

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**

**Автореферат  
диссертации на соискание ученой степени  
кандидата биологических наук**

Томск - 2022

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Томский национальный исследовательский медицинский центр Российской академии наук» (Томский НИМЦ), в Научно-исследовательском институте медицинской генетики, г. Томск.

**Научный руководитель:**

доктор биологических наук

**Харьков Владимир Николаевич**

**Официальные оппоненты:**

**Лавряшина Мария Борисовна**, доктор биологических наук, доцент; Федеральное государственное образовательное учреждение высшего образования «Кемеровский государственный медицинский университет» Министерства здравоохранения Российской Федерации, заведующая кафедрой молекулярной и клеточной биологии (г. Кемерово).

**Хидиятова Ирина Михайловна**, доктор биологических наук, профессор; Институт биохимии и генетики - обособленное структурное подразделение Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук, главный научный сотрудник лаборатории молекулярной генетики человека (г. Уфа).

**Ведущая организация:**

Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук (г. Москва).

Защита состоится «\_\_\_\_\_» \_\_\_\_\_ 2022 года в \_\_\_\_\_ часов на заседании диссертационного совета 24.1.215.03 на базе Федерального государственного бюджетного научного учреждения «Томский национальный исследовательский медицинский центр Российской академии наук» по адресу г. Томск, Московский тр., 3, Научно-исследовательский институт медицинской генетики Томского НИМЦ.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного научного учреждения «Томский национальный исследовательский медицинский центр Российской академии наук», адрес сайта <http://tnimc.ru/>

Автореферат разослан «\_\_\_\_\_» \_\_\_\_\_ 2022 года

Ученый секретарь  
диссертационного совета,  
кандидат биологических наук

Хитринская И.Ю.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**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 (Jeong C. et al., 2019; Gopalan S. et al., 2022;). 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 (Prahlow J.A. et al., 2017, Грачева А.С. и др., 2020;).

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 (Palamara P.F. et al., 2012;).

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 (B. L. Browning & Browning, 2011; Kivisild T., 2021; Finke K. et al., 2021;). 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 (Albrechtsen A. et al., 2010; Han L. & Abney M., 2011;). Сегменты IBD могут использоваться для оценки демографической истории, включая эффект бутылочного горлышка и потоки генов в популяциях (Gusev A. et al., 2012; Sikora M. et al., 2019; Ferrando-Bernal M., et al., 2020;). 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 (Hochreiter S., 2013;).

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 (Pemberton T.J. et al., 2012; Ceballos F.C. et al., 2018).

Выявленные закономерности распределения RON в геноме человека на полногеномном уровне позволяют лучше понять генетико-демографическую историю популяций, оценить уровень инбридинга, выявить регионы подверженные отбору (Biagini S. A. et al., 2019). Данные по распределению RON в популяциях могут служить основой для улучшения специфичности идентификации регионов, связанных с заболеваниями, в контексте картирования гомозиготности (Ropers H.H., 2007; Steinhaus R. et al., 2022).

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 (Hancock A.M. et al., 2010; Cardona A. et al., 2014; Чередниченко А.А. и др., 2015; Степанов В.А. и др., 2019; Rocha J., 2020) 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 (Derenko M. et al., 2007; Duggan A.T. et al., 2013; Malyarchuk B.A. et al., 2012; Харьков В.Н. и др., 2008, 2013, 2020, 2021).

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 (Mathieson I. et al., 2015). 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 (Pakendorf B. et al., 2006; Starikovskaya E.B. et al., 2005; Tamm E. et al., 2007; Харьков В.Н. и др., 2021; Gubina M.A. et al., 2022) и Y-хромосомных маркеров (Derenko M. et al., 2007; Grugni V. et al., 2019; Karmin M. et al., 2015; Харьков В.Н. и др., 2013, 2020, 2021)), 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 (Flegontov P. et al., 2016; Sikora M. et al., 2019; Triska P. et al., 2017). 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 (Sikora M. et al., 2019).

**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 (RON) 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 (RON).
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.** 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 использовании 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.

#### **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  $ROH > 1,5$  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 материалов диссертации.** Основные положения диссертационной работы были представлены на следующих конференциях: Международная научная конференция молодых ученых «Актуальные проблемы медицинской генетики» (Томск, 2018); Международная конференция «VII Съезд Вавиловского общества генетиков и селекционеров, посвященный 100-летию кафедры генетики СПбГУ, и ассоциированные симпозиумы» (Санкт-Петербург, 2019); Всероссийский Конгресс молодых ученых «Актуальные вопросы фундаментальной и клинической медицины» (Томск, 2020);.

