

## ВЗАИМОДЕЙСТВИЕ ГЕНОФОНДОВ РУССКОГО И ФИННОЯЗЫЧНОГО НАСЕЛЕНИЯ ТВЕРСКОЙ ОБЛАСТИ: АНАЛИЗ 4 МЛН SNP-МАРКЕРОВ

О. П. Балановский<sup>1,2,3</sup> ✉, И. О. Горин<sup>1,2</sup>, Ю. С. Записецкая<sup>2</sup>, А. А. Голубева<sup>2</sup>, Е. С. Кострюкова<sup>4</sup>, Е. В. Балановская<sup>2,3</sup>

<sup>1</sup> Институт общей генетики имени Н. И. Вавилова, Москва, Россия

<sup>2</sup> Медико-генетический научный центр, Москва, Россия

<sup>3</sup> Биобанк Северной Евразии, Москва, Россия

<sup>4</sup> Федеральный научно-клинический центр физико-химической медицины, Москва, Россия

Генофонды популяций Тверской области (русских и карел) изучены по широкогеномной панели из 4 млн аутосомных SNP-маркеров, типированной на суммарной выборке из 41 образца. Эти данные по популяции тверских карел ( $n = 11$ ) и русских из западных, центральных и восточных районов Тверской области ( $n = 30$ ) проанализированы на широком фоне русских популяций соседних областей, карел Карелии и других популяций Северо-Восточной Европы с целью изучения феномена взаимопроникновения генофондов славянского и финноязычного населения. Такое изучение генофондов населения России по наиболее обширной из существующих широкогеномных панелей важно для каталогизации геномного разнообразия населения России и характеристики региональных генофондов и имеет практическое применение в фармакогеномике и судебной медицине. Методами главных компонент, ADMIXTURE, d- и f3-статистик показано, что генофонд тверских карел, несмотря на их проживание среди преобладающего русского населения в течение 3–5 веков и 20-кратное сокращение численности в течение последнего столетия, сохраняет наибольшую близость к генофонду карел Карелии. Но при этом генофонд тверских карел более сходен с русским генофондом, чем генофонд других карельских популяций. Сближение генофондов русских и карел Тверской области происходит за счет более интенсивного потока генов от русских к карелам и при малозаметном потоке генов от карел к русским: тверские русские оказались столь же генетически отличны от карел, как, например, псковские. Сходство тверских карел с карелами Карелии по аутосомным маркерам (при небольшом смещении в сторону русского генофонда) согласуется с опубликованными данными по Y-хромосоме (отсутствие детектированного смещения тверских карел с русскими).

**Ключевые слова:** широкогеномные панели, SNP-маркер, чип Illumina, генофонд, карелы, русские, Тверская область, центральная Россия

**Благодарности:** мы благодарим всех доноров образцов, которые принимали участие в данном исследовании, АНО «Биобанк Северной Евразии» за предоставление коллекций ДНК и члена-корреспондента РАН В. В. Напольских за консультации при интерпретации результатов.

**Финансирование:** исследование выполнено при финансовой поддержке Министерства науки и образования РФ (Госконтракт # 011–17 от 26.09.2017) в рамках научно-технической программы Союзного государства «ДНК-идентификация» (работы по генотипированию), Государственного задания Министерства науки и высшего образования РФ для Медико-генетического научного центра им. Н. П. Бочкова (биоинформатический анализ данных), гранта Российского фонда фундаментальных исследований № 20-09-00479 а (анализ генеалогической информации, интерпретация результатов, написание текста).

✉ **Для корреспонденции:** Олег Павлович Балановский  
ул. Губкина, д. 3, г. Москва, 119991; balanovsky@inbox.ru

**Статья получена:** 10.10.2020 **Статья принята к печати:** 27.10.2020 **Опубликована онлайн:** 25.11.2020

**DOI:** 10.24075/vrgmu.2020.072

## INTERACTIONS BETWEEN GENE POOLS OF RUSSIAN AND FINNISH-SPEAKING POPULATIONS FROM TVER REGION: ANALYSIS OF 4 MILLION SNP MARKERS

Balanovsky OP<sup>1,2,3</sup> ✉, Gorin IO<sup>1,2</sup>, Zapisetskaya YuS<sup>2</sup>, Golubeva AA<sup>2</sup>, Kostryukova EV<sup>4</sup>, Balanovska EV<sup>2,3</sup>

<sup>1</sup> Vavilov Institute of General Genetics, Moscow, Russia

<sup>2</sup> Research Centre for Medical Genetics, Moscow, Russia

<sup>3</sup> Biobank of North Eurasia, Moscow, Russia

<sup>4</sup> Federal Research and Clinical Center of Physical and Chemical Medicine, Moscow, Russia

This study explored the gene pools of Russian and Karelian populations of Tver region. Forty-one samples representing Tver Karels ( $n = 11$ ) and Russians residing in the Western, Central and Eastern districts of Tver region ( $n = 30$ ) were genotyped using a genome-wide panel of 4,559,465 SNPs. In order to investigate the phenomenon of genetic admixture between Slavic and Finnish-speaking populations, the obtained results were compared to the data on the Russian populations inhabiting the neighboring territories, Karels from Karelia and other North Eastern Europeans. Studying the gene pools of Russian populations with a genome-wide SNP panel is essential for cataloging their genetic diversity and identifying the distinct features of regional gene pools; in addition, it provides valuable data for practical pharmacogenomics and forensics. Using the principal component analysis, the ADMIXTURE method and D- and f3-statistics, we demonstrated that the gene pool of Tver Karels is closest to the gene pool of Karelian Karels, despite a long (300 to 500 years) history of living among the larger Russian population and the twentyfold population decline during the 20th century. At the same time, the gene pool of Tver Karels exhibits more pronounced similarity to the gene pool of the studied Russian populations than does any other Karelian population. The genetic admixture between Tver Russians and Tver Karels occurred due to a more intense gene flow from Russians to Karels whereas the gene flow from Karels to Russians was much weaker: Tver Russians turned out to be as genetically different from Karels as Pskov Russians. The genetic similarity of Tver Karels to Karelian Karels assessed with the autosomal SNP panel exhibits a slight shift towards the Russian gene pool and is consistent with the previously published analysis of Y-chromosome lineages in these populations that detected no admixture between Tver Karels and Russians.

