARD4-AS1,ZBTB38</i>
Бесермяне	<i>ARHGAP26,CPA6,FOXK1,GABRB2,KIAA1217,LINC01194,LOC100130207,LOC101928516,LOXL2,MCC,NRXN3,OPRD1,PTPRN2,SMYD3,TP53BP1,WWOX,XIRP2,YTHDF2</i>
Чуваши	<i>CAPN14,CD226,CUX2,FAF2,FLJ22447,GABBR1,GRM5,HIST1H3I,LINC01229,LINC01266,NRXN3,RARG,RHOJ,ZSCAN31</i>
Карелы	<i>ANO3,ARHGAP10,ARHGAP26,CSMD1,GPATCH1,GUSBP2,LDHAL6A,LRP2,MIR4500HG,NFIA,NKAIN3,OR10C1,PPM1L,PREP,ROR2,RTTN,SLC14A2</i>
Коми	<i>ABLIM2,CFAP299,DGKI,DNAJC5B,ELMO1,GALNT16,LINC00682,LOC105374960,MCC,NGF-AS1,NRG3,PLXNA4,PREP,PSTPIP2,PTPRB,PTPRR,SLC22A23,SLC24A5,SUGCT,UBE2Q1</i>
Марийцы	<i>ARHGAP10,CSMD1,DPY30,ELMO1,GALNT2,GLDN,KIAA1324,LIMCH1,LINC00707,LINC01229,LINC01252,LINC01498,LINC02603,LOC101929705,LRP1B,MALRD1,MECOM,NALCN-AS1,NDST4,NREP,NRXN3,PDE1C,RANBP3L,TAP1</i>
Мордва (Э)	<i>ACTN2,ARHGAP26,COL6A5,DLG2,IRS1,LINC01727,MTTP,SLC22A23</i>
Мордва (М)	<i>ABI3BP,ANKRD44,ARHGAP10,ARPP19,CFAP77,CTNNA3,DLG2,FTO,GRM4,HLA-DMA,IL1RAP,KCNH7,LINC01622,LINC02255,LINC02268,LOC285766,LOXNRF1,LSM2,MARCHF8,MYPN,NALCN-AS1,PRMT8,PSD3,RSPO4,SAMD4A,STK32B,TVP23A</i>
Мордва (Ш)	<i>CDH13,COL6A5,FAM184B,H2AFY2,LTBP1,TBC1D9,ZNF536</i>
Татары (К)	<i>CACNB4,CUX2,DACH1,DIRC3,DLC1,DPH6,EPB41L3,IL12A-AS1,IL1RAP,JAK1,LOC101927272,LOC101927960,LOC101928622,LRIG1,LRIG3,LRP1B,MAP3K5,NDST4,NRXN3,OR10C1,PIKFYVE,PPM1L,RSPO4,SLC45A2,XIRP2</i>
Удмурты	<i>CDH16,ELMO1,HPS1,LINGO2,MALRD1,MAP3K5,MYO3B,NDUFA4,PRDM16,SLC24A5,TNFRSF21,TNRC18</i>
Удмурты (Б)	<i>AIFM2,ARHGAP26,ARL15,BACH2,BRD9,CDYL2,CHRM3,CLSTN2,DGKB,DGKZ,ELMO1,GNA14-AS1,LINC00578,LOC643542,MAN2B2,NKAIN2,PSPC1,PTPRM,RORA-AS1,SH2D4A,SPATA5,TMEM117,TMEM260,TP63</i>
Удмурты (Ш)	<i>CNBP,COL11A1,GRM4,GUSBP2,ICA1L,KCNAB1,KCNH7,LOC101927066,LOC646588,POU2F3,PREP,PRMT8,RBMS3,TMTC2,TRABD2A</i>
Вепсы	<i>ABL1,ARHGAP26,CADM1,CDH26,CFAP77,CSMD1,CTIF,DIRC3,ELMO1,EPCAM-DT,ESR1,GRIK2,ICA1,IRS1,ITGA9-AS1,KIAA0319L,LINC00702,LINC01584,LINC02241,LOC285766,LRP1B,LRR52-AS1,MAGI2,MGAT5,MTUS1,MYOM2,NFIB,PISD,PPL,RHO,SMIM28,SNTG1,ST18,STK32B,TBX19,TMTC2,UPK1B,UTRN,ZMYM4</i>
Агулы	<i>AOPEP,LINC01622,LUM,NAALADL2,ROR2,STARD13</i>

Продолжение таблицы А.6

Популяции	Гены
Ахвахцы	ADCY1,ANTXR1,APOD,CHST11,CUEDC1,FAM155A-IT1,FSTL4,GK5,GREM2,GRHL2,GRIN2B,IGF1R,KIFC3,LINC01229,LINC02147,LOC100288728,LOC100507053,LOC101927410,LOC105373805,LOC400655,LOXL1-AS1,MCTP1,MIR1268A,MTMR7,MYB,NFIA,PALD1,PARD3B,PATJ,PAX5,PDZD2,PFKFB3,PPP2R2C,RXYLT1-AS1,SLC45A2,SLC5A8,SMYD3,ST6GALNAC1,TENM2,TMEM192,TSPAN33,TSPEAR
Андийцы	ABL1,ARHGAP31,CFAP299,DGKB,DLG2,EIF5A2,FRMD4A,GABRG3,HIST1H3I,IMMP2L,KANK1,LIN7A,LINC01622,LINC02046,LOC105377590,LRMDA,MACROD2,MCM3AP-AS1,MTUS2,NKAPL,NRG3,NRXN1,PGBD1,PPP2R5C,PRMT8,QPCT,RHOJ,RSU1,SLC4A4,STK10,TOB2P1,TRPM8,ZKSCAN3,ZKSCAN4,ZKSCAN8,ZNF192P1,ZSCAN16-AS1,ZSCAN26,ZSCAN31,ZSCAN9
Арчинцы	ABCA4,ACVR1,ASPH,CELSR1,CNTNAP2,CSMD1,DLC1,DST,EYA4,ICA1,LINC01317,LRIG3,MLPH,MTUS2,NRXN3,PLEKHG1,PPL,RNF6,SH2D4B,SLC35F4,SLC38A8
Аварцы	BAG2,CACHD1,EPB41L4A,FAM13A,FAM186B,HUS1B,KCNH7,LCORL,LINC00355,LOC101928622,LOC101929705,LRP1B,METTL24,RFX8,TMBIM6,TNFRSF21,TOX2
Багулалы	CFAP299,CR1,ERC1,H1FOO,PTPRE
Бежтинцы	ABLIM3,CALCOCO1,CNTNAP2,CSMD1,DIRC3,DRD3,GPC5,GRM7-AS3,KCNE4,LINC01359,PPM1H,PREX1,SLC25A48,SPTB,TRAPPC9
Ботлихцы	CPNE8,DLG2,MSI2,ZSCAN31
Чамалалы	ADGB,AFF1,ARHGAP26,CREB3L2,CRMP1,GCOM1,MYZAP,KIAA0319L,NKAIN2,NPSR1-AS1,PTRHD1,SCART1
Даргинцы	CPNE4,GRM5,INPP5D,LINC00423,PLCXD2,PLXNA4,SLIT3,TENM3,TMTC2,TRPM8,ZNF436

Продолжение таблицы А.6

Популяции	Гены
Гинухцы	ABR,ACTR6,ADAM33,ARFGEF3,ARHGAP42,ARHGAP8,PRR5-ARHGAP8,ATF7,ATF7-NPFF,BLM,BNC2,BTLA,CADPS2,CAPN14,CAPN5,CBX2,CD200R1,CD44,CD96,CDH4,CHST11,CNTN5,COMMD1,CRADD,CSNKA2IP,CTDSP2,DD B2,DIRC3,DTNA,DTNB,EFCAB8,ERN2,EYS,FAM217A,FAM221B,FBN2,FGF12,FLI1,FLJ22447,FLRT2,FLT1,FRMD6,FRRS1L,FSTL4,GNAL,GNG4,GPC6,GRID1,GRK5,GYPC,HS3ST2,ICA1,IL17B,IL1RAP,ITPR1,KCNAB1,KCNIP4,KCNMA1,KCNN3,LAMP3,LINC01058,LINC01170,LINC01182,LINC01303,LINC01599,LINC01630,LINC01924,LINC02046,LINC02232,LINC02572,LINGO2,LMO1,LOC101929570,LOC730100,LOXL2,LPCAT1,LPP,LRP1B,LTK,MACROD2-IT1,MADD,MAGI2,MANSC1,MCF2L2,MECOM,MICAL2,MIR4535,MTUS2,MYO7A,NAALADL2,NDUFAF6,NEGR1,NELL1,NLRP11,NLRP9,NTM,NTNG2,NUB1,NUDT15,OR51E2,OTOF,PAMR1,PCDH15,PDE10A,PHACTR2,PON1,PPP2R5C,PRELID2,PRKAR2B,PTPRD,PTPRE,PTPRT,RAPSN,RBFOX1,RBFOX3,RHOJ,RNF13,RORA,RPS6KA2,SCHLAP1,SLC45A2,SLC45A4,SLC9A9,SLIT3,SNAP25,SPTB,SPTBN5,SUCLA2,TAS1R2,TBX19,TBXT,TENM4,TGM3,THBS4,THUMPD3,TMEM132B,TMEM132D,TNFAIP8,TNIK,UMODL1,UPP2,USH2A,WDR70,ZBP1,ZFAT,ZNF385B
Гунзибцы	ABCC2,ACOX2,ADARB2,ANKS1B,ARHGAP26,ASAP1,ASTN2,BCL2,CACNA1C,CACNA1E,CADM2,CADM2-AS2,CCR6,CLASP1,CNKSR3,CNPY1,CNTN5,CNTNAP5,CRHR1,CSMD2,DNAH10,DPH2,FAM163B,FAM49B,FAM81A,FER1L6,FGF12,FGGY,FLJ22447,FSHR,GABBR2,GABRB2,GALNT16,GRIK2,HRAT92,HUNK,IGSF21,INTS6,KCND2,KIF3C,LINC02210-CRHR1,LINC02348,LINC02542,LINC02650,LINC02743,LNX1,LOC100506023,LOC101928510,LOC105373764,LRMMDA,LRRFIP1,LRRTM4,MEGF9,MIA2,MIR100HG,MSRA,MT1A,MTRR,NECAB2,NREP,OVCH1-AS1,PAPSS2,PAQR9-AS1,PAX7,PCARE,PLXDC2,PPFIA4,PRELID2,PREX2,PRKG1,PTPRG-AS1,PTPRN2,PTPRS,RARB,RGS14,RPS6KA2,SCARA5,SH2D3A,SLC22A18AS,SMAD7,SPPL3,SPTBN1,ST6GALNAC3,SYT6,UBAP2L,UBASH3B,VWF,WDR70,ZHX2,ZNF717
Караногайцы	ACVR1,BTN3A2,CELSR1,CRISPLD1,CUEDC1,DLG2,KCNE4,KEL,KIAA1324,KIAA1549,LINC02147,LINC02208,LLCFC1,LOC101927267,LOC153910,PDGFRL,PPP2R5C,RARG,RFX3,SDK1,TEAD4,TRABD2A
Каратинцы	ABI2,ALK,ARHGAP26,CARD11,CCDC83,CDH13,CLSTN2,COL6A6,DLG2,DPP6,ERBB4,EYA2,F13A1,GABRB2,HIST1H2AK,HIST1H3I,HMGCR,LINC01807,LRP1B,MAB21L3,MOXD1,NELL1,NFIA,PCGF5,PGBD1,PPARA,RRP8,RSPO4,SHANK2,SLC4A10,SLC4A4,SPSB4,STK10,TENM2,ZKSCAN3,ZKSCAN4,ZSCAN23,ZSCAN26,ZSCAN31
Хваршины	CTNND2,FBN2,LINC02247,LOC105616981,LRRC52-AS1,PLEKHG1,RAP1GAP2,ZNF211
Кумыки	ARHGAP26,DACH1,LRP1B,ZBTB38

