**Введение­­**

Изучение взаимоотношений различных этнических групп на протяжении эволюционной истории человека является одним из ключевых направлений популяционной генетики и основывается на анализе особенностей геномного разнообразия отдельных популяций. Обширная территория Сибири и Дальнего Востока и низкая численность коренного населения способствовала формированию значительной территориальной обособленности. Коренное население, зачастую, развивалось в условиях длительной генетической изоляции от основных миграционных потоков на территории Евразии, что делает его генофонд особенно интересным для изучения этапов заселения человеком современного типа не только Сибири и Дальнего Востока, но и Нового Света. В виду этого, коренные народы очень разнообразны по антропологическому и языковому составу, а также этнической истории. Так, нивхи – малый коренной народ нижнего течения реки Амур и Сахалина на протяжении нескольких столетий поддерживал свой уникальный язык и культуру в относительной изоляции от соседских популяций.  
 На сегодняшний день накоплено достаточное количество материала о генетическом составе популяций данного региона (ссылка). Однако полученные результаты не могут точно описать реальную структуру генофонда и взаимоотношения популяций Сибири. Главным образом это связано с проблемой репрезентативности малочисленной выборки по отношению к всему генофонду изучаемых народов. В следствии чего по – прежнему остаются вопросы о в кладе различных предковых в современный генофонд Сибири и Дальнего Востока.

Для получения наиболее достоверных результатов необходимо объединить как можно больше доступных генетических данных по изучаемым этносам из разных источников. Информативной и удобной маркерной системой для характеристики структуры генофондов современных популяций и анализа эволюции популяций человека являются однонуклеотидные полиморфизмы (SNP) (Kim S., Misra A. 2007).

Цель работы: на основе разнообразия генофондов по однонуклеотидным полиморфизмам осуществить реконструкцию генетических взаимоотношений между коренными этническими группами Сибири и Дальнего Востока (какие).

Задачи:

1. Объединить данные высокоразрешающего генотипирования ядерного генома с литературными данными

**Обзор литературы**

**1. Генетическое разнообразие населения восточной Сибири и Дальнего Востока**

В последнее десятилетие, внимание исследователей было сосредоточено на реконструкции процессов происхождения человека и его расселение по земному шару[1],[2]. Однако сейчас становятся все более актуальными работы, связанные с конкретизацией эволюционных процессов отдельных этносов (ссылка). Современный этнический состав коренного населения сложился в результате длительных исторических событий, зафиксированных археологами, этнографами и лингвистами. Однако генетическое разнообразие генофонда малых коренных народов Сибири и Дальнего Востока является главной характеристикой их эволюционного процесса.

# Датировки заселения Сибири анатомически современными людьми, основанные на костных останках, найденных вблизи села Усть-Ишим в западной Сибири, оцениваются в 45 тысяч лет назад (Fu et al,2014). Популяции Северо-Восточной Сибири отделились от основной линии других Восточно-Азиатских популяций около 8.8–11.2 (Reconstructing genetic history of Siberian and Northeastern European populations, 2016)

Тысяч лет назадБольшое количество исследований, направлено конкретно на изучение истории Сибирского населения основываются на анализе митохондриальной ДНК (мтднк) и Y-хромосомных маркеров (Володько и соавт. 2008; Малярчук и соавт. 2011; Derenko и соавт. 2012; Dulik и соавт. 2012; Малярчук и соавт. 2012; Sukernik и соавт. 2012; Duggan et al. 2013; Федорова и соавт. 2013; Dryomov et al. 2015). Однако в последнее время становится более актуальный анализ ядерного генома коренных малых народов восточной Сибири: рассматривается не только генетическая история народов Северо-Восточной Сибири, а также на адаптации представителей Сибирских популяций к холоду и богатому мясной пищей рациону (fedorova), (Cardona), (Clemente et al. 2014). В исследования показано генетическое сходство популяций Центральной и Южной Сибири, в то время как население Северо-Восточной Сибири содержит в своем генофонде уникальные генетические компоненты и отражает сигналы региональной селекции.