**Keywords:** genome-wide genotyping, SNP, Illumina array, gene pool, Karelians, Russians, Tver region, Central Russia

**Acknowledgement:** we thank all the donors who took part in this study, the Biobank of North Eurasia for DNA collections and Napolskikh VV, the corresponding member of RAS, for his contribution to data interpretation.

**Funding:** the study was supported by the Russian Ministry and Science and Higher Education (Government Contact # 011–17 dated September 26, 2017). Genotyping and manuscript preparation were done under the *DNA-based identification* Research and Technology Project of the Union State. Bioinformatic analysis and interpretation of the obtained results were carried out under the State Assignment of the Russian Ministry of Science and Higher Education for Bochkov Research Centre for Medical Genetics.

✉ **Correspondence should be addressed:** Oleg P. Balanovsky  
Gubkina, 3, Moscow, 119991; balanovsky@inbox.ru

**Received:** 10.10.2020 **Accepted:** 27.10.2020 **Published online:** 25.11.2020

**DOI:** 10.24075/brsmu.2020.072

Тверь и ее окрестности, лежащие на границе центральной и северо-западной России, играли важную роль не только в истории русского народа, но и в его взаимодействии с западным финноязычным населением. До славянской колонизации этот регион населяли финно-угорские племена, в основном меря, но с середины I тысячелетия его стали интенсивно осваивать славяне. В начале XII в. в устье Тверцы был известен крупный торгово-ремесленный посад, а с середины XIII в. Тверь стала одним из трех великих княжеств Руси монгольского периода, в течение двух веков соперничала с Москвой за лидерство в объединении русских земель и являлась одним из центров, вокруг которого концентрировалось население.

В XV–XVI вв. началось переселение карел с Карельского перешейка и из Приладожья в северо-восточные районы тверских земель. В XVII в. оно стало массовым из-за событий русско-шведской войны — к 1670 г. в тверские земли переселилось 25–30 тыс. православных карел. Беженцы были расселены на государственных землях Тверской земли, опустевших из-за голода и разорений Смутного времени. Карелы формировали собственные поселения, компактно расположенные отдельно от русских деревень. Затем последовали еще несколько волн менее значительных миграций карел [1, 2]. В итоге всех переселений сформировалась этнографическая группа тверских карел, сохранявшая свой язык (карельский язык относится к финской подгруппе финно-угорской языковой группы). В 1937 г. в ареале их компактного проживания в Тверской области был создан Карельский национальный округ со столицей в Лихославле. Однако в 1939 г. он был расформирован, а активисты карельского движения арестованы. Это могло привести к смене этнической самоидентификации части тверских карел. Их численность, согласно переписям, сократилась в XX в. в 20 раз: от 150 тыс. человек в 1930 г. (причем 95% населения говорило на карельском языке) до 7 тыс. в 2010 г. [3], однако компактный ареал карел оставался прежним [4].

Проживание бок о бок в течение трех веков популяций двух этносов разного происхождения — тверских русских и тверских карел — ставит вопрос о степени взаимопроникновения их генофондов. Отчасти ответ был нами дан при анализе генофонда тверских карел по панели 49 SNP-маркеров Y-хромосомы, включившей наиболее информативные для Восточной Европы генетические линии. Результаты [5] убедительно продемонстрировали генетическое сходство по Y-хромосоме тверских карел с коренным населением северо-востока Европы, особенно с южными карелами и вепсами Карелии. Полученные результаты указывают, что популяция тверских карел, несмотря на 20-кратное сокращение численности и проживание среди многочисленного русского населения на протяжении более десяти поколений, сохранила свой предковый Y-хромосомный генофонд. Однако резкое сокращение численности тверских карел могло быть связано с изменением их самоидентификации и ассимиляцией тверских карел русским населением. В этом случае в современных популяциях русских Тверской области можно ожидать повышенную долю генетического компонента, характерного для Северо-Востока Европы (и карел в частности). Известно также, что при межэтнических браках соседних этносов Y-хромосомный генофонд более устойчив, чем аутосомный, поскольку большинство браков патрилокальны (женщины переезжают в селения мужчин), что означает миграции в отношении митохондриальной ДНК и аутосом, но отсутствие миграций в отношении

Y-хромосомы. Оба эти фактора могут привести к тому, что аутосомные генофонды тверских карел и тверских русских испытали более интенсивное взаимодействие и сблизились больше, чем по Y-хромосоме.

Изучение аутосомных генофондов различных этнических групп коренного населения по столь подробной «широкогеномной» панели маркеров важно как для каталогизации геномного разнообразия населения России, так и для накопления данных об особенностях генофондов региональных популяций. Это требуется в том числе для исследований по фармакогеномике и для определения происхождения в рамках судебно-медицинских экспертиз. Значимость таких данных для фармакогенетики связана с тем, что подавляющее большинство фармакогенетических протоколов разработано для популяций европейского происхождения, поэтому они могут работать недостаточно эффективно из-за того, что в популяциях России встречаются иные аллели, а частоты хорошо изученных аллелей значительно различаются в разных этнических группах (аналогичная ситуация показана для популяций Восточной Азии и Африки [6, 7]; исследования российских популяций по частотам фармакогенетических маркеров обобщены в недавно проведенном обзоре [8]). Данные о генофондах народонаселения приобретают значимость и для судебно-медицинских экспертиз в тех случаях, когда перед криминалистом ставится задача определить вероятное происхождение индивида по следовым количествам его ДНК, для этого существует ряд тест-систем и идет разработка новых, но принципиальным вопросом является доступность генетических данных о возможных популяциях происхождения [9, 10].

Поэтому целью данного исследования стала характеристика по широкой (4 млн SNP) панели аутосомных маркеров генофондов тверских карел и тверских русских и анализ их взаимопроникновения. Проведение такого анализа на примере Тверской области и на широком фоне генофондов других групп населения Европейской части России, служит и более общей цели — изучению феномена взаимодействия славянского и финноязычного населения.