Продолжение таблицы А.6

Популяции	Гены
Лакцы	AGMO, ARHGAP26, ARHGAP26-AS1, CEP128, EYA4, FAM238C, FAM71D, FLJ22447, FMO3, GNAT3, HCN4, LINC01303, LINC01376, LINC02107, LRRC4C, MAGI1, MCTP1, NELL1, PLXNA4, PRDM16, PRMT8, PTPRD, PTPRK, SEMA6A, SLC01B1, TMTC2
Лезгины	ACSL3, ASCL1, BARX2, CLSTN2, ETV7, HECW2, HSPG2, HYDIN, IL16, KCND2, LINC00682, LOC100996447, LOC101929705, NFIA, NKAIN2, PLXNA4, RARG, SCOC, SCOC-AS1, TMTC2, UVRAG, ZBTB38
Рутульцы	CTNNA3, CYR1-AS1, FLJ22447, GRID1, KIF6, LINC01844, LMF1, LOC339874, NUDT16P1, MCC, MECOM, MIR31HG, OR12D3, OR2J3, PIGG, PNLIPRP3, SNHG26, SOBP, WT1, ZNF536, ZSCAN23
Табасараны	CELSR1, CFAP70, DAB1, ESR1, GRIK1, GRIK1-AS2, LYRM4-AS1, NAALADL2, PPP2R5E, PSTPIP2, ROR2, TMTC2, TRPM8, ZBTB20
Тиндинцы	ACTN2, ACVR1, B3GNTL1, BDH1, BZW1, C5orf46, DUT, ESR1, FMN2, GRIK1, LINC01266, LINC01378, LOC101928622, LOC285766, MTMR7, NAALADL2, NPSR1-AS1, OLFML2B, PIGG, PXDNL, SLC12A1, ST3GAL3, SYNE2, TRPM8, ZNF12
Цахуры	CFAP299, DOCK1, FGF2, FLOT1, LINC00506, TUBB
Цезы	CAMKMT, CCDC149, EDC4, ESR1, GRIK2, GRM3, GSAP, NLGN1, P3H2, SLC39A11, STK39, STXBP5-AS1, TMEM196, TMPRSS11D, TSPOAP1-AS1
Абхазы	ADGRA3, ANKRD44, AOPEP, ARHGAP26, GPHB5, HEATR1, IL1RAP, JHY, LINC00923, LINC02444, LOC101927078, LOC101927267, LOC105377590, LRP1B, MACROD2, MLPH, PPP2R5E, PSTPIP2, PTPRD, RPF2, STK32B, TENM2, TMEM154, TNIP3
Адыгейцы	KCND2, LPL
Балкарцы	ANXA4, KHDC1, PLXNA4
Чеченцы	ACBD6, ARHGAP10, CELSR1, CRYBG1, LRIG3, MAB21L3, MTURN, RARB, RARG, SLC5A4-AS1, SOAT2
Черкесы	APLP2, ARHGAP26, CYTIP, EPB41L4A, KIAA1549, LINC00381, LINC01170, PPP2R5E, SLC12A1, SLC7A6, SPTB, UVRAG, ZSCAN23
Ингуши	ARHGAP26, COL22A1, DAB1, DUT, FGF2, GCSIR, GRK5, KCND2, LINC01844, LINC02789, LOC730100, NRXN3, PGBD2, PPP2R5C, PSD3, PTPRM, RHOJ, RSPO4, SPATS2L, STK32B, TEAD4
Карачаевцы	BCAS3, CTNNA3, GRIK1, HIST1H3I, IL12A-AS1, INPP4B, KCND2, MTMR2, OR2B2, ZSCAN31
Мегрелы	C12orf50, CNKSR3, ERICH6, FHIT, GABRB2, INPP4B, KIF26B, LDB2, LGI1, LINC01250, LUM, MCM3AP-AS1, PCDH18, PLXNA4, PPARD, PTPRK, PTPRM, SERPINI1, SLC39A11, SLC45A2, SORCS1, WWOX
Ногайцы	DIRC3, DIRC3-AS1, ICA1L, IRS1, LINC02046, LUM, MCC, NAV2, NRCAM, PLXNA4, PPL, SH3GL2, STK32B, TMEM154, TRPM8, ZNRD1ASP
Осетины (С)	ANO2, ARHGAP26, LYRM4-AS1, OPCML
Осетины (Ю)	ABO, ARHGAP26, CFAP70, DEPDC5, FMNL2, GPHB5, GRIK1, LINC01378, LOC100506497, PXDNL, RSPO4, SLC12A1, ZNF804B
Алтайцы (Б)	ADCY1, DHX37, DYNC1H1, FOXP1, LINC00682, LINC02299, LOC101926960, MGLL, PDE10A, PPP2R2B, PTPRG, STK32B, SUCLG2-AS1

Продолжение таблицы А.6

Популяции	Гены
Алтайцы (К)	ADGRB3,ARL15,C5orf66,CELF2,LINC01378,LINC01622,RAPGEF2,RBFOX1,RBMS3,SOX5
Буряты (А)	STAG1
Буряты (К)	GLO1,PCDH15,PLSCR1,PRR16,ST8SIA6,TNFR
Чукчи	ACTA2,ADAMTS18,ANKRD60,CACNA2D3,CDCA7L,CDK5R2,CNTN5,FAM135B,FHIT,GABRG3,GMFG,HS1BP3-IT1,HYDIN,ITSN2,LOC100507283,MOBP,NALCN-AS1,NCALD,NLRP5,NRXN3,OPCML,PAPPA,PCLO,PDLIM5,PGR,PRKACB,SAR1A,SCUBE2,TDRP,TENM2,TMEM65,TNFR,TSBP1-AS1,VAV2,VSTM2A
Чулымцы	ANKRD33B,CFAP299,DPP6,GRIA2,HLA-DQA1,LAMA1,LINC00473,LINC01435,LOC100128164,MEIS1,MIR646HG,MYH11,PREP,PTPRN2,RDH10-AS1,RGMB-AS1,SAMD7,SPATS2,STARD4-AS1,TMEM244,TMSB10,TMTC2,TSBP1-AS1,USH2A,XKR4,ZNF423
Эвенки (Я)	ADGRB3,AUTS2,DOCK8,DPP6,GABRB3,GAP43,LINC02294,LINC02743,LOC101928855,MYO15B,NR5A2,NTF3,PTPRN2,TEC,TENT5A
Эвенки (З)	ADH5,ANO3,FAM228B
Калмыки	CUEDC1,ERC1,MGST3,ZNF407
Кеты	LOC101928304,MCF2L2
Хакасы (Ш)	ANO3,CCDC138,CDH13,CUEDC1,EDAR,LINC00974,LINC02568,PRKG2,STK32B
Хакасы (Т)	AATK,CYP1B1-AS1,FOXP2,GRIK4,HLA-B,KIAA1324,LINC00448,LINC01630,LRP1B,MCPH1,MRC1,TMEM236,NCKAP5L,NCOA7-AS1,NRCAM,NTRK2,PALLD,PLD5,PPM1L,RARB,SDK2,ST6GALNAC1,TRPC6,UBB
Ханты (К)	ARFGEF3,ATP2C1,BTBD9,CCDC33,CD226,CDH13,CSMD2,DGKH,DNAH11,EGFLAM,FAM135B,GAP43,GDI2,GRIN2B,HCG9,ITPR3,KCNAB1,KDM1B,LARGE1,LINC00473,LINC01006,LINC01339,LINC02147,LOC100507283,MAPRE2,PDZD2,PRIMA1,ROR1,RSPO4,SAA3P,SLC6A15,SMARCA2,SMOC1
Ханты (Р)	ADGRA3,AFF3,ALX1,ANK1,ANKRD33B,APBA1,ATP4B,ATXN1,CCND2,CDH13,CDKAL1,CNIH3,CNTN3,CSMD2,CXADR,DCANP1,TIFAB,DDX55,DSCAM,EPHA6,ERC1,FARS2,FAT2,GAREM1,GATAD2A,GDPD5,GLT1D1,GNAL,KIF6,KSR2,LINC00929,LINC01088,LINC01339,LINC02006,LINC02147,LINC02618,LOC101927630,LOC101929595,LOC105370586,LOC105374428,LOC339862,LRFN2,LYRM4-AS1,MTCH1,MYOM1,NAV2,OR5D16,OSBPL6,PBX4,PCDH15,PGR,PIP4K2A,PLEC,PLXNA2,PPP1R16B,PPP2R5C,RASGRP3,RDH13,RFC1,RMND1,RNF7,ROBO2,RYR1,SAA3P,SDK1,SLC14A2,SLC35G2,SLC46A2,SLX4IP,SNAP25,SOX1-OT,ST18,STAG1,SUGP1,TAS1R2,TCERG1L,TCF7L1,TCF7L2,TM6SF2,TMEM163,TSHZ2,TSNAX-DISC1,WWOX
Коряки	AGBL1,ATP8B1,CACNA2D3,CACNA2D3-AS1,CCDC169-SOHLH2,SOHLH2,CD96,CHI3L1,CLCN1,DOK7,FLJ22447,GACAT2,HDAC9,ITPR1,KCNAB1,LIMS1,LINC02542,LOC100507053,LOC105369911,LRRN2,MMP16,MREG,NLRP11,NRG1,OSTM1-AS1,PCDH15,PRKG1,SLC10A7,SLC16A12,SPIRE2,TMPRSS13,TNIP1,ZBED9