**Личное участие 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.

## **МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ**

Материал исследования составили образцы ДНК, выделенные из лейкоцитов крови, принадлежащих к 20 популяциям коренного населения Сибири включающих 15 этнических групп. На рисунке 1 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).

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. Харьковским), 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).

С характеристикой представленных популяций Сибири и популяций Средней Азии, Волго-Уральского региона, Восточной Европы, Северного Кавказа и Дагестана можно ознакомиться в тексте диссертации.

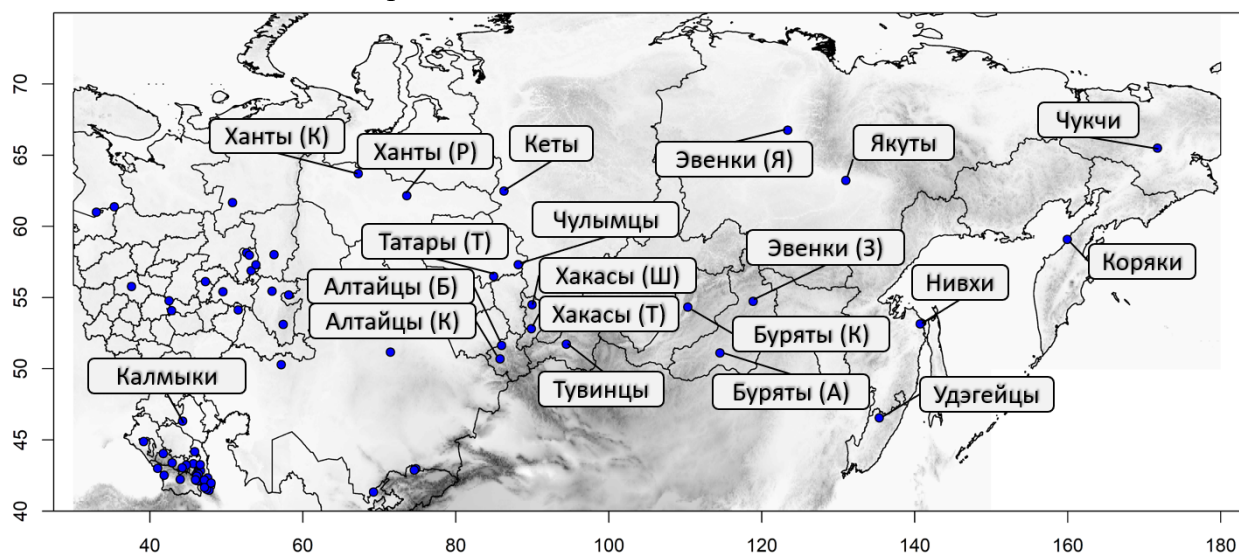


Рисунок 1 - Локализация исследованных этнических выборок коренного населения Сибири.

**Подготовка материала для генотипирования.** Выделение геномной ДНК из периферической крови проводилось с помощью метода фенол-хромофорной экстракции.

**Генотипирование.** Полногеномное генотипирование ДНК-маркеров было проведено с использованием платформы Infinium Multi-Ethnic Global BeadChip (Illumina). Платформа Infinium Multi-Ethnic Global BeadChip включает свыше 1,7 миллиона SNP-маркеров. Кластеризацию массива генотипов и контроль качества выполняли с использованием протокола, опубликованного в статье Guo Y. (Guo et al., 2014) с использованием GenomeStudio (модуль генотипирования v2.0.3), программного пакета, который Illumina разработала для анализа полученных геномных данных.

**Оценка качества генотипирования.** 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 аутосомных 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.

**Фазирование.** Фазирование генотипов проведено с использованием программного обеспечения Beagle 4.1 (S. R. Browning & Browning, 2007).

**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) (Menozzi P. et al., 1978; Patterson N. et al., 2006). Для выполнения задачи по выявлению и оценке индивидуальных компонентов и количества примесей в генофонде изучаемых популяций была использована программа Admixture (Alexander D.H. et al., 2009). Для визуальной дифференцировки и оценки кластеров была использован метод стохастического вложения соседей с t-распределением (англ. t-distributed Stochastic Neighbor Embedding, t-SNE) (Maaten L., 2008).