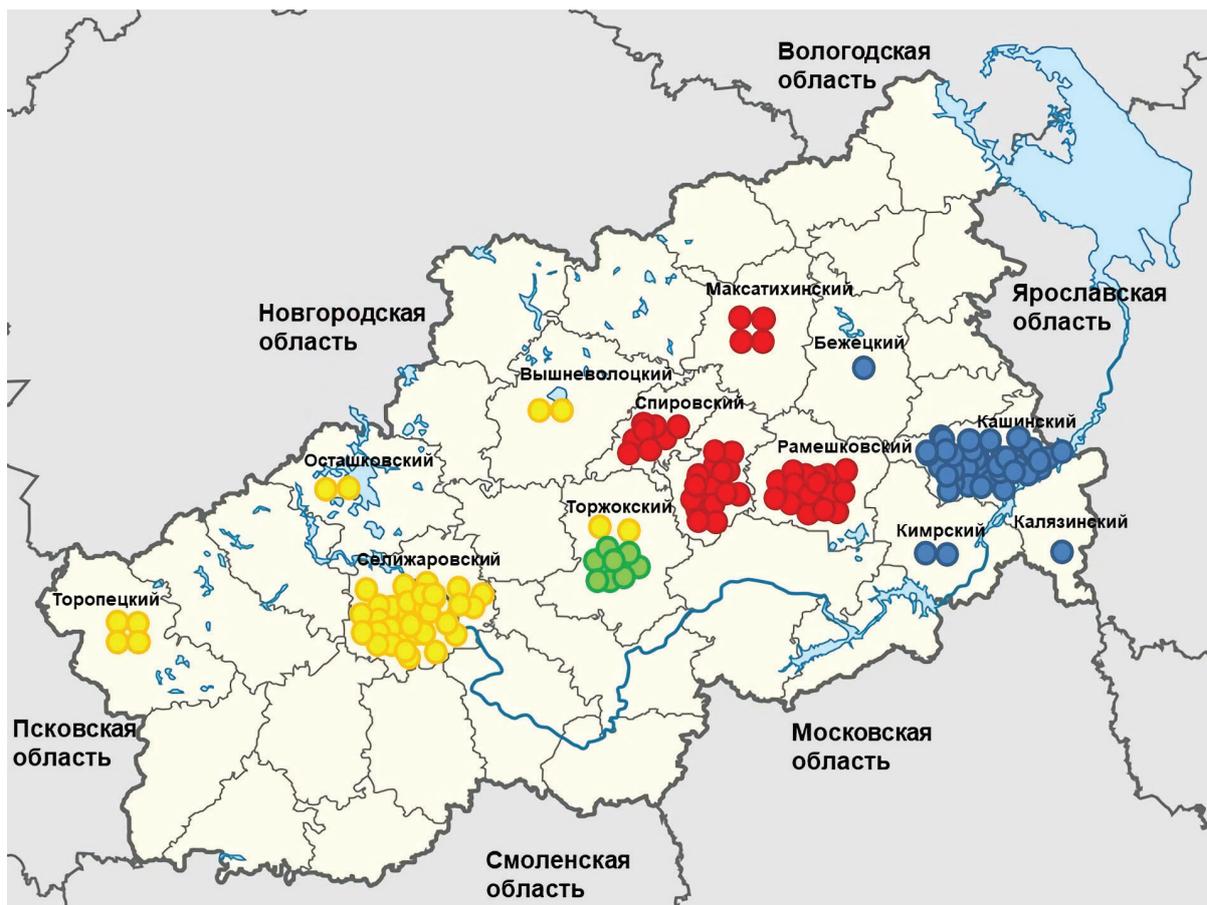
## ПАЦИЕНТЫ И МЕТОДЫ

Экспедиционное обследование и русского, и карельского населения Тверской области проведено по единой методологии, детально описанной в [11]. В исследование включены только неродственные между собой индивиды (на глубину до трех поколений, по данным анкетирования), все предки которых на глубине минимум трех поколений родились в Тверской области, причем все четверо бабушек и дедов относили себя либо к русским, либо к карелам, и не помнили, чтобы их более отдаленные предки были иной этнической принадлежности.

Критерии включения в исследование: самоидентификация четырех предков обследуемого (двух дедов и двух бабушек) как принадлежащих к данному народу; письменное информированное согласие на участие в обследовании.

Критерием исключения образцов было качество выделенной ДНК или ее концентрация, недостаточные для проведения полногеномного гентипирования.

Тверские карелы представлены в анализе аутосомного генофонда 11 индивидами, происхождение которых охватывает все центральное ядро ареала тверских карел: Лихославльский ( $n = 4$ ), Максатихинский ( $n = 1$ ), Спировский ( $n = 2$ ) и Рамешковский ( $n = 4$ ) районы Тверской области. В 1930 г. в этих четырех районах проживало 88 тыс. тверских



**Рис. 1.** Карта изученных популяций Тверской области. Кружками отмечены места рождения каждого из четырех предков (двух бабушек и двух дедов) обследованного; тверские карелы показаны *красным* цветом; восточные тверские русские — *синим*; южные тверские русские — *зеленым*; западные тверские русские — *желтым*

карел, т. е. более половины (58%,) их общей численности (в Лихославльском — 15%, Максатихинском — 19%, Спировском — 8%, Рамешковском — 16%). В 2010 г. в этих четырех районах проживало 5 тыс. тверских карел, составляющих 78% их общей численности (в Лихославльском — 36%, Максатихинском — 13%, Спировском — 15%, Рамешковском — 14%).