Продолжение таблицы А.6

Популяции	Гены
Нивхи	<i>HCG9,LRRC52-AS1,TMEM237</i>
Татары (Т)	<i>ELMO1,LOC101927822</i>
Тувинцы	<i>ADH5,ANO3,BCAR3,CUEDC1,LIMS1,LOC101927822,LYRM4-AS1,RARB,RBFOX1,SULT1C4</i>
Удэгейцы	<i>HPSE,LOC101928764,RBMS3,ST8SIA1</i>
Якуты	<i>ADGRB3,CRTAC1,GLI3,HECW2,LCORL,NTRK3,PRDM6,RHOJ,SMOC1</i>
Дунгане	<i>BDH1,CENPT,CUEDC1,CYP1B1-AS1,EDAR,LINC00448,PRKN,SHC3,SV2B,TSNAXIP1,XKR6,ZNF322</i>
Казахи (М)	<i>CLMP,CUEDC1,CYP1B1-AS1,DLG2,EPB41L3,LINC01229,LOC100420587,LOC101927822,NPHP4,PXDNL,SDK1,TRABD2A</i>
Казахи(С)	<i>BACE2,CUEDC1,DAB1,ELMO1,LIMCH1,LINC00682,NAV2,PPP2R5E,PRKN,RHBDD1,SPRYD3,STK10,TMBIM6,WWOX</i>
Киргизы	<i>BDH1,CADM3,CDH13,COL21A1,CUEDC1,ERC1,FBXO25,LAMA1,LIMCH1,LINC02147,LOC101927822,PPP2R5C,RNF14,STK10,TRABD2A</i>
Узбеки	<i>ANO3,ATP13A3,C5orf66,CNN1,CYB561A3,DMRT1,ELMO1,NRG3,STARD13</i>

Таблица А.7 Результаты поиска сигналов направленного отбора с помощью теста iHS

Популяции	Гены
Абхазы	A4GALT,ALK,CALM2,COLEC11,EFCAB6-AS1,ENPP7,EXOC6B,FAM20C,FOXK1,GAS6-AS1,LINC01844,LINC02444,LOC101929006,MRC1,TMEM236,MRPL23,PLXNA4,PRKAR1B,RPF2,SERPINE2,SLC2A9,STK32B,TAFA1,TENM2,TMEM255B,XIRP2
Адыгейцы	ABCA1,ABCC10,ABHD12,ADAMTS16,ADAMTS9-AS2,ADAMTSL2,ADCY8,ADGRF2,ADGRV1,ADPGK,AFTPH,AGBL1,AGFG1,ALKAL1,ANKS1B,APLF,ARHGAP19-SLIT1,ARSG,ART3,ART4,ATE1,ATP6V1G3,ATP8A2,ATP9B,BAG6,BAIAP2L1,BBS4,BCAR3,BCAS3,BPTF,C2CD3,C3orf67,C4orf47,CCDC146,CCDC178,CCDC3,CCDC91,CCP110,CDH10,CDK2AP2,CDK6,CELF2,CERT1,CFAP52,CFAP99,CFH,CHIC2,CHRNA3,CIB3,CLDN14,CLGN,CLTC,CMSS1,FILIP1L,CNTLN,CNTN5,CNTROB,COG1,COL24A1,COL25A1,COL6A4P2,COL6A5,CRELD2,CSMD2,CSMD3,CSNK2B,CSTF3,CTDSPL,CTIF,CTNNA2,CUEDC1,CYP19A1,DCAF4L1,DCAF5,DDB1,DDHD1,DDX1,DERA,DIP2B,DIRC3,DIRC3-AS1,DLC1,DLG2,DMRT3,DNAH6,DNAJA3,DOCK10,DOCK3,DOCK4,DOCK7,DSCAM,DSCAML1,DST,DTX4,ECD,EFCAB10,EFCAB2,ELF1,ELP1,EP400P1,EPB41L1,EPB41L4A,EPHA4,ERBIN,EVI5,EXD3,EXOC6B,FAM169A,FAM186B,FBN1,FHAD1,FMNL3,FOXO1,FRA1S1,FSIP1,FUT9,GABRA5,GALK2,GANC,GAS7,GATAD2A,GATAD2B,GEN1,GFM2,GJA9,GKAP1,GNAI3,GPATCH8,GPR39,GPR55,GRIK2,GRIN2A,GRIN3A,GRM8,GTF2IRD1,GULP1,HDAC11,HEATR1,HECTD2-AS1,HIPK2,HIPK3,HIST1H2AK,HMBOX1,HPSE2,HS2ST1,HTR7,IARS,IFI16,IMPACT,INPP4B,INVS,IRGM,ISM1,ISSY1-RAB43,RAB43,ITCH,ITFG1,KCND2,KCNQ3,KCNS3,KCTD10,KDM1B,KIRREL2,KLHL1,KLHL24,KMT2E,KNL1,LAMA3,LCE2A,LDB2,LGALS14,LGALS8,LIN28B,LIN7A,LINC00261,LINC00402,LINC00452,LINC00469,LINC00504,LINC00565,LINC00682,LINC00871,LINC00907,LINC00923,LINC01019,LINC01250,LINC01331,LINC01362,LINC01376,LINC01606,LINC01609,LINC02046,LINC02231,LINC02389,LINC02436,LINC02585,LINC02624,LINC02664,LINGO2,LIPE-AS1,LNPEP,LOC100129620,LOC100288728,LOC100507291,LOC101927045,LOC101927141,LOC101927668,LOC102724084,LOC105369911,LOC105371849,LOC105616981,LOC145694,LOC285766,LOC339166,LOC400655,LOC646588,LOC728755,LPL,LRIG3,LRMDA,LRR52-AS1,LRR52,LSAMP,LUZP6,MTPN,MACROD2,MAGI2,MAP2,MAP3K19,MAPKAPK5,MARCHF6,MB21D2,MBOAT2,MC4R,MCTP1,MELK,MIER1,MIG7,MIOS,MIR99AHG,MMP16,MOBP,MPP5,MRC1,MROH2B,MRPS15,MTHFD2L,MTRF1,MTUS2,MUC21,MYBPHL,MYO9A,MYOF,MYRIP,NCAM1,NCBP1,NFE2L1,NFIA,NIF3L1,NKAIN2,NMBR,NOL8,NPM1,NR5A2,NSA2,NTRK3,NUDT13,NUP50,OPA1,OPCML,OR52D1,ORC2,ORC4,OSTM1,PAN3,PARP4,PCDHA1,PCDHA2,PCDHA3,PCNT,PCNX4,PCSK5,PDCD10,PDE11A,PDE1A,PEX1,PHACTR1,PHKB,PHLDB2,PLCXD2,PIK3CB,PIKFYVE,PIP4K2A,PISD,PJA2,PKDREJ,PKHD1L1,PLA2G4A,PLCB4,PLCE1,PLEKHA8,PLG,PLXNA4,PPARGC1A,PPP1R12A,PPP1R12A-AS1,PPP2R5E,PPP4R2,PPP4R3A,PRH1-PRR4,PRKAR2A,PRKCQ,PRR14L,PRRC2A,PTPN14,PTPRB,PTPRK,PTPRN2,PTPRT,PTRH2,QSER1,RAB3GAP2,RACK1,RAD51AP1,RANBP3L,RAP1GAP2,RAVER2,RBM33,RHBD1,RIC3,RNF121,RNF14,RNF169,RPL39L,RPS6KA2,RSU1,RUFY3,RYR2,RYR3,SCEL,SCHLAP1,SCP2,SEC23B,SELENOF,SEMA5B,SETD3,SFI1,SGCZ,SHLD1,SHOC1,SLC14A1,SLC14A2,SLC28A3,SLC2A9,SLC30A9,SLC4A4,SLC4A7,SLC9A8,SLCO1B3,SLCO1B7,SLK,SMAD5,SNX29,SP3,SPATA5,SPIN1,SPTB,SRP14,STIM1,STK31,SUFU,SULF1,SYBU,SYNE2,TBC1D22A,TBC1D22B,TCF7,TCFL5,TCTN2,TDRD9,TENM2,TENT2,TEX2,THEG,TM6SF2,TM9SF4,TMCC1,TMEM220,TMEM38B,TMPR