**2. Однонуклеотидный полиморфизм (SNP)**

Однонуклеотидный полиморфизм (SNP) - вариабельные однонуклеотидные позиции генома человека, в которых частота минорного аллеля составляет не менее 1% (Brookes 1999).

Для каждой позиции генома можно заффиксировать замену на любой из четырех возможных нуклеотидов, однако на практике би-аллельные встречаются чаще всего (менее 0.1% трех-аллельных полиморфизмов от всех SNP человека (Lai 2001)). Среди генетических вариаций, аллели встречающиеся у более чем 0,01% человечества индексируются и заносятся в базу данных (dbSNP NCBI USA). Геном человека содержит 3,2 миллиарда пар нуклеотидов, в котором на 1000 оснований приходится 1 SNP ( Luger K, Mäder AW, Richmond RK et al. Crystal structure of the nucleosome core particle at 2.8 A resolution // Nature. 1997). Одиночные полиморфизм встречаются как в кодирующих, так и в некодирующих облостях генома и могут отвечать за определенные фенотипические признаки или представлять нейтральную вариацию, которая может быть использована при анализе полиморфизма в контексте эволюционных процессов (Kwok et al., 1996)([https://cdn.intechopen.com/pdfs-wm/28889.pdf](https://vk.com/away.php?to=https%3A%2F%2Fcdn.intechopen.com%2Fpdfs-wm%2F28889.pdf&cc_key=)).  
  
Испоьзование SNP, как наиболее многочисленных и стабильных маркеров [Evaluation of coverage variation of SNP chips for genome-wide association studies], позволяет анализировать большое число аллелей в различных областях медицины [Microarray-based SNP genotyping to identify genetic risk factors of triple-negative breast cancer (TNBC) in South Indian population], популяционной экологии [Genetic variation and clonal diversity in introduced populations of Mimulus guttatus assessed by genotyping at 62 single nucleotide polymorphism loci] и эволюционной генетике [Genomic insights into the origin of farming in the ancient Near East]. Также повсеместное использование однонуклеотидных полиморфизмов обусловлено одновременным развитием технологий в смежных областях исследования (анализ генома и связанных с ним технологий, биоинформатика и вычислительная техника, популяционной генетика человека на глобальном уровне).

Днк-микрочипы основанные на использовании SNP как молеклярно-генетические маркеров применяют для выявления генетических вариаций между представителями различных этносов [Selecting SNPs to Identify Ancestry]. Частота встречаемости определенного полиморфизма в конкретном генотипе зачастую отличается у представителей различных народов, что может указывать на медецинские особенности определенных этносов. Кроме того, доказано влияние однонуклеотидных полиморфизмов на устойчивость организма к внешним воздействиям среды (Sachidanandam R, Weissman D, Schmidt SC et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms // Nature. 2001)

Большинство готовых днк-микрочипов были разработаны для медицинских генетических экспериментов и не рассчитаны на глобальные популяционные исследования. Это послужило причиной для разработки ДНК-чипов для реконструкции эволюционной истории человечества [4], [axiom]. Получение все большего количества данных, основанных на генотипировании SNP различных этносов, позволяет выявить достоверную генетическую дифференциацию и отнести конкретных людей к определенным популяциям, а также помогает реконструировать маршруты распространения людей по земному шару.

Большинство исследований в области популяционной генетике человека были посвящены, лавным образом, европейскому населению [Ancient human genomes suggest three ancestral populations for present-day Europeans], [Discerning the Ancestry of European Americans in Genetic Association Studies], в то время как малочисленные популяции по всему миру нуждаются в дополнительном генетическом анализе.