Общая выборка русских Тверской области, представленная в анализе аутосомного генофонда, составила 30 человек. Обследование русского населения Тверской области проводили целенаправленно для изучения взаимодействия генофондов русского и карельского населения: выборки были собраны нами так, чтобы они не перекрывали ареал расселения тверских карел, но находились вблизи. Такое планирование представляется оптимальным для определения степени потока генов от русских к карелам: при изучении отдаленных русских популяций, которые не могли непосредственно контактировать с карелами, уровень сходства тверских карел с ними мог бы быть занижен за счет различий разных русских популяций друг от друга, а при изучении русских, проживающих непосредственно в «карельских» селах, показатель сходства мог бы быть, наоборот, завышен за счет того, что русские в этих селах могли оказаться потомками карел, сменившими этническую самоидентификацию. Для дополнительного контроля были обследованы не одна, а несколько популяций русских, находящихся на разном удалении от ареала тверских карел. Восточная популяция тверских русских примыкает к ареалу тверских карел (рис. 1). В анализе аутосомного генофонда она представлена 13

индивидами — уроженцами Кашинского района Тверской области. Западная популяция тверских русских выбрана так, чтобы она находилась на большем, чем восточная, удалении от ареала тверских карел. Она представлена в анализе аутосомного генофонда 15 индивидами — уроженцами Селижаровского района Тверской области. В отдельную группу выделены два индивида из Торжокского района, находящегося прямо на юге от Лихославля, столицы тверских карел. Таким образом, суммарно в данной работе представлены генотипы 41 образца из Тверской области, изученных по широкогеномной панели маркеров. На рис. 1 указаны места происхождения каждого из четырех предков этих индивидов.

Анализ генофонда Тверской области проведен в сравнении как с русскими популяциями соседних областей (Архангельской, Вологодской, Воронежской, Курской, Новгородской, Псковской, Смоленской, Ярославской), так и с северными и южными карелами Карелии ( $n = 16$ ). Суммарно анализировали 27 геномов карел, 100 геномов русских, а также ряд других популяций Восточной Европы (белорусы, вепсы, воль, ижора, литовцы, украинцы), изученных по той же широкогеномной панели маркеров. Большинство этих популяций были ранее изучены и по маркерам Y-хромосомы [5, 12, 13].

Генотипирование всех образцов ДНК (как Тверской области, так и популяций сравнения) проводили по панели Illumina, включающей 4,5 млн SNP-маркеров. Генотипирование выполняли с использованием набора реактивов Infinium Omni5Exome-4 v1.3 BeadChip Kit (Illumina; США) на приборе iScan (Illumina; США). Первичный анализ

и оценку качества проводили в программе GenomeStudio v2011.1 (Illumina; США). Для исследованных образцов показатель CallRate составлял не менее 0,99. В результате для изученных выборок были получены генотипы по 4 559 465 маркерам.

Полученные генотипы размещены в базе данных GG-base [14] и открыты для скачивания для популяций Тверской области (RussiansTverKashin, RussiansTverSelizharovo, RussiansTverTorzhok, TverKarelians).

Для первичного общего анализа применяли классический метод главных компонент, который позволяет увидеть общую структуру изучаемых генофондов. Для более подробного изучения «генетического расстояния» между популяциями использовали анализ  $f_3$ -статистик, а анализ  $d$ -статистики — для определения направления потоков генов между изученными популяциями и образцами.

Фильтрацию данных проводили с помощью PLINK 1.9 [15, 16]. Конкретные фильтры описаны ниже для каждого метода отдельно.

Так, перед расчетом главных компонент (PCA) были исключены полиморфизмы, прочитанные менее чем у 95% образцов (geno 0,05) и встречаемые с частотой минорного аллеля менее 1% (maf 0,01); были исключены образцы, у которых прочитано менее 90% полиморфизмов (mind 0.1); исключение тесно сцепленных полиморфизмов ( $r^2 > 0,2$ ) проведено методом движущегося окна, включающего на каждом шаге 1500 полиморфных сайтов и сдвигающегося за один шаг на 150 полиморфных сайтов (indep-pairwise 1500 150 0.2). После фильтрации осталось 274 036 полиморфизмов и 127 образцов (из 131). Расчет значений главных компонент проводили с помощью утилиты smartpca из программного пакета EIGENSTRAT [17, 18] с пятью итерациями исключения outliers. Результаты работы smartpca визуализировали с помощью Python 3, в том числе с использованием библиотек pandas [19, 20], matplotlib [21] и seaborn [22].

Для анализа методом ADMIXTURE фильтрацию проводили с такими же параметрами (mind 0,1, geno 0,05, maf 0,01). После этого исключали сцепленные полиморфизмы с коэффициентом  $r^2$  более 0,2. Отфильтрованный датасет анализировали с помощью программного обеспечения ADMIXTURE v1.3.0 [23] и рассчитывали кроссвалидацию для вероятностной оценки каждого  $k$ .

Метод  $f_3$ -статистик позволяет количественно оценить общий дрейф генов (т. е. степень общего происхождения) между двумя популяциями относительно дальней популяции (outgroup). Расчет  $f_3$ -статистик проведен с помощью программы qp3Pop из пакета AdmixTools [24]. В качестве дальней популяции (outgroup) применяли образцы популяции йоруба из данных проекта «1000 геномов» [25]. Вместе с образцами популяции йоруба использовали 668 образцов по 3 757 004 маркерам. Применяли следующие фильтры: mind 0,1, geno 0,05, maf 0,01, исключение сцепленных полиморфизмов проведено с  $r^2 > 0,5$ . После фильтрации остались 1 144 136 маркеров у 635 образцов.

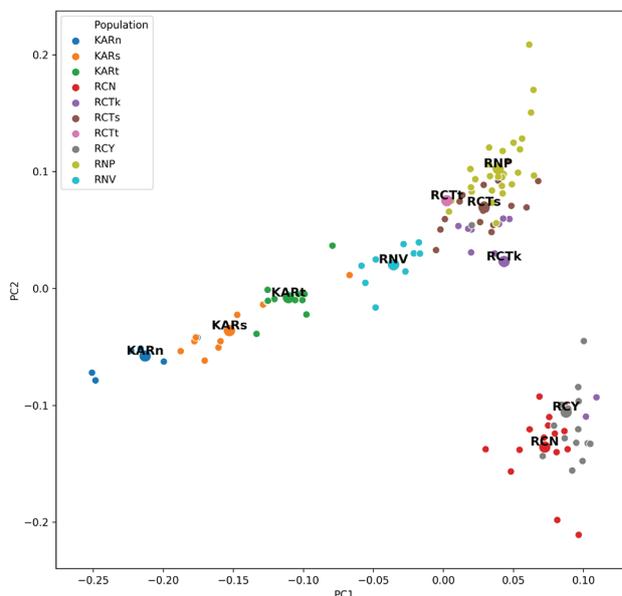
Метод  $d$ -статистик является тестом на генетическое смешение между четырьмя популяциями. Классически его применяют с одной дальней популяцией (африканской). В этом случае он позволяет понять направление потока генов между оставшимися тремя популяциями. Расчет  $d$ -статистик проводили с помощью программы qpDstat из того же пакета AdmixTools с использованием йоруба в качестве дальней популяции. Всего использовали 748 образцов по 3 757 004 маркерам. Параметры фильтрации: mind 0,05; geno 0,2; maf 0,01;  $r^2 > 0,6$ . После фильтрации остались 1 355 253 маркера у 633 образцов.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Положение тверских русских и тверских карел в генетическом пространстве главных компонент, рассчитанных по широкогеномной панели из 4,5 млн маркеров, представлено на рис. 2. Выборка тверских карел оказалась близка в карелам Карелии и отдалена от всех проанализированных русских популяций (тверских, новгородских, вологодских и ярославских). Лишь один образец тверских карел на диаграмме генетически приближается к русским Вологодской области. Все остальные образцы тверских карел расположены очень компактно, демонстрируя генетическое сходство. Такое расположение согласуется с результатами, полученными при анализе Y-хромосомных линий [5]: диаспора тверских карел сохранила свой предковый генофонд.