Продолжение таблицы А.7

Популяции	Гены
	SS11D,TMTC3,TNR,TPRG1,TPST1,TRPM3,TSBP1-AS1,TSHR,TTK,UBE3B,UHRF1BP1,UNC5D,URB1,USP40,VWA2,WDR74,WFDC3,WWC2,WWOX,ZBTB7C,ZBTB8A,ZFP90,ZFPM2,ZMYM5,ZNF148,ZNF333,ZNF398,ZNF521,ZNF562,ZNF577,ZNF609,ZNF763,ZNF782,ZNRD1ASP,ZSCAN9,ZYG11B
Агулы	A4GALT,AGBL4,BST2,CMYA5,COLEC11,ELMO1,FAM20C,GAS6-AS1,GRK1,LOC100507065,LOC101928622,LRRMT4,MRC1,NRG3,TAFA1,TMEM255B
Ахвахцы	ABRAXAS2,ADCY1,CUEDC1,EIF4B,GRIN2B,KRTAP12-2,LINC01951,LINC02147,LOC100507053,LOC105374972,MIA3,OR6C2,RUFY1,SLC45A2,SLC5A8,TENM2,TRIM10,TSPEAR,WWOX
Алтайцы (Б)	FAM20C,FSTL1,LINC00565,LINC02299,MRC1,TMEM236,RNASEH1,RNASEH1-AS1,RPS7,SUCLG2-AS1,TMEM255B,TTL1
Алтайцы (К)	ATP4B,CCDC146,ERC1,FSIP2,GABBR1,GAS6,ISX-AS1,LIMS1,NUDCD1,PKIA,RNASEH1-AS1,SOX5,TMEM255B
Андийцы	A4GALT,ARHGAP31,BABAM2,DLG2,EBF3,EPG5,ERG,EXD3,GABRG3,LINC00452,LINC02046,LINC02516,LOC202181,MBD4,MIR2052HG,MOG,MRC1,MTUS2,MYO1B,MYO1H,MYRIP,NPSR1-AS1,NR5A2,OPCML,PDE9A,PGBD1,PPP2R2C,PRMT8,RASA3,RNASEH1-AS1,RYR2,SHQ1,SOX2-OT,SPTB,STK10,TEX36,TMEM255B,TTC1,VTI1A,ZFYVE16,ZKSCAN4,ZKSCAN8,ZNF192P1,ZSCAN16-AS1,ZSCAN26,ZSCAN31
Арчинцы	A4GALT,ASPH,CCHCR1,FAM20C,GAS6,GAS6-AS1,ISX-AS1,LINC00452,MRC1,PACSIN2,PRKAR1B,RNASEH1-AS1,RPS7
Аварцы	A4GALT,APOL1,BAG2,COLEC11,FAM13A,FAM186B,FAM20C,GALNT9,ISX-AS1,KCNH1,LCORL,MRC1,POLN,PRKAR1B,RPS7,SUCLG2-AS1,TMEM255B
Багулалы	CDK17,COLEC11,CR1,DNAH14,FAM20C,FLI1,GAS6-AS1,LINC00452,LINC00565,LINC02538,MED6,MRC1,TMEM236,RNASEH1,RPS7,SIVA1,TMPRSS11E
Балкарцы	ENPP7,MRC1,PLXNA4,RNASEH1,SLC2A9,TCTE1,ZNF12
Башкиры (Б)	COLEC11,DCAF4L1,FAM20C,GRK1,MRC1,NRG3,PSORS1C1,SPATA13,TSBP1,TSBP1-AS1,TWIST2,VAV2,YTHDC1
Башкиры (П)	ABLIM2,AFDN,CCDC102B,CCDC146,DHX16,DSCAML1,FAM120A,FAM20C,GABBR1,GAS6,GAS6-AS1,GSAP,HCG17,HCG18,INMT-MINDY4,KCNH1,KCNH7,LOC100507065,MANBAL,MIR2052HG,NFIB,NIF3L1,NUP210,ORC2,PDE1A,RBFOX2,RHEX,SLC2A9,SMAD5,SPATA5,STAB2,TM9SF4,TMEM255B,TOX2,XIRP2,XKR6,ZBTB38,ZNF804A
Башкиры (С)	ARHGAP26,ATRNL1,B4GALT5,CAMSAP2,CDC16,CDH13,CDYL,COPS3,CUBN,CYP20A1,DLC1,DMBT1,EGFLAM,ELMO1,ERC1,FAM20C,FHIT,GCNT2,GPX5,GPX6,GRID1,HRH1,IMMP2L,KDM2A,KEL,LINC00967,LINC01250,LINC01322,LINC01356,LINC01435,LINC02110,LINC02624,LIPE-AS1,LOC145694,MRC1,MTHFD1L,NCOA1,NCOA3,NRXN3,PCDH9,PLCB4,PTDS2,RBFOX1,RHOJ,RPS6KA2,SIGLEC12,SLC25A21,SLC9A9,SPAG16,ST7,STARD4-AS1,STK32B,TET2,TMEM236,TWIST2,USP12,YJEFN3,ZBTB38
Бесермяне	CCDC102B,CDC14A,CPA6,CSMD1,CYP20A1,DDR1,DIRC3,DIRC3-AS1,FAM126B,FAM20C,FBN1,FO XK1,FYB1,GAS6-AS1,GRK1,KCNMB2-AS1,LINC02624,LOC100128317,MALRD1,MRC1,TMEM236,NECTIN3,NELL1,OPRD1,ORC2,PDE1C,PLCB4,PPP1R11,PTPRN2,RGS5,RNASEH1,RNF39,RUBCN,TMEM117,TMEM255B,TRIM26,TTC1,WWOX,ZBTB10,ZNRD1ASP

Продолжение таблицы А.7

Популяции	Гены
Бежтинцы	ANKRD44,CDH17,CNTNAP2,CSMD1,DRD3,DTNBP1,DUSP8,FAM20C,LMF1,MRC1,MTNR1A,MYRIP,RNASEH1,RNASEH1-AS1,SLC25A48,ZNF12
Ботлихцы	ABCB1,CCL15-CCL14,CDKN2B-AS1,CR1,EXD3,FAM20C,GALNT2,GSAP,HLA-DRB1,ISX-AS1,LIMCH1,LINC01162,LINC01250,LINC02234,LOC105378127,MECOM,MRC1,MSI2,NLRP11,RANBP3L,RAPH1,RPS12,RPS7,SH3RF1,SQSTM1,TIGAR,TMEM14A,TMEM255B,TWIST2,ZNF12,ZNF192P1,ZSCAN16-AS1
Буряты (А)	ANKS1B,ATG10,BCAR3,ERC1,FAM184B,ISX-AS1,LINC01029,LOC101928937,MED28,MPPED1,RIMS2,STAG1,TMEM255B,TLL1,TWIST2,YTHDC1
Буряты (К)	ADI1,COLEC11,ERC1,FAM20C,GAS6-AS1,GLO1,LINC00452,RNASEH1,RNASEH1-AS1,VWA3A
Чамалалы	A4GALT,COLEC11,KIAA0319L,MRC1,PRKAR1B,RNASEH1,RNASEH1-AS1,RPS7,SPIN1,TMEM131L,TMEM255B
Чеченцы	ACBD6,ARHGAP10,EFCAB6,FAM20C,LINC00452,LINC02516,LOC101928565,MRC1,PAN3,PPP1R12A,PRKAR1B,RARG,RNASEH1,RNASEH1-AS1,RPS7,SLC22A15,SLC2A9,SNX24,USP40
Черкесы	APLP2,BCAS4,COLEC11,EFCAB6,FAM20C,FBN1,FMNL3,HECTD4,LINC00381,LINC01170,LMO4,LOC101928416,MANSC1,MKLN1,MRC1,TMEM236,NEK11,PDE9A,PISD,PLCB4,RAVER2,SLC12A1,SLC2A9,SLC44A1,SP3,SPTB,TMEM255B,TWIST2,UNC5C,UVRAG,ZBTB38
Чукчи	AKAP13,ANO2,CDCA7L,ENOX1,GALC,GMFG,HS1BP3-IT1,ISX-AS1,ITSN2,LINC02538,LOC100507283,MOBP,MRC1,TMEM236,NRXN3,OPCML,PRKACB,PRKAR1B,SCUBE2,TMEM65,TSBP1-AS1
Чулымцы	LAMA1,LINC00454,LOC643542,MIR646HG,PLD6,RALYL,SPATS2,TSBP1-AS1
Чуваши	A4GALT,CD226,COLEC11,EGFLAM,FLJ22447,GABBR1,HIF1A-AS3,HLA-V,LIMCH1,LOC101929935,LOC105378127,MPPED1,MRC1,PRKAR1B,RNASEH1-AS1,TMEM255B,ZNF99
Даргинцы	ARHGAP44,D2HGDH,FAM20C,ISX-AS1,LOC256880,MRC1,PLXNA4,RNASEH1,RNASEH1-AS1,RPS7,SNX4,TMEM255B,TMTC2,TSBP1,TSBP1-AS1
Дунгане	ADI1,CADPS,CCDC138,CCHCR1,CENPT,CPEB4,CREB5,CUEDC1,D2HGDH,DOCK7,DOCK8,EDAR,EFCAB6-AS1,FAM20C,GH2,GRID1,HDAC7,HSPC324,IFI30,KCNK10,KCNQ1OT1,LIMS1,LOC102724659,LRR4C,MEG8,MRC1,NRN1L,NUTF2,PAPPA2,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PDE3A,PPT2-EGFL8,RANBP10,RNF14,SCHLAP1,SIK3,SMC6,SPTLC2,SULT1C4,SV2B,TAF1A1,THAP4,TMEM117,TMEM236,TMEM255B,TSBP1-AS1,TSNAXIP1,ULK4,VPS37C,XKR6,ZBTB38,ZKSCAN7,ZNF322
Эвенки (Я)	ADGRB3,ARFGAP3,DOCK8,DPP6,DSCAM,ERC1,FAM20C,GAS6-AS1,HDAC7,LOC101928324,MYT1,SCHLAP1
Эвенки (3)	ADH5,FAM20C,TMPRSS11E