**Выявления общих по происхождению блоков сцепления (IBD-сегментов).** Для анализа идентичных по происхождению блоков генома использовали алгоритм Refined IBD (B. L. Browning & Browning, 2013). 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.

**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» (Purcell S. et al., 2007) мы удалили SNP с высоким неравновесием по сцеплению (LD). 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).

**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).

**Идентификация регионов высокой гомозиготности с помощью программы GARLIC.** Для оценки возраста ROH применяли программу GARLIC (Szpiech Z.A. et al., 2017), 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.

**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 A.1), при этом, если расстояние между SNP было меньше 300 т.п.н., регионы были объединены в один регион.

**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).

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

**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. 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. 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. Начиная от  $K=8$  в общем массиве анализируемых популяций, отчетливо выделяется компонент, практически полностью доминирующий у монголоязычных бурятов (63%) и калмыков (54%). Наиболее четко он проявляется при  $K=12$  (рисунок 1). 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.

При задании числа предковых компонент больше 7 в большинстве южносибирских популяций выявляется специфичный для них генетический компонент, наиболее отчетливо проявляющийся на анализируемом массиве популяционных выборок при  $K = 10$ , 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%. Около трети генофонда тувинцев 28,4% и практически четверть генофонда хакасов качинцев 24,6% имеют в своем генофонде элементы енисейского компонента. 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.

Европеоидный компонент среди сибирских популяций выявлен у томских татар 40,3%, и группы индивидуумов внутри популяции эвенков, что также подтверждается наличием длинных блоков IBD европейского происхождения и значительного снижение длины R<sub>OH</sub> 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) (доминирует у хантов и составляет 95%). 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%.

**Оценка распределения регионов высокой гомозиготности с помощью программы «Plink».** Общая длина и количество R<sub>OH</sub> для каждого индивидуального генома колеблется от 41,2 до 687,7 м.п.о. и от 61 до 394, соответственно. Медианное число R<sub>OH</sub> на индивидуальный геном в исследованных образцах составило 135,4, граница первого квартиля 121,8, граница третьего квартиля 151,5. Медианная длина R<sub>OH</sub> 128,4 м.п.о., граница первого квартиля 103,7 м.п.о. третьего 139,3 м.п.о.. Среднее значение совокупной длины R<sub>OH</sub> 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 R<sub>OH</sub> на человека наблюдается у узбеков (103,7), максимальное – среди коряков 312.

Доля R<sub>OH</sub> 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 R<sub>OH</sub> 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 R<sub>OH</sub> 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 R<sub>OH</sub>, 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 ROH (15.6%). Это полностью соответствует полученным ранее данным о наибольшем генетическом разнообразии и многокомпонентности генофонда среднеазиатских популяций в сравнении с другими этносами Северной Евразии.

**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 сибирских чукчей, коряков, нивхов для всех классов длин наблюдается значительное увеличение суммарной длины и количества ROH на человека по сравнению с другими популяциями, что говорит о продолжительном инбридинге, а также эффекте «бутылочного горлышка» (Рисунок 2).

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 (Харьков В. Н. и др., 2008, 2013). 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. (Харьков В. Н. и др., 2013, 2014). 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. 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 ROH добавляются удэгейцы и ханты. Отражая высокий уровень инбридинга внутри популяции, популяция кетов в распределении длинных ROH попадают в группу с эвенками(Я), якутами, хантами, чулымцами.

Распределение ROH связано с размером популяции и их эффективной численностью, при этом относительно небольшие популяции Сибири имеют ROH большего количества и размера, чем большие по численности популяции этих регионов. Инбридинг добавляет небольшое количество очень длинных ROH для тех, кто является потомком кровнородственных браков, увеличивая дисперсию суммы ROH, видимую как сдвиг влево на графике длин ROH по сравнению с количеством ROH. 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).

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 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 ROH 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 ( $F_{ROH}$ ).