**3. Биоинформатический анализ геномных данных**

**3.1. Современные технологии днк-микрочипов**

В последние 10 лет технологий высокопроизводительного секвенирования (NGS) стали мощными, гибкими и широко распространёнными методами в различных областях биологии и медицины. В эволюционной биологии благодаря NGS – технологиям стали возможными широкомасштабные геномные исследования на популяционном уровне[5]. Однако использование ДНК-микрочипов (DNA microarray) обходится дешевле и требует меньше вычислительных ресурсов по сравнению с NGS- секвенирования. Первоначально разработанные для определения уровня экспрессии генов, микрочипы являются удобным аналитическим инструментом для определения геномных особенностей человека [6]. ДНК-микрочип содержит на своей поверхности зонды – специфические олигонуклеотиды или протяженные фрагменты ДНК/РНК, способные гибридизоваться с кодирующей ДНК или матричной РНК по принципу комплементарного связывания нуклеотидов. Для подтверждения связывания гибридизации ДНК-чип сканируется до тех пор, пока не будет получен весь паттерн гибридизации. На площади чипа в несколько квадратных сантиметров помещаются десятки тысяч элементов, что позволяет сконструировать матрицы для эффективного проведения индивидуальных исследований. Сейчас подобные технологии являются бурно развивающимся направлением - биологические фирмы предлагают микрочипы, содержащие до нескольких сотен тысяч зондов [ссылка на illlumina, affymetrix].

На сегодняшний день в исследованиях широко распространены 2 типа ДНК-чипов в зависимости от природы зондов. В первом случае ДНК-зонды, синтезируются непосредственно на поверхности чипа (in situ). Второй тип объединяет в себе независимо синтезированные зонды, которые наносят на поверхность чипов(ссылка). По технологии изготовления ДНК-чипы можно разделить на: печатные (printed), фотолитографические (in situ synthesized) и высокоплотного кварцевого нанесения (high-density bead arrays) (Кулемин).

Первыми и наиболее известными днк-чипами высокой плотности (high-density microarrays), изготовленные по In situ технологии, являются чипы фирмы Affymetrix (ссылка в вк последняя). Технология Affymetrix GeneChip разработана в 1990-х годах командой Стива Фодора, основанная на комбинации фотолитографии и методов комбинаторной химии [7]. Данная технология обеспечивает высокую плотность размещения зондов на поверхности кварцевого стекла при анализе точечных мутаций, однонуклеотидных полиморфизмов и анализе экспресси генов [Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada CG. The affymetrix GeneChip platform: an overview. Methods Enzymo].

Технологическая платформа Illumina представляет чипы высокоплотного кварцевого нанесения (технология BeadChips). На поверхности чипа располагаются кремниевых бусины с закрепленными олигонуклеотидными зондами. Каждая бусина кроме специфического набора олигонуклеотидов имеет также и кодирующий элемент, для ее однозначного идентификации.

При работе с микрочипами используется технология APEX (arrayed primer extension) [Tebbutt SJ. Genotyping of single nucleotide polymorphisms by arrayed primer extension. Methods Mol Biol., Pullat J, Metspalu A. Arrayed primer extension reaction for genotyping on oligonucleotide microarray. Methods Mol Biol]. Данный подход комбинирует метод полимеразной реакции и гибридизации. Олигонуклеотид, закрепленный на чипе, комплементарен целевой молекулы и меньшее нее ровно на один нуклеотид (SNP). При добавлении к гибридизованным комплексам (кодирующая ДНК- олигонуклеотид) ДНК-полимеразы и меченых ддНТФ происходит присоединение к олигонуклеотиду-приймеру меченого дидезоксинуклеотидтрифосфата, который комплементарен нуклеотиду в кодирующей ДНК. Нуклеотид, вступивший в реакцию определяется по цвету флуорисцентной метки, тем самым определяя SNP в данном сайте. Далее, чип сканируется на яркость и цвет свечения для определения статуса полиморфизма.

Проект HapMap[8],основываясь на распределении блоков неравновесия по сцеплению, позволил отобрать максимально информативные полиморфные маркеры генома человека, доказав, что генотипирование каждого SNP – избыточно.

Основным критерием при выборе коммерческого чипа является общее покрытие (global coverage) генома. Тем не менее данная характеристика не обеспечивает одинаковую плотность покрытия на протяжении всего генома и общее количество маркеров не может выступать единственным критерием при выборе днк-чипа [Evaluation of coverage variation of SNP chips for genome-wide association studies].