Но при этом аутосомные маркеры (широкогеномная панель) выявляют и частичное сближение тверских карел с русскими. На рис. 2 виден четкий клинальный градиент: северные карелы – южные карелы – тверские карелы – русские популяции. При этом к карелам наиболее генетически близки русские не Тверской, а Вологодской области, а тверские русские образуют единое генетическое облако вместе с псковскими и с популяциями центральной России. Генетические различия между западной и восточной группой тверских русских невелики, но заметны и согласуются с географией: западные тверские русские на диаграмме занимают общую зону с образцами из Псковской области, в то время как восточные (Кашинский район) находятся на их периферии. Причем два образца из восточной популяции перешли в новгородско-ярославский тандем, который вторая компонента отличает от остальных русских популяций (см. рис. 2).

В целом анализ главных компонент выявляет наибольшее сходство генофонда тверских карел с карелами Карелии, а не с русскими популяциями, но позволяет выдвинуть



**Рис. 2.** Диаграмма главных компонент изменчивости изученных популяций. KARn — карелы северные, KARs — карелы южные, KART — карелы тверские, RCN — русские новгородские, RCTk — русские тверские кашинские (восточные), RCTt — русские тверские торжокские (южные), RCTs — русские тверские селижаровские (западные), RCY — русские ярославские, RNP — русские псковские, RNV — русские вологодские. Индивидуальные образцы отмечены малыми кружками, центроиды (центр тяжести для образцов каждой популяции) отмечены увеличенными кружками того же цвета

гипотезу незначительного сближения генофондов русских и карел Тверской области. Поскольку из трех популяций карел в сторону русских популяций сдвинута только тверская популяция карел, а среди всех изученных русских популяций тверская не сдвинута в сторону карел, можно предполагать наиболее интенсивный поток генов от русских к карелам, а не в обратном направлении. Анализ  $f_3$ -статистики уточняет степень генетического сходства тверских карел с различными популяциями Восточной Европы: наиболее близки к ним генофонды народов Балтийского региона. По степени генетической близости к тверским карелам они расположились в таком порядке: ижора, водь, южные карелы, вепсы, литовцы, северные карелы. Сходство меньшее, чем для популяций балтийского региона, обнаруживается для русских популяций, среди которых тверских русских характеризует отнюдь не самое выраженное сходство с тверскими карелами (по убыванию сходства: псковские русские, новгородские, тверские западные, смоленские, курские, тверские восточные, ярославские, вологодские, воронежские, русские северо-востока Архангельской области).

Для уточненного описания генофонда мы применили метод ADMIXTURE, позволяющий дать качественную и количественную оценку вкладов различных «предковых популяций» в генофонды изучаемых популяций. Он позволяет варьировать число предковых популяций  $k$  для обнаружения общих предковых компонентов разной степени дробности.

При  $k = 5$  (рис. 3; таблица) мы видим у всех рассматриваемых популяций основной вклад лишь двух компонентов: обозначенный синим цветом (максимальные значения которого обнаруживаются у народов уральской языковой семьи) и обозначенный рыжим цветом (характерным для литовцев, украинцев, белорусов и большинства русских популяций). Синий компонент «А» преобладает у карел Карелии (85%; см. таблицу). И если у них компонент «Б» встречается лишь у единичных образцов, то у тверских карел он присутствует во всех образцах, составляя в среднем 41% их геномов (см. таблицу). У тверских русских мы видим более высокую частоту компонента «Б» — его в 2 раза больше (в среднем 80%), чем у тверских карел. Таким образом, результаты ADMIXTURE при  $k = 5$  не противоречат гипотезе частичного потока генов от русских к тверским карелам.

При  $k = 6$  (рис. 3) картина становится более точной благодаря тому, что карельские геномы диагностируются компонентом, окрашенным на графике в ярко-желтый цвет («В»; см. таблицу): у карел Карелии он достигает

практически 100%, у тверских карел его в два раза меньше (52%), а у тверских русских (8%), как и у псковских (4%), он крайне редок, указывая, что поток генов к ним от карел был либо незначительным, либо отсутствовал (в последнем случае наличие компонента «В» у русских объясняется более древним общим родством всех популяций Восточной Европы). Значительно больший вклад компонента «В» неожиданно обнаруживается у других соседей Тверской земли — у новгородских (39%), ярославских (30%) и вологодских (20%) русских.

Но при  $k = 8$  (см. рис. 3) эта картина меняется за счет дифференцировки компонента «В». Если у карел и вологодских русских ярко-желтый (условно «западно-финский») компонент сохраняет свои прежние позиции (на данном графике он обозначен как компонент «Е» и составляет 96% в Карелии, 53% — у тверских карел, 20% — у вологодских русских), то в остальных русских популяциях его доля минимальна. Наличие этого компонента у русских популяций может отражать не недавнее взаимодействие с карелами, а более древние события, например происхождение самих русских популяций в результате смешения славян с дославянским населением.