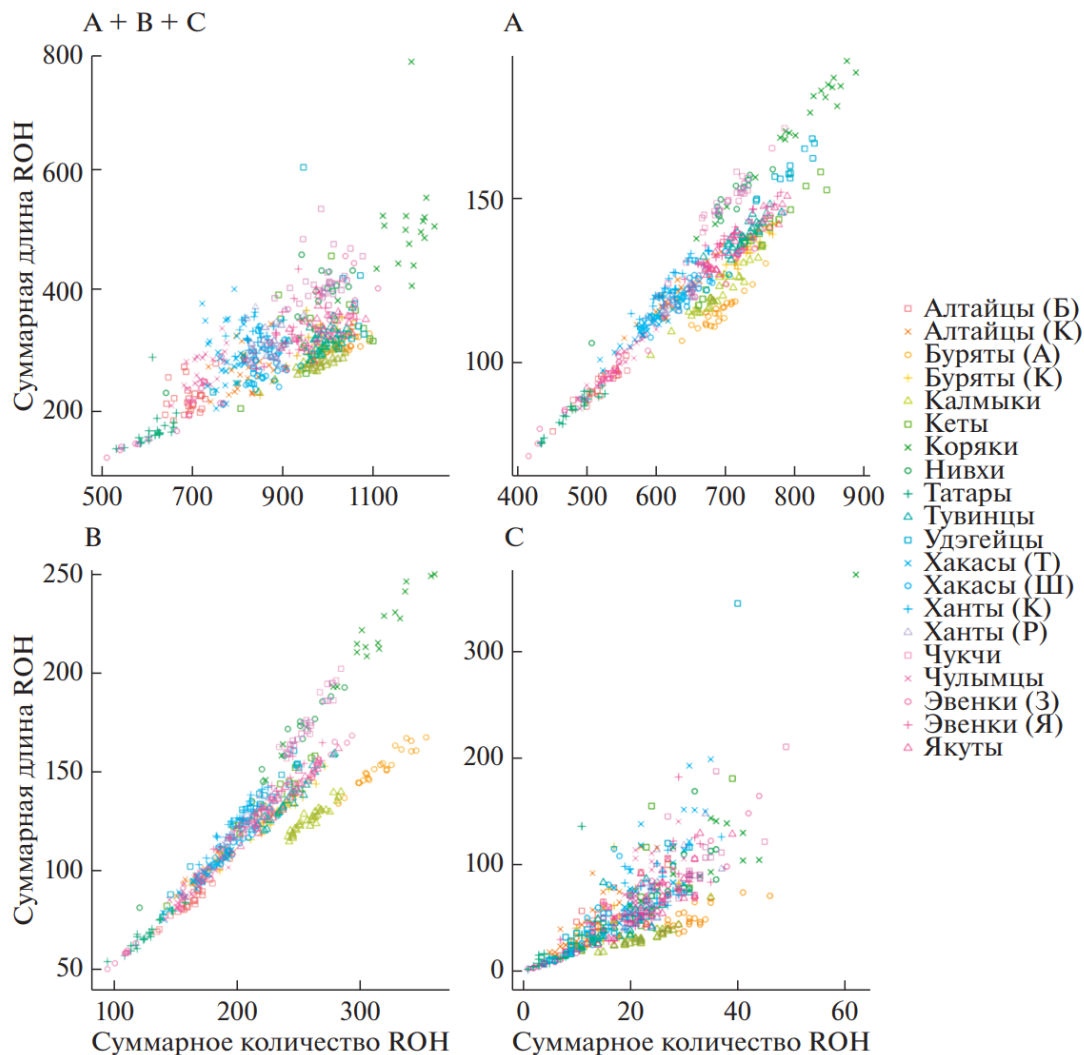


Рисунок 2 - Отношение суммарной длины ROH к суммарному количеству ROH в исследуемых популяциях Сибири.

**Оценка уровня инбридинга  $F_{ROH}$  в исследуемых популяциях.** 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 1 диссертационного исследования.

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).

Несмотря на то, что  $F_{\text{РОН}} > 5$  и  $F_{\text{РОН}} > 1,5$  в популяциях Дагестана значительно превышает значения  $F_{\text{РОН}}$  в популяциях Сибири, по более коротким участкам гомозиготности ( $F_{\text{РОН}} > 0,5$ ) наблюдаются схожие значения. Это может свидетельствовать о значительных эффектах инбридинга для этих популяций Сибири в прошлом с постепенным уменьшением доли кровнородственных браков к настоящему времени.

Наблюдается высокая корреляция для  $F_{\text{РОН}} > 1,5$  с суммой средних длин сегментов  $\text{IBD} > 1,5$  cM между парами индивидов в сибирских популяциях ( $r = 0,9246$ ,  $p < 5,612 \times 10^{-9}$ ).