В таблице 1 приведена краткая информация о наиболее часто используемых днк-биочипах в области популяционной генетике человека [9],[10],[3],[11].

|  |  |
| --- | --- |
| Название биочипа | Особенности |
| Infinium OmniExpress-24 BeadChip | * Технология сканирования: HiScan, [iScan](https://emea.illumina.com/content/illumina-marketing/emea/en_GB/systems/array-scanners/iscan.html) * Количество маркеров: фиксировано – 713,599 шт., дополнительно – 30000 шт. * Особенности: олигонуклеотиды на чипе подобраны таким образом, чтобы охватить наибольшее количество распространенных ОНП. |
| Infinium OmniExpress-12 BeadChip | * Технология сканирования: HiScan, [iScan](https://emea.illumina.com/content/illumina-marketing/emea/en_GB/systems/array-scanners/iscan.html) * Количество маркеров: фиксировано – 730,525 шт., дополнительно – 200000 шт. * Особенности: возможность включения большого числа специфических маркеров при генотипировании |
| Human660W-Quad BeadChip | * Технология сканирования: HiScan, [iScan](https://emea.illumina.com/content/illumina-marketing/emea/en_GB/systems/array-scanners/iscan.html) * Количество маркеров: фиксировано –657366 шт. * Особенности: обеспечивает высокое покрытие широко распространенных SNP |
| Human1M-Duo BeadChip | * Технология сканирования: HiScan, [iScan](https://emea.illumina.com/content/illumina-marketing/emea/en_GB/systems/array-scanners/iscan.html) * Количество маркеров: фиксировано –1140419 шт., * Особенности: обеспечивает высокое покрытие вдоль всего генома |
| Genome-Wide Human SNP Array 6.0 | * Технология сканирования: GeneChip® Scanner 3000 7G * Количество маркеров: фиксировано – 934946 шт., * Особенности: выокая специфичность связывания молекул, высокая плотность покрытия |

По мере того, как стоимость секвенированию и генотипированию продолжает падать, объем генерируемых данных увеличивается, все большее значение преобретает правильный подход к хранению и обработке данных[12].

**3.2 Методы анализа ядерной ДНК**

**3.2.1 Метод главных компонент**

Метод главных компонент (Рrincipal component analysis (PCA)) – широко распространенный метод в популяционной генетике для визуализации данных, позволяющий уменьшить их размерность, потеряв наименьший объем информации. Вместо начального количества исходных признаков вычисляется такое же количество новых признаков – главных компонент. Первая главная компонента скоррелирована с исходными данными. Следующая компонента характеризует вторую по значимости долю изменчивости и при этом не имеет корреляции с первой компонентой. Данный метод может быть использован при анализе нескольких сотен исходных признаков, который сводится к рассмотрению первых двух-трех главных компонент. Подробные детали методы описаны в статье [Patterson et al., 2006].

**3.2.2 Метод предковых компонент (**Admixture)

Анализ, демонстрирующий смешанность состава популяции на основе данных генотипа. Используя данные о известных генотипах и количестве предполагаемых предковых популяций, программа строит свою модель того, какой вклад внес каждый предполагаемы предок в уже известные популяции. Чем больше вклад каждой предполагаемой предковой популяции в индивида, тем больше общего количество SNP имеет индивид с представителями предполагаемого предка.

**3.2.3 Анализ неравновесного сцепления генов (ALDER)**

ALDER (Admixture-induced Linkage Disequilibrium for. Evolutionary Relationships) в своем анализе популяционной истории использует оценку двух параметров: рекомбинации и неравновесия по сцеплению. Рекомбинация – процесс обмена генетическим материалом между разными молекулами ДНК. Неравновесие по сцеплению – явление передачи нескольких участков ДНК одним блоком, которые формируются неодинаково в разных популяциях из-за наследования разных комбинаций сегментов ДНК. Основа метода заключается в выявлении специфических блоков участков ДНК для каждой популяции и на оценке доли общих сегментов в определенной выборке популяций. ALDER определяет возможность того, что две выбранные группы являются предковыми по отношению к тестируемым популяциям. Метод позволяет отследить время смешения, анализируя доли рекомбинации на одно поколение. Стоит отметить, что ALDER-статистика имеет смысл только при достаточном количестве образцов (n > 20) и при анализе событий, имевших место меняя 5500 лет назад [Loh et al., 2013; Pickrell et al., 2014].