Таким образом, при  $k = 8$  из рассмотренных русских популяций только у вологодских можно проследить заметный (пятая часть генома) вклад условно «западно-финского» компонента «Е». Вместо него в остальных русских популяциях выделяется иной генетический пласт (обозначен светло-серым компонентом «И»). Он достигает предельного максимума у новгородских (91%) и ярославских (90%) русских, составляя почти весь их геном. Больше трети генома светло-серый компонент «И» занимает у тверских (39%), псковских (36%) и вологодских (34%) русских. Этот (условно «новгородский») компонент «И» обнаружен во всех остальных изученных нами популяциях центральной и южной России, составляя в среднем около трети (38%) их геномов.

По пропорциям условно «новгородского» компонента «И» и условно «южнорусского» компонента «К» различаются две группы тверских русских, причем не совсем в ладах с географией: если на западе Тверской области, граничащей с Новгородской областью, «южнорусский» компонент резко преобладает над «новгородским» («К»/«И» = 63/27), то у восточной группы тверских русских он равноправен с «новгородским» компонентом («К»/«И» = 42/42), а у представителей центральной тверской группы «новгородский» компонент «И» составляет весь генофонд (100%).

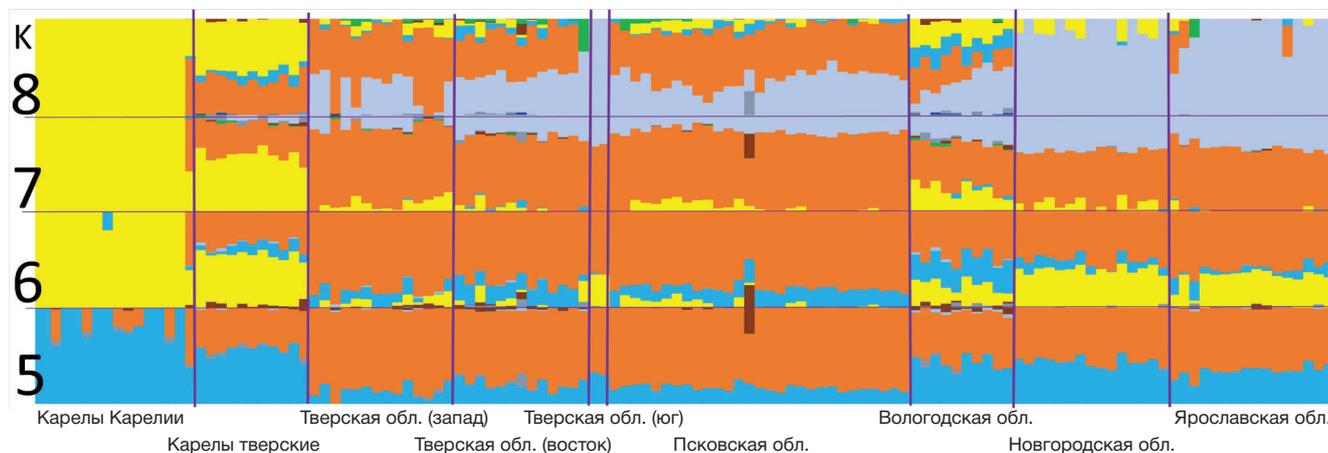


Рис. 3. Диаграмма результатов ADMIXTURE (вклад предковых компонентов в изучаемые популяции при разном числе компонентов  $k$ ). Профили индивидуальных образцов показаны в отдельных столбиках, индивиды разных популяций разграничены вертикальными линиями

Таблица. Величины вкладов предковых компонентов ADMIXTURE в изучаемые популяции (при разном числе компонентов  $k$ )

Уровень	Код компонента	ЭТНОС	КАРЕЛЫ									РУССКИЕ																	
			Популяции			Тверские			Тверские (запад)			Тверские (восток)			Тверские (юг)			Псковские			Вологодские			Новгородские			Ярославские		
			Объем выборки			$n = 11$			$n = 14$			$n = 13$			$n = 2$			$n = 29$			$n = 10$			$n = 15$			$n = 16$		
			Цвет компонента на рис. 3	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN
%	%	%		%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%		
К																													
5	A	Синий	85	37	100	55	41	63	13	0	24	20	10	29	31	30	33	17	11	25	42	30	52	43	34	53	34	18	45
	Б	Рыжий	13	0	60	41	35	50	85	71	96	75	65	90	69	67	70	83	52	89	51	42	61	56	47	65	65	55	77
6	В	Ярко-желтый	95	36	100	52	38	60	6	0	17	5	0	22	33	32	33	4	0	15	20	8	31	39	29	46	30	3	37
	Г	Синий	2	0	20	9	3	14	10	0	18	18	6	27	1	1	2	16	8	24	28	20	35	9	0	14	8	2	20
	Д	Рыжий	3	0	56	35	27	45	81	68	97	73	62	84	65	64	66	79	50	86	47	37	60	51	41	62	60	49	72
8	Е	Ярко-желтый	96	38	100	53	40	59	7	0	19	3	0	16	0	0	0	5	0	13	20	10	30	9	0	23	1	0	13
	Ж	Синий	0	0	3	8	2	13	2	0	7	6	0	15	0	0	0	3	0	11	18	12	22	0	0	4	1	0	7
	И	Светло-серый	0	0	0	0	0	0	27	0	47	42	31	66	100	100	100	36	13	57	34	8	55	91	72	100	90	43	100
	К	Рыжий	3	0	58	37	29	47	63	36	98	42	0	54	0	0	0	54	36	70	24	0	38	0	0	0	6	0	41

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Итак, анализ широкогеномной панели маркеров (аутосомных) подтвердил выводы, сделанные ранее по Y-хромосоме [5], что тверские карелы генетически более сходны с карелами Карелии, чем с окружающими русскими популяциями Тверской области. Важно, что этот вывод не только следует из описательных методов (PCA, ADMIXTURE), но подтверждается и значениями  $d$ -статистики. Метод  $d$ -статистик ( $f_4$ -статистик) классически применяют с одной дальней популяцией (африканской). Он позволяет понять направление потока генов между оставшимися тремя популяциями, причем поток генов считают достоверным при параметре  $Z$ , по модулю превышающим 3. Метод  $d$ -статистики в форме  $d$  (Yoruba, TverKarelians; SouthKarelians, TverRussians) для восточной популяции тверских русских характеризуется величиной  $Z = -6,9$ , для западной  $Z = -5,0$ . Это доказывает, что генофонд тверских карел ближе к генофонду карел Карелии, чем к генофонду тверских русских. Но при этом у тверских карел генетическое сходство с русскими популяциями все же более выражено, чем у южных (и тем более северных) карел Карелии. Соответствующие  $d$ -статистики тверских карел принимают достоверные значения при включении в анализ целого ряда более южных, чем Тверь, русских популяций: например, для русских юга Смоленской области  $d$ -статистика в

форме  $d$  (Yoruba, RussiansSmolensk; TverKarelians, Karelians) характеризуется значением  $Z = -3,4$ . Это показывает, что генофонд тверских карел при общем сходстве с карелами Карелии по сравнению с ними достоверно приближен к генофонду смоленских и других русских популяций.