Продолжение таблицы А.7

Популяции	Гены
Гинухцы	ABCC5,ABCG8,ACBD3,ACP2,ACTR6,AFAP1,AGPAT4,ARFGEF3,ARHGAP42,ARHGAP8,PRR5-ARHGAP8,ARHGEF10,ASH1L,ASTN2,ATF7,ATF7-NPFF,ATP4B,BBX,BHLHE40-AS1,BLM,BMS1P21,BNC2,BOD1L1,C1QTNF3-AMACR,CA10,CABP1,CACNB2,CACNG2,CADPS2,CAPN5,CBX2,CD200R1,CD247,CDH13,CDH4,CDYL,CFAP299,CFDP1,CHST6,COBLL1,COMMD1,CRADD,CSGALNACT1,DIRC3,DLC1,DNAH17,DOCK7,DRC1,DSCAM,DTNA,DTNB,DYNC1I1,DYRK4,EBF3,EFCAB8,EHBP1,EIF3A,ENTPD4,EPB41L4A,EPHA4,EPN1,EYS,FAM217A,FAM221B,FLRT2,FMN2,FRMD6-AS2,FRRS1L,FSTL4,GALM,GPATCH2,GPC6,GRK5,GSG1L,HSPA12A,HTR4,ICA1,IGF2BP2,IL17B,IL1RAP,IL21R,IL4R,IRF6,ITPR1,KATNAL1,KCNA6,KCNH5,KCNIP4,KCNK2,KCNMA1,KCNMA1-AS1,KIAA1211,KIRREL3,LINC00384,LINC00519,LINC01182,LINC01303,LINC01411,LINC01630,LINC02046,LINC02232,LINC02245,LINC02337,LOC101927421,LOC101929095,LOC101929294,LOC101929415,LOC105377716,LOC105378098,LOC374443,LOC730100,LOXL2,LPP,LRP1B,LRRC4C,LSG1,LY75,LY75-CD302,LY86-AS1,LYAR,LYST,MACROD2-IT1,MAGI2,MCF2L2,MECOM,MIR4535,MMD2,MMEL1,MOSMO,MTHFD1L,MYO3B,NAALADL2,NDUFAF6,NLRP11,NLRP9,NMU,NPR3,NTM,NUDT15,OR51B5,OR51E2,OSBP2,OTOF,PAMR1,PDE10A,PDE4D,PDE6G,PDZD2,PICALM,PIEZO2,PIGN,PKP4,PPARGC1A,PRELID2,PRKAR2B,PRKCB,PRKCH,PRKN,PSME4,PTPRD,PTPRE,PTPRT,RAB2B,RAPH1,RBFOX1,RBFOX3,RGS7,RNF13,RNF213-AS1,RPL13AP17,RPS6KA2,RPTOR,RYR2,SAE1,SAMD5,SCHLAP1,SCNN1G,SLC13A3,SLC1A1,SLC35F1,SLC9A9,SLIT3,SPATC1L,SPHKAP,SPTB,SPTBN5,SRBD1,ST20-AS1,STARD13,STK32B,STX18-AS1,SUCLA2,SUN1,SV2B,SVIL,SYCP2L,TBXT,TENM4,TFDP1,THBS4,THUMPD3,TMEM132B,TMEM132D,TMEM150C,TMEM207,TNIK,TNRC6A,TRA2B,TRIM9,TSC22D2,TSPAN10,TTC34,TXLNB,UBASH3B,UMODL1,UQCRC2,USH2A,WASF3,WDR70,XIRP2,XRCC5,ZDHHC14,ZFAT,ZKSCAN2,ZNF112,ZNF407,ZNF516,ZNF555
Гунзибцы	AACSP1,AATK,ADAM23,ADI1,AP3S1,ARHGAP26,ARHGEF37,ASAP1,BUD13,BZW1,CACNA1C,CASC21,CASC8,CBR3-AS1,CCDC60,CCR6,CCSER1,CLASP1,CLSTN2,CNPY1,CNTNAP5,COL4A1,CRYBG1,CSMD1,CSNK1G1,DNAH10,DYRK1A,ENDOD1,ERG,EXOC1L,FABP2,FAM20C,FAM49B,FAM81A,FGF12,FGGY,FHIT,FKBP1A-SDCBP2,SDCBP2-AS1,FLJ22447,FSTL1,GABBR2,HLA-DPA1,HLA-DPB1,HOPX,HRAT92,HTT,IL12A-AS1,LINC01100,INTS6,JPT1,KIF14,LINC01182,LINC01478,LINC02380,LINC02542,LINC02743,LOC100505502,ZEB1,LOC101927314,LOC102724511,LRRC4C,LRRFIP1,LRRTM4,MAD1L1,MEGF9,MIA2,MINDY2,MIR100HG,MOK,MSRA,MTRR,MYO16,MYT1L,NACA,NIPBL,NKAIN2,NPAS3,NREP-AS1,ODF2L,ORC2,PAPSS2,PARVB,PAX3,PCARE,PDE10A,PIEZO2,PLA2G5,PPFIA4,PRKAG2,PRKAR1B,PRKG1,PRORP,PTPN14,PTPRD,PTPRG-AS1,PVT1,RGS14,RORA-AS1,RPS6KA2,RREB1,SCAF8,SCD5,SCFD1,SENP7,SESN1,SLC14A2,SLC14A2-AS1,SMG6,SNTB1,SNX29,SORCS2,SPPL3,ST6GALNAC3,STK33,SYCP1,SYT6,TAF1B,TCTE1,THUMPD3,TIMMDC1,TMPRSS11E,TTL1,UBAP2L,UBASH3B,UBE2O,USH2A,WDR20,WDR64,WDR70,YIPF1,ZFAND3,ZNF664-RFLNA,ZNF717
Ингуши	ACSBG2,BCAR3,COLEC11,EYA3,FAM20C,LINC00452,LINC01250,LOC100128340,LOC730100,MRC1,PAMR1,PGBD2,PPP1R11,PTPRM,RNASEH1,SLC2A9,SYT14,TEAD4,TMEM255B,TRAPPC10,TWIST2

Продолжение таблицы А.7

Популяции	Гены
Калмыки	<i>BBX, C4orf47, COL22A1, CUEDC1, EGFLAM, ERC1, FAM20C, GAS6, KCNC2, KDM4B, LINC00261, MPPED1, MRC1, ORC2, PAPP2, PAX2, RHEX, SPATS2, SUCLG2-AS1, TNFR, UFSP2, USP20, ZKSCAN7, ZNF407, ZNF445, ZNF546</i>
Карачаевцы	<i>AGR2, EFCAB6, GRIN2B, GRK1, INPP4B, IQCA1, MPPED1, MRC1, PRKAR1B, RYR2, SLC2A9, TMEM236, TMEM255B, TMPRSS11E, TTLL1, ZKSCAN4</i>
Караногайцы	<i>ADAMTS6, AHI1, ANKS1B, APOM, ARPC1B, ASH2L, ATP6V1G2-DDX39B, C12orf50, C18orf61, CDH7, CPA6, CPNE1, CRMP1, CSMD2, CUEDC1, DDH55, EHD3, EIF2B1, EIF3E, ENPP1, ESF1, FAM20C, FAM71D, FGF9, FRMD5, FSTL1, G3BP2, GALNT9, GAS7, GGCX, GPATCH2, GRID2, GRK3, GSG1L2, GTF2F2, HS2ST1, IL1RAP, KEL, KIAA1324, KIAA1549, KIF9-AS1, KRT40, LINC00448, LINC00967, LINC01088, LINC01250, LINC01630, LOC101929532, LRIG1, MAGI2, MAPK14, MBP, MCTP2, MPP5, MRC1, NRG3, OR4E2, OSTM1, PAK5, PALLD, PDGFRL, PHIP, PIP4K2A, PLCB4, PTH2R, PTPN4, PTPRG, RAB35, RBF, FOX1, RPSAP52, RSPO3, SLC12A1, SLC16A12, SLC25A21, SLC25A26, SLC2A9, SLC34A3, SPEN, SUSP6, SYT9, TCTN2, TEAD4, TMEM236, TOX2, TRABD2A, TRAPPC10, TRPM8, TSHR, TULP3, TWIST2, USP20, ZKSCAN4, ZNF717, ZPBP</i>
Каратинцы	<i>ANHX, COLEC11, FAM20C, HIST1H2AK, LOC105378127, MPPED1, MRC1, NOTCH2, PRKAR1B, RNASEH1, RNASEH1-AS1, SLC22A15, TMEM255B, TMPRSS11E, ZKSCAN3</i>
Карелы	<i>A4GALT, ADGRA3, ANO3, CDH4, COLEC11, EFCAB6, EFCAB6-AS1, ERC1, FAM20C, GALNT9, GPATCH1, GRK1, ISX-AS1, KIAA1217, LINC00693, LINC01844, MAS1L, MRC1, MUC4, NFIA, NUTM2B-AS1, OR10C1, OR2H1, PISD, POU5F1, TCF19, PPARGC1A, PPP4R2, PREP, R3HDM1, RHPN2, ROR2, SHISA9, SLC14A2, STX18, TMEM255B</i>
Казахи (М)	<i>ACOT7, ARHGAP42, COLEC11, CSMD1, CUEDC1, ELMO1, EPB41L3, ERC1, FSTL1, GRM6, KCNH1, LINC01229, LOC100420587, LOC101927822, LOC101928416, ME2, MPPED1, MRC1, MYO1B, NPHP4, SP3, SULF1C4, TEF, TMEM117, TRABD2A, UMOD</i>
Казахи(С)	<i>AGFG1, APOL1, COLEC11, CUEDC1, DAB1, ELMO1, ERC1, FAM20C, FHIT, GABRA5, GALNT9, GAS6-AS1, IMMP2L, LINC01497, LIPA, MIR548F3, MIR646HG, MLPH, MRC1, TMEM236, NAV2, POLE2, PPP2R5C, RNASEH1, RNF39, STK10, TMEM255B, TWIST2</i>
Кеты	<i>ARID1B, ATE1, CDKAL1, CYP4X1, DMBT1, EFNA5, ERI2, FAM20C, GAS6-AS1, GNL3, GRID2, HCG17, HCG18, HLA-L, LAMC1, LINC01500, LINC02234, LINC02487, LINGO2, LOC100506679, LOC101928304, MRC1, MYO16, PDE3A, PDZRN4, RABGAP1L, SLC2A9, STK38, TOM1L1, TTL1, TWIST2, WWOX, ZDHHC14</i>
Хакасы (Ш)	<i>ADI1, ANO3, CCDC138, COLEC11, DENND5A, EDAR, EGFLAM, FAM20C, GABBR2, GAS6-AS1, MRC1, PPEF2, PRKAR1B, RNASEH1, TMEM255B, ZNRD1, ZNRD1ASP</i>
Хакасы (Т)	<i>CSMD1, DIAPH3, FAM186B, GAS6, GAS6-AS1, GNL1, GPHN, ISX-AS1, LOC440084, MCPH1, MRC1, TMEM236, PPP1R10, TMEM255B</i>
Ханты (К)	<i>ATP2C1, CDH13, COLEC11, CSMD2, DNAH11, FAM20C, GAS6, GAS6-AS1, GDI2, LINC00452, LOC100506679, MAPRE2, MRC1, RNASEH1, RNASEH1-AS1, RPS7, TBC1, TSBP1, TSBP1-AS1</i>
Ханты (Р)	<i>ADGRA3, AFF3, ATP4B, C9orf129, CCND2, DDX55, DSCAM, EEF1AKMT2, FAM53B, GALNT9, GAREM1, LINC02147, LINC02538, LRRIQ1, PAPP2, PBX4, PGR, PIP4K2A, PLEC, PRKAR1B, RDH13, RNASEH1, RNASEH1-AS1, RNF7, SNAP25, STAG1, TCERG1L, TMPRSS11E, ZNF236</i>