**«Горячие точки» РОН.** При анализе распределения частот РОН для всех популяций было найдено 56 регионов, содержащих «горячие точки». Среди 56 регионов 19 включают гены, совпадающие с данными из предыдущих исследований «горячих точек» РОН: 20 регионов для Дагестанских популяций (Karafet T.M. et al., 2015), 3 для европейских (Nothnagel M. et al., 2010) и 1 регион для мировых популяций (Pemberton T.J. et al., 2012). 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 (Степанов В. А. и др., 2019). Для кросс-популяционного теста (ХР-ЕНН) 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 РОН (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 хр-ehh.

Также было выявлено, что некоторые из этих генов несут сигналы направленного отбора в коренных популяциях Сибири (Cardona A. et al., 2014) 8 генов по тесту PBS (SLAIN2, SELP, F5, CCDC146, FGB, PLRG1, HLA-DPB1, HLA-DPA1) и 3 гена по тесту iHS (GCC2, LIMS1, SNTG1).

**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).

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 3).

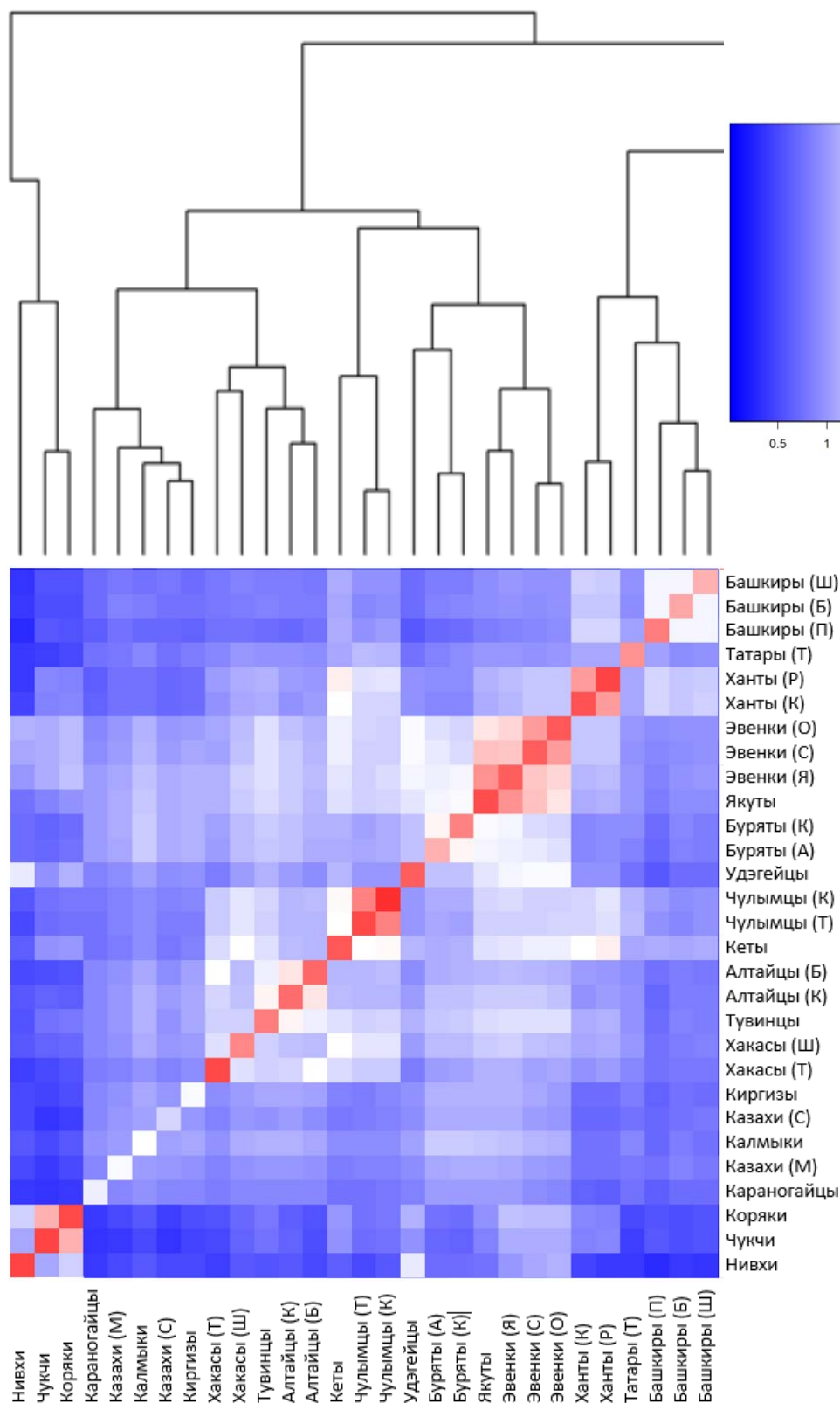


Рисунок 3 - Тепловая карта с дендрограммой на основании логарифма от суммы средних длин сегментов IBD ( $>1.5$  cM) между парами индивидов.