Таким образом, если предполагать, что изначально предки тверских карел и карел Карелии представляли собой единую популяцию [1, 2, 4], то  $d$ -статистики доказывают, что затем предки тверских карел приняли генетический вклад популяций, населяющих более южные территории Восточно-Европейской равнины. Для населения Восточной Европы характерна столь сложная сеть предшествующих и последующих перемещений населения, что генетические данные не позволяют однозначно идентифицировать источник этой южной примеси у тверских карел, но исторически на эту роль лучше всего подходят русские популяции Тверской области и соседних областей.

## ВЫВОДЫ

Мы изучили генофонды популяций карел и русских Тверской области по панели 4,5 млн аутосомных SNP-маркеров, проанализировав их на широком фоне карел Карелии и русских областей, граничащих с Тверской (Псковской, Новгородской, Вологодской,

Ярославской). Применение различных методов анализа (главных компонент, ADMIXTURE, d- и f3-статистик) дало согласованные результаты.

Генофонд тверских карел сохраняет наибольшую близость к генофонду карел Карелии, несмотря на их проживание среди многочисленного русского населения в течение 3–5 веков и 20-кратное сокращение численности в течение последнего столетия. При этом генофонд тверских карел все же более сходен с русским генофондом, чем генофонд других обследованных карельских популяций. Сопоставление данных, полученных по аутосомным маркерам (частичное смещение в сторону русского генофонда) с опубликованными данными по Y-хромосоме (отсутствие детектированного смешения с русскими),

позволяет предположить, что основной формой потока генов от русских к тверским карелам были браки карелов с русскими женщинами.

Демографические данные (резкое сокращение численности тверских карел) и исторические события позволяли предполагать изменение самоидентификации тверских карел и их ассимиляции русским населением. В этом случае в геномах тех тверских русских, которые были потомками карел, сменивших этническую самоидентификацию, должна была быть велика доля генетического компонента, характерного для карел. Однако это явление не обнаружено: тверские русские оказались столь же генетически отличны от карел, как, например, псковские.

## Литература

1. Савинова А. И., Степанова Ю. В. Карельская диаспора южных районов Тверского Поволжья: история формирования и историческая судьба. CARELiCA. 2018; 1 (19): 26–37.
2. Степанова Ю. В., Савинова А. И. Расселение карел в Верхневолжье в середине — второй половине XVII в.: опыт изучения с применением гис-технологий. Историческая информатика. 2018; 4: 57–72.
3. Вишневецкий А. Г., редактор. Переписи населения Российской Империи, СССР, 15 новых независимых государств. Демоскоп Weekly [Internet]. [cited 2020 Oct 9]. Available from: [http://www.demoscope.ru/weekly/ssp/census\\_types.php?ct=6](http://www.demoscope.ru/weekly/ssp/census_types.php?ct=6).
4. Головкин А. Н. История Тверской Карелии. Тверь: Студия-С, 2008; 432 с.
5. Агджоян А. Т., Дараган Д. М., Схаляхо Р. А., Реутов П. П., Балановский О. П., Балановская Е. В. и др. Возможность сохранения генофонда в диаспоре на примере тверских карел. Генетика. 2018; 54: 91–94.
6. Rajman I, Knapp L, Morgan T, Masimirembwa C. African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development. EBioMedicine. 2017; 17: 67–74.
7. Jing L, Haiyi L, Xiong Y, Dongsheng L, Shilin L, Jin L, et al. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. Journal of Medical Genetics. 2014; 51 (9): 614–22.
8. Mirzaev KB, Fedorinov DS, Ivashchenko1 DV, Sychev DA. ADME pharmacogenetics: future outlook for Russia. Pharmacogenomics. 2019; 20 (11): 847–65.
9. Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, et al. Progress toward an efficient panel of SNPs for ancestry inference. Forensic Science International: Genetics. 2014; 10: 23–32.
10. Nassir R, Kosoy R, Tian C, White PA, Butler LM, Silva G, et al. An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. BMC Genetics. 2009; 10: 39.
11. Балановская Е. В., Агджоян А. Т., Чухряева М. И., Маркина Н. В., Балаганская О. А. Балановский О. П. и др. Популяционные биобанки: принципы организации и перспективы применения в геногеографии и персонализированной медицине. Генетика. 2016; (12): 1371–87.
12. Агджоян А. Т., Схаляхо Р. А., Балаганская О. А., Козлов С. А., Палипана С. Д., Балановский О. П. и др. Генофонд новгородцев: между севером и югом. Генетика. 2017; 53 (11): 1338–48.
13. Чухряева М. И., Павлова Е. С., Напольских В. В., Гарин Э. В., Балановский О. П., Балановская Е. В. и др. Сохранились ли следы финно-угорского влияния в генофонде русского населения Ярославской области? Свидетельства Y-хромосомы. Генетика. 2017; 53 (3): 378–89.
14. GG-base [Internet]. [cited 2020 Oct 9]. Available from: <https://www.gg-base.org/>.
15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007; 559–75.
16. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 2009; (19): 1655–64.
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015 Feb 25; 4 (7). Available from: DOI: 10.1186/s13742-015-0047-8.
18. Price A, Patterson N, Plenge R. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006; 38: 904–9.
19. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLOS Genetics [Internet]. 2006; 2 (12): e190 [cited 2020 Oct 9]. Available from: <https://doi.org/10.1371/journal.pgen.0020190>.
20. McKinney W. Data structures for statistical computing in Python. SciPy 2010: Proceedings of the 9th Python in Science Conference; 2010 Jun 28 – Jul 3. Austin, Texas. Available from: <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>.
21. Reback J, McKinney W, jbrockmendel, Augspurger T, Cloud F, Mehyar M, et al. pandas-dev/pandas: Pandas 1.0.3. Version 1.0.3 [software]. Zenodo. 2020 Mar 18 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.3715232>.
22. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. IEEE Xplore. 2007; 9 (3): 90–95.
23. Waskom M, Botvinnik O, O’Kane D, Hobson P, Lukauskas S, Qalieh A, et al. mwaskom/seaborn: v0.8.1 (September 2017). Version 0.8.1 [software]. Zenodo. 2017 Sep 3 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.883859>.
24. Alexander DH, Novembre J, Lange K. ADMIXTURE Software. Version 1.3.0 [software]. 2020 May 3 [cited 2020 Oct 9]. Available from: <http://dalexander.github.io/admixture/index.html>.
25. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. GENETICS. 2012; 192 (3): 1065–93.
26. Auton A, Abecasis G, Altshuler D, et al. A global reference for human genetic variation. Nature. 2015; (526): 68–74.