Продолжение таблицы А.7

Популяции	Гены
Хваршины	B4GALT1,CTNND2,GRK1,KCNK2,NCALD,NCOA7,NIF3L1,NRG3,NXN,ORC2,PD E1A,PHKB,RAP1GAP2,RFX8,RNASEH1- AS1,SCFD1,SERPINB2,TAOK3,TIGAR,TMEM255B,TPP2,ZNF211
Коми	ACOT7,COL6A5,CPLANE1,CTSE,EGFLAM,EPHB1,FAM107B,FAM20C,IL12RB2, MRC1,NRG3,OR2H1,PRDM16,PRKAR1B,RNASEH1,RNASEH1- AS1,RPS7,SLC30A9,TMEM117,TMEM255B,TWIST2
Коряки	ADH5,AFG1L,ANO3,BRINP1,CABP5,CACNA2D3,CACNA2D3-AS1,CCDC169- SOHLH2,CCDC85A,CLCA4- AS1,CORO1B,CYFIP2,CYP2C19,CYTH4,DLG1,FAM228B,FLJ22447,FSTL1,GAS6 ,INPP5D,KIF20B,LGR6,LIMS1,LINC02227,LINC02532,LINC02538,MARCHF4,M ETTL5,NLRP11,PCDH15,PRKG1,PRKN,RABL3,RNASEH1,RNF217,RPN1,RPS6K B2,RTN4IP1,SUGCT,SYNE2,TMEM169,TRAPPC10,TRIM10,XRCC5,ZBED9,ZC3 H14,ZFAND3,ZNF236,ZNF333,ZNF804B
Кумыки	AGBL4,COLEC11,DACH1,EFCAB6,EFCAB6- AS1,FAM168A,FAM20C,MRC1,TMEM236,PTPRM,RNASEH1,SLC25A26,SLC2A9, TMEM255B,TWIST2,USP25,ZBTB38
Киргизы	ANKFN1,BDH1,CD226,CEACAM18,CTIF,CUEDC1,DDX55,DNAH10,DNAH6,DN AJC5B,DYM,EFCAB6-AS1,ERC1,FAM20C,FYTTD1,HIPK2,HS2ST1,ISX- AS1,LINC02299,LOC101927822,LOC440982,MPPED1,MRC1,PIKFYVE,PTPRG,R NASEH1,RNASEH1-AS1,SLC2A9,SLC30A9,SRP54- AS1,STK10,TMEM117,TMEM255B,TRABD2A,TLL1
Лакцы	AFAP1,AFAP1- AS1,COLEC11,D2HGDH,EYA4,FAM20C,FAM71D,FMO3,GNAT3,L3MBTL4,LIN C02538,PRKCH,PRMT8,RNASEH1,RNASEH1- AS1,SLCO1B1,TENT4B,TMEM255B
Лезгины	ANHX,COLEC11,CTIF,EFCAB6- AS1,IL16,LOC100507065,MRC1,NFIA,RNASEH1,RNASEH1- AS1,RPS7,SCHLAP1,SCOC,SLC27A4,SLC2A9,SLC30A9,TMTC2,UVRAG,ZBTB38
Марийцы	ADI1,COLEC11,EFCAB6,FAM20C,GAS6,HIF1A- AS3,LIMCH1,MALRD1,MPPED1,MRC1,PSMB8-AS1,RANBP3L,TAP1,TMEM236
Мегрелы	ATP4B,C12orf50,C8A,CRYL1,ENPP7,FAM20C,FHIT,IKBKB,INPP4B,LDB2,LOC3 39166,MCM3AP-AS1,MPP5,MYO16,NEBL,PTPRK,REC114,RNASEH1,RNASEH1- AS1,RSPO4,SERPINI1,SIDT1,SLC2A9,SLC39A11,TMEM255B,WWOX,ZNF546,ZN F780A
Мордва (Э)	ABCC1,ADAMTS18,ADGRL2,ANO6,ANP32A,ASIC2,ATAD2B,BST2,CACNA1A,C DKAL1,CNIH1,COL6A5,COLEC11,CTNNA3,DGKI,DHX16,EFCAB6- AS1,EFCAB7,EGFLAM,EIF2S1,FAM71D,GMFG,GRIN3A,HCG15,HCG9,IL2RB,I RS1,JMY,KCNH1,KDM2A,KSR2,LARP1,LEKR1,LINC01229,LINC01237,LINC013 39,LINC01727,LINC01844,LINC02046,LINC02670,LOC100130987,MED6,MIR13 7HG,MKKS,MTMR6,MTTP,MTUS2,OPA1,OPA1- AS1,OR2J2,ORC4,PALLD,PCMTD1,PDE1C,PPP1R11,PPP1R12A,PRKCH,QSOX 1,RAB11FIP3,RHEX,RNF39,SEMA5A,SFSWAP,SLC30A8,SPAG16,SPATA13,STP G2-AS1,STX18- AS1,TENT2,TMPRSS11E,TRIM2,TTC37,TTC7B,TWIST2,UBN1,UBR1,USH1C,USP 8,ZBTB38,ZBTB8A,ZKSCAN4,ZKSCAN8,ZNF311,ZNF717,ZNRD1,ZNRD1ASP,ZS CAN16-AS1,ZSCAN31

Продолжение таблицы А.7

Популяции	Гены
Мордва (М)	ARHGAP10, ARID1B, COL6A5, COLEC11, EFCAB6, EXD3, GALNT9, HIF1A-AS3, HLA-DMA, INMT-MINDY4, KCNH7, LDLRAP1, LOC285766, MRC1, PDE4B, PPP4R2, PSD3, RALGPS1, RNASEH1, RNASEH1-AS1, RNF212B, RSPO4, SDR42E2, STK32B, TMEM255B
Мордва (III)	A4GALT, ABCA1, ABI1, AFTPH, AKT3, ARHGAP31, ARID1B, ATG3, ATP5MPL, CDH13, CDKAL1, COL6A5, CTNNA2, CYP2E1, CYP4Z1, DSCR10, EFCAB6-AS1, EIF2A, FAM20C, FLJ22447, GABBR1, GPATCH2, HLA-DPA1, IFT122, IL1RAP, IL6ST, INPP5E, KAT2B, KDM5A, LOC101929279, LOC339874, LTBP1, MAST4, MBD4, MRC1, TMEM236, MTRR, NUP210, ORAI1, OSBP1, OSCP1, PARP12, PDE4B, PDK1, PROX1-AS1, PPK, REC114, RGS6, RNASEH1, RNASEH1-AS1, RPS7, SEC16A, SFPQ, SNTG1, TBC1D31, ZBTB38, ZFH4, ZMAT4, ZMYM4, ZNF252P, ZNF536, ZNF732, ZNF778, ZNRD1, ZNRD1ASP
Нивхи	ABCB11, ADGRB3, APOL1, BZW1, CDH13, ENDOD1, EPHB1, G3BP2, GABBR1, KCNIP4, KPNA4, LCOL, LINC00457, LINC02343, LOC284930, MAP3K2, MIR4500HG, MIR4713HG, MRC1, TMEM236, MYOM2, NELFCD, NSUN7, OR4E2, ROBO2, SEMA5A, SLC35F3, SORCS2, TANC2, TCF7L2, TMEM236, VARS2, ZBTB38, ZNF398, ZNF568, ZNF804A, ZNRD1ASP
Ногайцы	ABCA1, ABCB1, ADGRB3, AGL1, ANKIB1, APOL1, ARID4B, B4GALT5, BCAR3, BCAS4, C1orf61, C4orf47, CASZ1, CCDC85A, CCHCR1, COLEC11, CSNK1G1, CTNNA2, CTNND2, DDX19B, DNAJC5B, DOCK7, EDEM1, EFCAB6, ENPP7, FAM20C, FKBPL, GALNTL6, GPATCH2, HCG17, HCG18, HLA-DPB1, HLA-L, HPF1, HUNK, ICA1L, ITIH5, JAKMIP1, KAZN, L3MBTL4, LINC01787, LINC02046, LOC105369911, MCC, MECOM, MIR2052HG, MOK, MPP5, MRC1, TMEM236, MYO9A, NACA, NAV2, NEK7, NLRP1, NRCAM, NRXN1, OR6C74, OR7G1, PBX4, PLG, PLXNA4, POFUT2, PPP1R11, PSORS1C3, PPK, RAP1GDS1, RASA3, RGL3, RHBDD1, SETX, SFTPA1, SH3PXD2B, SLC38A8, SLC44A1, SLIT3, SMURF2, SPSB4, ST8SIA2, STK32B, SYNJ1, TET2, TNXB, TRDN, TRIM26, TRPM8, TSPAN12, VPS13B, WDR36, WDR70, WWOX, ZNRD1, ZNRD1ASP
Осетины (С)	ANHX, CDYL, EFCAB6, EFCAB6-AS1, ENPP7, GPLD1, HLA-DQA2, HLA-DQB2, LOC101928622, MYT1L, OPCML, PRKAR1B, PSMB8-AS1, TNXB, TTL1, TWIST2, UHRF1BP1, ULK2
Русские	A4GALT, COLEC11, EFCAB6, EFCAB6-AS1, FAM20C, GRK1, LIMK2, LINC00452, LINC00565, MRC1, NAA25, PLCB4, RNASEH1, SNX11, TMEM255B, TNRC6B
Рутульцы	A4GALT, ARHGAP44, CCDC186, EFCAB6, EFCAB6-AS1, EPB41L4A, FAM20C, FHIT, LINC01844, LOC100129636, LOC100131257, MRC1, TMEM236, OR2J1, PLCB4, RNASEH1, RNASEH1-AS1, RPTOR, TMC01, TMEM131L, TMEM255B, TTL1, TWIST2, WT1, ZNF311, ZNF536, ZNF778
Осетины (Ю)	ABO, ADAMTS9, AGL1, ANP32A, ARHGEF11, ATP2B1, BST2, BZW1, C3orf79, LINC02006, CFAP70, CHD5, COL22A1, COL4A4, CTIF, CXADR, CYP20A1, DDR1, DSCAM, DSG2, EFCAB6, EFCAB6-AS1, EPHYC, FAM20C, FGF14, GUCY2C, HAO1, HCG17, HCG18, HUNK, IL1RAP, LGA1S8, LINC01250, LINC02691, LOC100505502, ZEB1, LOC100507065, LOC648987, LRRFIP1, MALRD1, MARCO, MELK, MIR548AU, MRC1, NAV2, NRG3, NRXN3, PHF21A, PLEKHG4B, PLXNA4, PPAT, PXDNL, RABL3, RNASEH1, RNF121, RSPO4, SHQ1, SLC12A1, SLC38A10, SLC9A9, SPINK6, TENM2, TMPRSS11D, TMTC2, TSBP1, TSBP1-AS1, ULK2, VPS13B, XIRP2, ZMIZ1-AS1, ZNF804B