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 cM), чем с забайкальскими эвенками (114,1 cM). 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 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 сМ (46–42%), что говорит о недавнем сильном инбридинге внутри популяции.

Популяции, характеризующиеся как наиболее гетерогенные среди сибирских популяций и имеющие минимальные значения по  $F_{\text{ST}}$ , также имеют минимальные значения по сумме средних длин сегментов IBD. Наиболее выражено это у популяций калмыков и томских татар.

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%). 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.

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.

**Поиск сигналов направленного отбора с помощью тестов на протяженную гомозиготность гаплотипов.** В результате поиска сигналов, сформированных под действием направленного отбора в изучаемых популяциях коренного населения Сибири, с помощью теста nSL было найдено 375 SNP ( $\text{fdr} \geq 0,05$ ), попадающих в функциональные участки 284 генов, и 363 SNP, расположенных в межгенных участках. 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.

Основными молекулярными путями были пути, отвечающие за синаптическую пластичность: долговременная потенцияция (LTP) и долговременная депрессия (LTD) – 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.

По результатам тестов nSL и iHS 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  $P = 0,0435$ ).

Всего было найдено 86 категорий онтологии генов в биологических процессах. 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, GPD5, TRPC6, MCPH1, AUTS2, ARID1B, APOD, DLG1, LRFN2, ITSN2, SYNE2, NTRK3, PARK2, PLXNA2, LRRN2, TNFR, 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, CH13L1, 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*).

Определенные функциональные полиморфизмы в ферментах *ADH*, которые вызывают различную эффективность превращения этанола в ацетальдегид, могут быть защитными. Голдман и Енох (Goldman D. & Enoch M.A., 1990) 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 (Racimo F., 2016).

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 (Grossman S.R. et al., 2010; Racimo F., 2016; Sabeti P.C. et al., 2007). 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 (Racimo F., 2016). Для популяций коряков, тувинцев (*nSL*, *iHS*) и алтайцев (К) (*iHS*) наблюдаются высокие значения по гену *LIMS1*, белок которого участвует в регуляции выживаемости, пролиферации и дифференцировки клеток.

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*) (Cardona A. et al., 2014). 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 (Р) – 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 (Р) 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 (Dauber et al., 2016). Интересно, что у 9 из 20 популяций по тесту *iHS* 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 (Tagliabracci V.S. et al., 2012). Мутации в этом гене ответственны за развитие синдрома Рейна (OMIM: 259775, неонатальная остеосклеротическая дисплазия костей).

**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).

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*, *WVOX*, *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 (Chaplin G. & Jablonski N.G., 2009; Jablonski N.G., 2012). Кроме того, в различных европейских популяциях было показано участие генетических вариантов *SLC45A2* в восприимчивости к меланоме (Ibarrola-Villava M. et al., 2012). 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 (Sakurai K. et al., 2011). Положительный отбор по *SLC45A2* особенно интересен, поскольку сцепленные вредные SNP могут быть связаны с заболеваниями человека (Chun S. & Fay J.C., 2011). 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 (Nijbroek G. et al., 1995; Schrijver I. et al., 1999), а один был обнаружен у пациентов с синдромом Марфана или родственными фенотипами (Liu W.O. et al., 1997).