## References

- Savinova AI, Stepanova UV. Kareĭskaya diaspora uzhnih rayonov Tverskogo Povolgya: istoria formirovaniya i istoricheskaya sud'ba. CARELiCA. 2018; 1 (19): 26–37. Russian.
- Savinova AI, Stepanova UV. Rasselenie karel v Verhnevolzhie v seredine — vtoroy polovine XVII v.: opit izucheniya s primeneniem gis-technologiy. Istoricheskaya informatika. 2018; 4: 57–72. Russian.
- Vishnevskiy AG, editor. Perepisi naseleniya Rossiyskoy Imperii, SSSR, 15 novykh nezavisimyykh gosudarstv. Demoskop Weekly. [Internet]. [cited 2020 Oct 9]. Available from: [http://www.demoscope.ru/weekly/ssp/census\\_types.php?ct=6](http://www.demoscope.ru/weekly/ssp/census_types.php?ct=6).
- Golovkin AN. Istoriya Tverskoy Karelii. Tver': Studiya-S, 2008; 432 s. Russian.
- Agdzhoyan AT, Daragan DM, Skhalyakho RA, Reutov PP, Balanovskiy OP, Balanovskaya EV, et al. Vozmozhnost' sokhraneniya genofonda v diaspore na primere tverskikh karel. Genetika. 2018; 54 (Application S): 91–94. DOI: 10.1134/S0016675818130027. Russian.
- Rajman I, Knapp L, Morgan T, Masimirembwa C. African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development. EBioMedicine. 2017; 17: 67–74.
- Jing L, Haiyi L, Xiong Y, Dongsheng L, Shilin L, Jin L, et al. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. Journal of Medical Genetics. 2014; 51 (9): 614–22.
- Mirzaev KB, Fedorin DS, Ivashchenko DV, Sychev DA. ADME pharmacogenetics: future outlook for Russia. Pharmacogenomics. 2019; 20 (11): 847–65.
- Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, et al. Progress toward an efficient panel of SNPs for ancestry inference. Forensic Science International: Genetics. 2014; 10: 23–32.
- Nassir R, Kosoy R, Tian C, White PA, Butler LM, Silva G, et al. An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. BMC Genetics. 2009; 10: 39.
- Balanovskaya EV, Agdzhoyan AT, Chukhryaeva MI, Markina NV, Balaganskaya OA, Balanovskiy OP, et al. Populyatsionnye biobanki: printsipy organizatsii i perspektivy primeneniya v genogeografii i personalizirovannoy meditsine. Genetika. 2016; (12): 1371–87. Russian.
- Agdzhoyan AT, Skhalyakho RA, Balaganskaya OA, Kozlov SA, Palipana SD, Balanovskiy OP, et al. Genofond novgorodtsev: mezhdru severom i yugom. Genetika. 2017; 53 (11): 1338–1348. Russian.
- Chuhriaev MI, Pavlova ES, Napolskih VV, Garin EV, Balanovskiy OP, Balanovska EV, et al. Sohranilis' li sledy finno-ugorskogo vliyaniya v genofonde russkogo naseleniya Yaroslavskoy oblasti? Svidetel'stva Y-chromosomi. Genetica. 2017; 53 (3): 378–89. GG-base [Internet]. [cited 2020 Oct 9]. Available from: <https://www.gg-base.org/>.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007; 559–75.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 2009; (19): 1655–64.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015 Feb 25; 4 (7). Available from: DOI: 10.1186/s13742-015-0047-8.
- Price A, Patterson N, Plenge R, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006; 38: 904–9.
- Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLOS Genetics [Internet]. 2006; 2 (12): e190 [cited 2020 Oct 9]. Available from: <https://doi.org/10.1371/journal.pgen.0020190>.
- McKinney W. Data structures for statistical computing in Python. SciPy 2010: Proceedings of the 9th Python in Science Conference; 2010 Jun 28 – Jul 3. Austin, Texas. Available from: <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>.
- Reback J, McKinney W, jbrockmendel, Augspurger T, Cloud F, Mehyar M, et al. pandas-dev/pandas: Pandas 1.0.3. Version 1.0.3 [software]. Zenodo. 2020 Mar 18 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.3715232>.
- Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. IEEE Xplore. 2007; 9 (3): 90–95.
- Waskom M, Botvinnik O, O'Kane D, Hobson P, Lukauskas S, Qaliev A, et al. mwaskom/seaborn: v0.8.1 (September 2017). Version 0.8.1 [software]. Zenodo. 2017 Sep 3 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.883859>.
- Alexander DH, Novembre J, Lange K. ADMIXTURE Software. Version 1.3.0 [software]. 2020 May 3 [cited 2020 Oct 9]. Available from: <http://dalexander.github.io/admixture/index.html>.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. GENETICS. 2012; 192 (3): 1065–93.
- Auton A, Abecasis G, Altshuler D, et al. A global reference for human genetic variation. Nature. 2015; (526): 68–74.