Продолжение таблицы А.7

Популяции	Гены
Табасараны	<i>EFCAB6-AS1, EXD3, FAM20C, FBN1, GRIK1, HLA-F-AS1, LINC02484, MIR2052HG, MPPED1, NRG3, PIGG, RNASEH1, RNASEH1-AS1, RPRD2, SETX, SLC2A9, TMEM255B, TRPM8, TWIST2</i>
Татары (К)	<i>ABCC1, COLEC11, DACH1, EPB41L3, FAM20C, INMT-MINDY4, LINC00504, LOC101928622, MRC1, TMEM236, OR2H1, PIKFYVE, RABL3, RNASEH1, SEMA3C, SLC2A9, TMEM236, TLL1, WWOX, XIRP2, ZBTB38, ZNF12</i>
Татары (Т)	<i>ADH5, AK4, ALMS1, APOD, ARHGAP26, BDNF-AS, BEND3P3, COLEC11, DOCK8, EFCAB6, ELMO1, GAS6, GPATCH1, HSD17B12, LOC101927822, LPIN1, LRP6, LRRC36, MPPED1, MRC1, TMEM236, NLRP1, NOL8, NOX5, PAPPA2, RNF121, RUFY3, SEMA5A, SENP7, SLC25A12, SLC38A10, SLC44A1, TMEM255B, TMPRSS11E, TMTC2, TOX2, XIRP2, ZBTB38, ZNF804A</i>
Тиндинцы	<i>ABCA1, ABI1, ACVR1, APLP2, BZW1, CRLS1, DIS3, DNAH14, DTX4, DUT, EBF3, EFCAB6, EXOC6B, FAM20C, FAT3, GAS6-AS1, HUNK, LOC101928622, LOC285766, MIR99AHG, MPPED1, MRC1, NBAS, NLN, NPSR1-AS1, PACSIN2, PROX1-AS1, RNASEH1, RSPH3, SLC12A1, SMYD3, SNTG1, ST3GAL3, SUCLA2, SYNE2, TACC2, TMEM255B, TRAPPC10, TSBP1-AS1, ZNF12</i>
Цахуры	<i>COLEC11, ENPP7, FLOT1, ISX-AS1, LOC100507065, PAN3, RNASEH1, RNASEH1-AS1, RPS7, TMEM255B, TMPRSS11E</i>
Цезы	<i>ADAT1, EDC4, EDIL3, GAS6, P3H2, RTTN, SLC22A15, TMPRSS11D</i>
Тувинцы	<i>ADH5, CUEDC1, FAM20C, GAS6-AS1, LIMS1, MOK, RNASEH1-AS1, SLC2A9, SULT1C4, ZBTB38</i>
Удэгейцы	<i>ACSL1, ANKS1B, CSMD3, DOCK10, LEKR1, LOC374443, MAP3K7, MPPED1, MRAP2, MST1R, OR11A1, PCDHGA8, PPP1R15A, PRKAR1B, RBMS3, SLC25A26, TMEM255B, VDR, ZNF398, ZNF804A</i>
Удмурты	<i>A4GALT, AFDN, CDSN, COLEC11, ELMO1, FAM20C, GAS6, IL6ST, LINC01497, LINC02538, MRC1, OPA1, PRDM16, PRKAR1B, RALYL, RNASEH1, RPS7, SRPK2, TENT4A</i>
Удмурты (Б)	<i>ADI1, COLEC11, ELMO1, FAM20C, GAS6, GAS6-AS1, ISX-AS1, LINC01339, LOC643542, PTPRM, RNASEH1, RNASEH1-AS1, RPS7, TMEM117, TMEM255B, TMEM260, TMPRSS11E</i>
Удмурты (Ш)	<i>A4GALT, AKR1C8P, ARID4B, CADM1, CCAR2, CDKN2B-AS1, CHD2, DSCR10, EFCAB6, ESR2, FAM104A, FAM20C, FERMT1, GHR, LEMD2, LINC00400, LOC101927066, LOC441204, MPP6, MUC16, NALCN, OR51I1, PREP, PRMT8, RNASEH1-AS1, RPS7, RSPO3, SCMH1, SLC30A8, SNTG1, SORCS1, SPOCK3, TMEM117, TMEM255B, TMTC2, TOX2, ZNF12, ZNF331, ZNF341</i>
Узбеки	<i>ABCA4, ANO3, APOL1, ATP2C1, CDH13, CDIP1, CPXM2, DMRT1, ELAVL3, ELMO1, ERC1, EXOC6B, FAM20C, FBXW8, GALNTL6, HMOX2, HYDIN, IFT140, ISX-AS1, JAKMIP2-AS1, LINC01844, LOC102724084, LOC400655, LOC730100, MPPED1, MRC1, TMEM236, MSR1, MYO10, NAALADL2, NMRAL1, NPAS3, NPSR1-AS1, NPTN, NRG3, NRXN1, OTOGL, PAAF1, PRKAR1B, PRMT8, PRSS1, PSMA8, RASA3, RIPK1, SETD5, SH3RF1, SLC12A1, SLC30A9, SPATA13, STARD13, TLE4, TMEM255B, TSBP1-AS1, VAV2, VWA3A, ZNF717</i>
Вепсы	<i>AGBL4, ARFGAP3, CLSTN1, COLEC11, CSMD1, EFCAB6, EFCAB6-AS1, EXD3, GRIK2, IL12RB2, IL6ST, IRS1, KIAA0319L, LINC00702, LOC100996583, MRC1, NACA, PISD, RNASEH1, RPS7, SRPK2, TMEM255B, TTN, UBN1, ZMYM4</i>
Якуты	<i>LCORL, MRC1, SACM1L</i>

Таблица А.8 Сумма средних длин сегментов IBD (>1.5 cM) между парами индивидов.

	Алтайцы (Б)	Алтайцы (К)	Буряты (А)	Буряты (К)	Чукчи	Чулымцы	Эвенки (З)	Эвенки (Я)	Калмыки	Кеты	Хакасы (Ш)	Хакасы (Т)	Ханты (К)	Ханты (Р)	Коряки	Нивхи	Татары (Т)	Тувинцы	Удэгейцы	Якуты
Алтайцы (Б)	467,2																			
Алтайцы (К)	68,8	410,7																		
Буряты (А)	12,8	17,7	149,3																	
Буряты (К)	13,4	17,2	55,4	312,4																
Чукчи	2,5	3,4	3,9	3,7	785,8															
Чулымцы	14,9	14,9	12,2	11,3	4,6	883,5														
Эвенки (З)	14,1	18,1	30,3	25,0	12,4	22,0	547,7													
Эвенки (Я)	16,4	20,5	35,6	37,0	12,9	25,7	114,4	548,3												
Калмыки	11,7	13,8	19,6	20,5	2,1	7,6	14,2	17,6	45,2											
Кеты	14,5	16,4	12,9	13,2	8,1	52,5	36,6	31,3	7,6	573,4										
Хакасы (Ш)	15,6	17,2	16,4	16,5	3,6	30,7	17,1	21,7	12,7	43,5	284,8									
Хакасы (Т)	43,1	24,5	9,5	10,3	2,1	20,0	10,9	13,5	7,9	24,9	29,0	722,9								
Ханты (К)	9,6	8,0	7,1	7,0	6,6	24,3	18,7	15,1	4,7	49,3	11,1	8,4	616,5							
Ханты (Р)	9,2	9,3	7,4	7,7	6,8	29,9	19,3	16,3	4,7	61,1	12,3	10,0	204,1	795,7						
Коряки	2,8	3,4	4,7	4,3	152,6	5,3	15,5	16,9	2,9	8,6	3,8	2,8	6,8	6,2	702,9					
Нивхи	2,1	2,9	4,4	4,2	11,0	2,9	11,4	9,2	3,0	3,2	2,9	1,7	1,9	1,6	22,2	766,5				
Татары (Т)	7,6	8,2	6,5	7,4	1,8	15,2	8,2	9,2	6,7	11,4	8,8	7,0	11,3	11,9	2,0	1,4	249,3			
Тувинцы	36,0	55,0	19,3	19,7	5,0	23,9	27,0	27,5	13,9	27,4	21,8	22,6	12,0	13,0	5,5	2,7	8,0	342,3		
Удэгейцы	7,5	9,9	17,6	17,1	8,0	9,0	43,0	37,1	10,7	14,1	9,5	6,0	8,0	8,3	13,7	31,6	4,8	13,2	548,4	
Якуты	15,0	19,8	36,9	40,2	6,3	23,0	119,1	249,1	18,7	27,2	19,5	13,4	12,7	13,8	7,9	4,9	9,0	26,3	30,2	687,3

Таблица А.9 Сумма средних длин сегментов IBD (1-4 сМ) между парами индивидов.

	Алтайцы (Б)	Алтайцы (К)	Буряты (А)	Буряты (К)	Чукчи	Чулымцы	Эвенки (З)	Эвенки (Я)	Калмыки	Кеты	Хакасы (Ш)	Хакасы (Т)	Ханты (К)	Ханты (Р)	Коряки	Нивхи	Татары (Т)	Тувинцы	Удэгейцы	Якуты
Алтайцы (Б)	100,0																			
Алтайцы (К)	27,4	83,9																		
Буряты (А)	11,6	15,6	83,9																	
Буряты (К)	12,1	15,0	31,3	103,9																
Чукчи	2,5	3,3	3,6	3,4	433,6															
Чулымцы	13,5	13,2	11,0	10,0	4,3	202,6														
Эвенки (З)	13,1	17,0	25,0	20,8	11,8	20,7	140,6													
Эвенки (Я)	14,9	18,4	28,5	28,0	11,7	24,5	71,0	207,7												
Калмыки	9,7	11,6	16,8	17,1	2,0	7,0	12,9	15,7	31,2											
Кеты	13,1	15,1	12,1	12,6	7,9	43,3	32,6	28,4	7,0	177,9										
Хакасы (Ш)	13,1	14,3	14,4	14,0	3,5	23,3	15,9	19,7	11,0	25,2	73,3									
Хакасы (Т)	24,8	16,2	8,7	9,3	2,1	17,9	10,5	12,5	7,0	19,3	18,3	212,9								
Ханты (К)	8,4	7,7	6,7	6,7	6,4	22,6	17,8	14,6	4,5	44,6	10,6	7,9	250,0							
Ханты (Р)	8,8	8,9	7,0	7,4	6,6	27,9	18,4	15,2	4,6	52,3	11,4	9,4	113,2	258,6						
Коряки	2,6	3,1	4,4	4,1	116,7	5,2	14,4	15,4	2,8	8,1	3,7	2,6	6,6	6,0	399,2					
Нивхи	1,9	2,7	4,0	4,0	10,4	2,6	10,5	8,0	2,8	3,2	2,6	1,6	1,8	1,5	21,4	449,9				
Татары (Т)	5,8	6,4	6,0	6,6	1,7	12,2	7,7	8,7	5,8	10,2	7,2	6,3	10,4	11,0	1,9	1,4	32,6			
Тувинцы	21,6	31,0	16,7	16,7	4,6	21,9	24,7	25,1	11,7	25,3	18,2	16,7	11,3	12,3	5,2	2,5	7,1	129,9		
Удэгейцы	7,3	9,2	15,2	14,8	7,5	8,7	34,4	27,6	9,9	12,9	8,9	5,3	7,8	8,0	12,6	26,6	4,6	12,1	142,9	
Якуты	13,4	17,9	28,1	28,9	6,1	21,4	67,4	113,2	16,3	25,1	17,6	12,0	12,2	13,1	7,5	4,3	8,4	23,5	20,7	267,8

Таблица А.10 Сумма средних длин сегментов IBD (4-10 сМ) между парами индивидов.