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  $F_{\text{ROH}}$ . 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 ROH 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 ROH 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 ROH выше, чем в остальных исследуемых популяциях Средней Азии, Восточной Европы и Дагестана. Количество и общий суммарный размер коротких и средних ROH показывают

более высокие значения для исследуемых популяций Сибири. Однако у коренного населения Дагестана, несмотря на меньшее количество коротких ROH, существенный вклад вносят более протяженные регионы высокой гомозиготности, что свидетельствует о сильном инбридинге. Среди коренного населения Сибири максимальные количество и длина ROH 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, iHS 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 XP-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 ROH на индивида по сравнению с другими географическими регионами. Практически для всех популяций Сибири сумма длинных ROH 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  $ROH > 1,5$  м.п.о. обладают популяции нивхов ( $F_{ROH} = 0,0389$ ), чукчей ( $F_{ROH} = 0,0431$ ) и коряков ( $F_{ROH} = 0,0446$ ), 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 РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

### Статьи в журналах и сборниках

1. Степанов В.А., Харьков В.Н., Вагайцева К.В., Хитринская И. Ю., Бочарова А.В., **Колесников Н.А.**, Зарубин А.А., Попович А.А., Марусин А.В., Спиридонова М.Г., Триска П., Татарина Т.В. Сигналы направленного отбора в популяциях народов Сибири и европейской части России // Генетика. - 2019. - Т. 55 - № 10. - С. 1198-1207.
2. **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.
3. **Колесников Н. А.**, Харьков В. Н., Раджабов М.О., Зарубин А. А., Литвинов С. С., Екомасова Н. В., Хуснутдинова Э. К., Степанов В. А. Различия по уровню инбридинга в популяциях Дагестана: анализ регионов высокой гомозиготности // Медицинская генетика. - 2020. - Т. 19. - № 7. - С. 21-22
4. **Колесников Н.А.**, Харьков В.Н., Зарубин А.А., Раджабов М.О., Воевода М.И., Губина М.А., Хуснутдинова Э.К., Литвинов С.С., Екомасова Н.В., Штыгашева О.В., Максимова Н.Р., Сухомясова А.Л., Степанов В.А. Особенности геномного распределения регионов высокой гомозиготности у коренного населения Северной Евразии на индивидуальном и популяционном уровнях на основе анализа SNP высокой плотности. Генетика. - 2021. - Т. 57. - № 11. - С. 1261-1275.
5. **Колесников Н.А.**, Харьков В.Н., Зарубин А.А., Воевода М.И., Губина М.А., Штыгашева О.В., Максимова Н.Р., Сухомясова А.Л., Степанов В.А. Сигналы направленного отбора в популяциях коренного населения Сибири. Генетика. - 2022. - Т. 58. - № 4. - С. 470-474.

### Материалы конференций

1. **Колесников Н.А.**, Зарубин А.А., Харьков В.Н., Степанов В.А. Особенности геномного распределения регионов высокой гомозиготности у коренного населения Сибири. Актуальные вопросы фундаментальной и клинической медицины: сборник материалов конгресса молодых ученых. 24-25 мая 2018 г., Томск. 102-104 с.
2. **Колесников Н.А.**, Харьков В.Н., Раджабов О.М., Степанов В.А. Распределение регионов высокой гомозиготности в геномах коренного населения Северной Евразии. Сборник тезисов: «VII Съезд Вавиловского общества генетиков и селекционеров, посвященный 100-летию кафедры генетики СПбГУ, и ассоциированные симпозиумы». 18-22 июня 2019 г., Санкт-Петербург, Россия)
3. **Колесников Н.А.**, Харьков В.Н., Зарубин А.А., Степанов В.А. Различия по уровню инбридинга в популяциях Сибири: анализ регионов высокой гомозиготности / Актуальные вопросы фундаментальной и клинической медицины: сборник материалов конгресса молодых ученых, 19–20 ноября 2020 г. [Электронный ресурс] / под ред. В.А. Степанова, Е.Л. Чойнзонова, С.В. Попова, Н.А. Бохана, В.В. Жданова; Томский национальный исследовательский медицинский центр Российской академии наук. – Томск: Изд-во Том. ун-та, 2020. – 534 с.

### СПИСОК СОКРАЩЕНИЙ

SNP - (Single-nucleotide polymorphism) однонуклеотидный полиморфный вариант  
 м.п.о. – миллион пар оснований (1000000 пар нуклеотидов)  
 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  
 $F_{ROH}$  – коэффициент геномного инбридинга  
 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

**Просьба высылать отзывы на автореферат по адресу:**

НИИ медицинской генетики Томского НИМЦ  
634050, г. Томск, ул. Набережная реки Ушайки, 10  
Ученому секретарю Диссертационного совета 24.1.215.03  
канд. биол. наук Хитринской И.Ю.  
Факс +7(3822)51-37-44  
E-mail: khitrinskaya@tnimc.ru