	Алтайцы (Б)	Алтайцы (К)	Буряты (А)	Буряты (К)	Чукчи	Чулымцы	Эвенки (З)	Эвенки (Я)	Калмыки	Кеты	Хакасы (Ш)	Хакасы (Т)	Ханты (К)	Ханты (Р)	Коряки	Нивхи	Татары (Т)	Тувинцы	Удэгейцы	Якуты
Алтайцы (Б)	161,8																			
Алтайцы (К)	26,9	125,2																		
Буряты (А)	1,1	2,0	52,1																	
Буряты (К)	1,3	2,0	18,7	119,5																
Чукчи	0,0	0,1	0,2	0,3	204,3															
Чулымцы	1,3	1,3	0,8	1,0	0,1	277,4														
Эвенки (З)	0,8	1,1	4,5	3,7	0,6	1,2	116,3													
Эвенки (Я)	1,4	2,0	6,3	8,1	1,1	1,2	35,8	206,7												
Калмыки	1,7	2,1	2,7	3,2	0,1	0,5	1,2	1,9	11,5											
Кеты	1,2	1,2	0,8	0,6	0,1	8,7	3,7	2,7	0,5	172,4										
Хакасы (Ш)	2,2	2,6	1,9	2,1	0,1	6,7	1,0	1,8	1,6	8,2	93,9									
Хакасы (Т)	15,1	7,0	0,7	1,0	0,0	2,1	0,4	1,0	0,9	4,2	8,1	283,3								
Ханты (К)	1,0	0,3	0,4	0,3	0,2	1,7	0,8	0,4	0,2	4,6	0,5	0,5	176,9							
Ханты (Р)	0,4	0,4	0,3	0,3	0,2	2,0	0,8	0,9	0,2	7,8	0,9	0,6	56,1	200,8						
Коряки	0,1	0,2	0,3	0,2	33,2	0,1	1,0	1,3	0,0	0,4	0,1	0,2	0,2	0,2	148,8					
Нивхи	0,1	0,2	0,4	0,2	0,6	0,3	0,8	1,2	0,2	0,1	0,3	0,1	0,1	0,1	0,8	180,8				
Татары (Т)	1,5	1,7	0,5	0,8	0,1	2,8	0,5	0,5	0,8	1,1	1,4	0,7	0,8	0,8	0,1	0,1	56,2			
Тувинцы	10,9	19,0	2,2	2,6	0,4	1,9	2,2	2,3	1,9	1,9	3,4	4,8	0,7	0,7	0,2	0,2	0,8	121,2		
Удэгейцы	0,2	0,7	2,0	1,9	0,5	0,3	7,5	8,0	0,8	1,1	0,5	0,6	0,3	0,3	1,0	4,2	0,2	5,8	132,7	
Якуты	1,5	1,8	7,6	9,6	0,2	1,6	41,8	106,3	2,3	2,2	1,8	1,3	0,4	0,7	0,4	0,5	0,5	2,7	7,5	268,2

Таблица А.11 - Сумма средних длин сегментов IBD (>10 cM) между парами индивидов.

	Алтайцы (Б)	Алтайцы (К)	Буряты (А)	Буряты (К)	Чукчи	Чулымцы	Эвенки (З)	Эвенки (Я)	Калмыки	Кеты	Хакасы (Ш)	Хакасы (Т)	Ханты (К)	Ханты (Р)	Коряки	Нивхи	Татары (Т)	Тувинцы	Удэгейцы	Якуты
Алтайцы (Б)	205,4																			
Алтайцы (К)	14,5	201,6																		
Буряты (А)	0,0	0,2	27,9																	
Буряты (К)	0,0	0,2	5,4	89,0																
Чукчи	0,0	0,0	0,0	0,0	147,9															
Чулымцы	0,1	0,1	0,1	0,0	0,0	403,0														
Эвенки (З)	0,2	0,0	0,7	0,5	0,0	0,1	290,9													
Эвенки (Я)	0,1	0,1	0,7	0,9	0,1	0,0	7,6	133,9												
Калмыки	0,3	0,1	0,1	0,1	0,0	0,1	0,1	0,0	2,6											
Кеты	0,2	0,1	0,1	0,0	0,0	0,5	0,3	0,2	0,1	223,0										
Хакасы (Ш)	0,3	0,3	0,1	0,3	0,0	0,8	0,1	0,2	0,1	10,1	117,6									
Хакасы (Т)	3,2	1,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,4	2,5	226,6								
Ханты (К)	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,1	0,0	0,0	189,7							
Ханты (Р)	0,0	0,0	0,1	0,0	0,0	0,1	0,0	0,1	0,0	1,0	0,0	0,0	34,8	336,1						
Коряки	0,0	0,0	0,1	0,0	2,7	0,0	0,1	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	154,9				
Нивхи	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	113,6			
Татары (Т)	0,3	0,1	0,0	0,0	0,0	0,3	0,0	0,0	0,1	0,0	0,3	0,0	0,1	0,0	0,0	0,0	159,6			
Тувинцы	2,9	5,0	0,4	0,2	0,0	0,0	0,1	0,1	0,2	0,1	0,2	0,9	0,0	0,0	0,0	0,0	0,1	87,7		
Удэгейцы	0,0	0,0	0,3	0,4	0,0	0,0	1,0	1,4	0,0	0,1	0,1	0,0	0,0	0,1	0,1	0,8	0,0	0,1	251,5	
Якуты	0,1	0,0	1,2	1,6	0,0	0,0	9,9	29,6	0,1	0,0	0,1	0,1	0,0	0,0	0,1	0,1	0,0	0,1	1,9	148,7

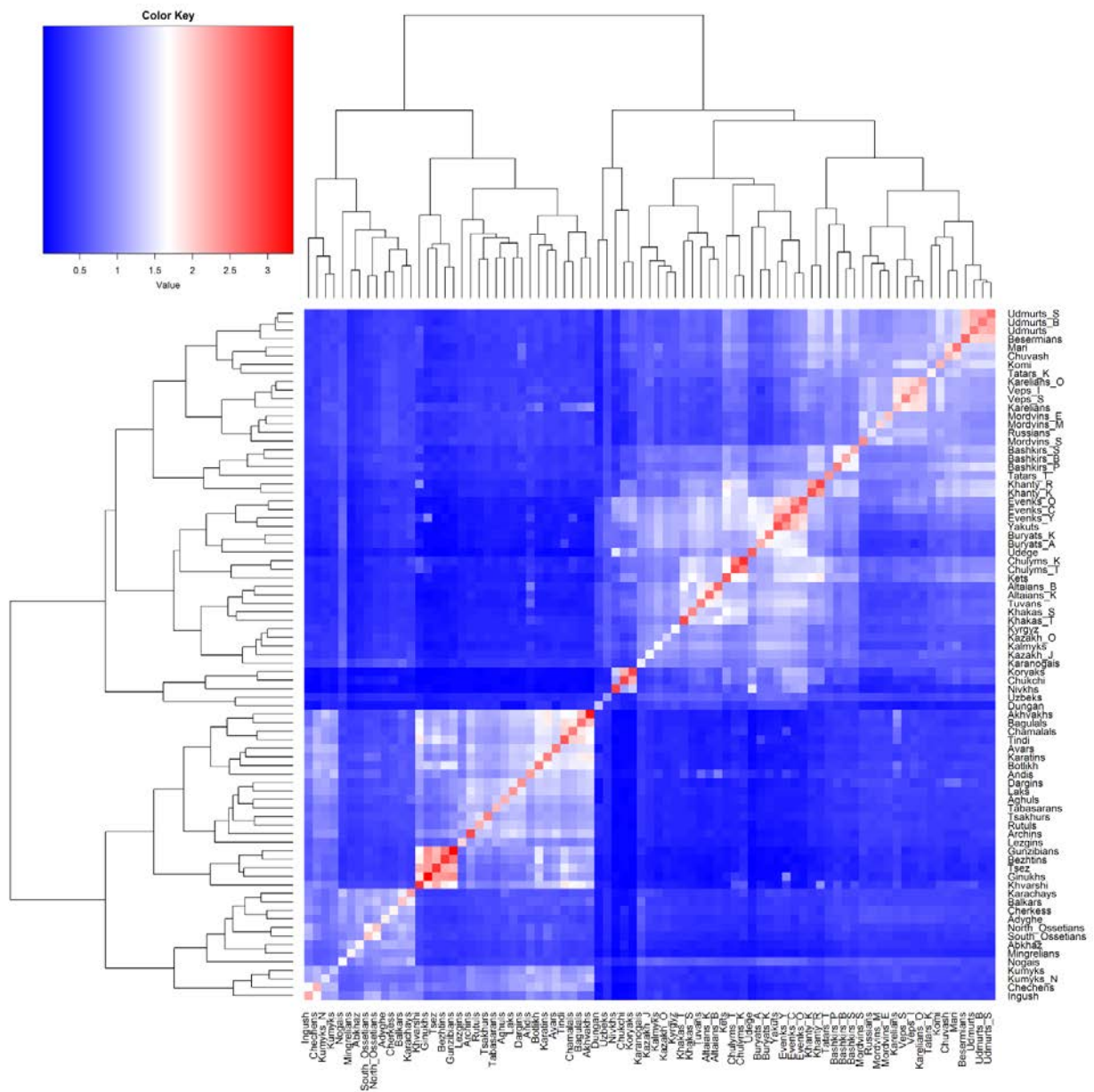


Рисунок А.1 - Тепловая карта с дендрограммой на основании логорифма от суммы средних длин сегментов IBD (1,5-4 сМ) между парами индивидов.

БЛАГОДАРНОСТИ

Выражаю глубокую признательность научному руководителю, доктору биологических наук, ведущему научному сотруднику лаборатории эволюционной генетики Харькову Владимиру Николаевичу.

Выражаю огромную благодарность директору Томского НИМЦ и руководителю лаборатории эволюционной генетики, академику РАН, доктору биологических наук, профессору Вадиму Анатольевичу Степанову за предоставленную возможность выполнения диссертационной работы, критические замечания и мудрые советы, помощь при планировании, выполнении и обсуждении результатов исследования.

Хочу поблагодарить сотрудников лаборатории эволюционной генетики, за конструктивные замечания и рекомендации, всестороннюю помощь и поддержку во время выполнения исследования.

Также признателен сотрудникам Научно-исследовательского института медицинской генетики за ценные советы при обсуждении результатов работы.

Благодарю ученого секретаря диссертационного совета, кандидата биологических наук Ирину Юрьевну Хитринскую за методическую помощь в подготовке документов к защите.