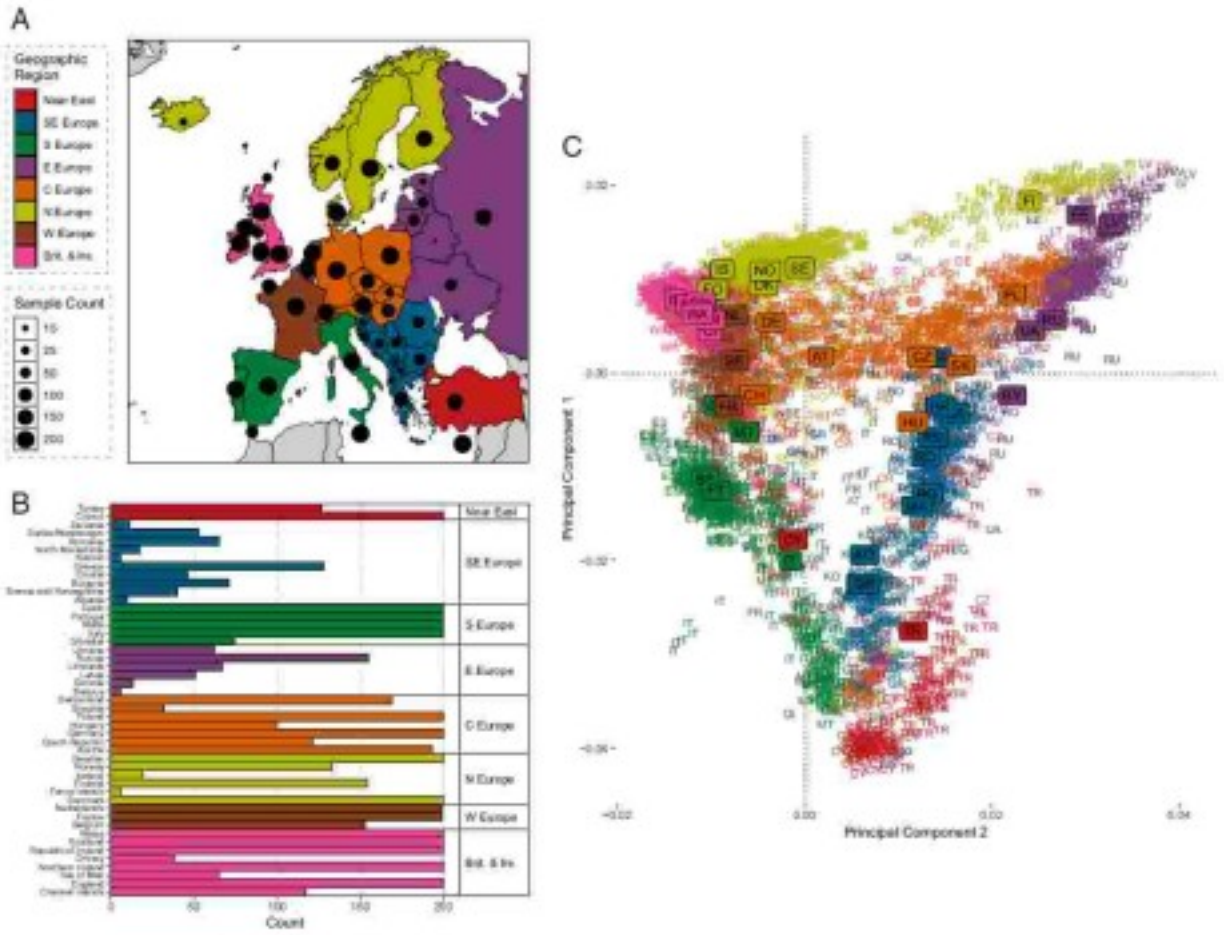
Анализ образцов ДНК из Биобанка Великобритании позволил уточнить геномный ландшафт Европы

Анализ гаплотипов образцов из Биобанка Великобритании позволил получить информацию о различиях в демографической истории популяций разных европейских регионов и уточнить геномный ландшафт в масштабах всей Европы. Статья ирландских специалистов опубликована в журнале PNAS**.**

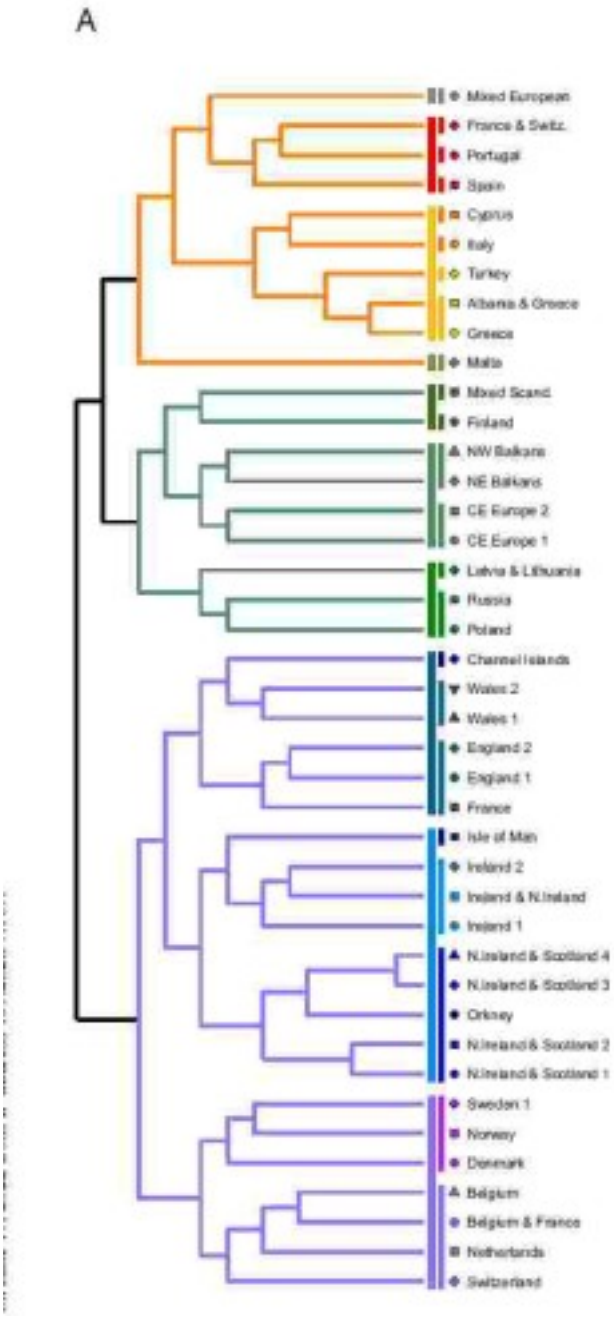
Авторы изучили геномы 5500 человек, образцы ДНК которых хранились в Биобанке Великобритании. Для анализа использовали метод IBD (identical-by-descent), который выявляет в геномах индивидов из разных популяций сегменты общего происхождения, что говорит об общих предках в недавнем прошлом. Исследователи отобрали 5500 индивидов европейского происхождения, которые родились не только в Великобритании и Ирландии, но и в других регионах: Северная Европа, Западная Европа, Центральная Европа, Восточная Европа, Южная Европа, Юго-Восточная Европа, Ближний Восток, всего 47 стран, из каждой было представлено от 10 до 200 образцов. В каждом геноме анализировали около 500 тысяч SNP.

Анализ главных компонент по 5500 образцам провели с применением алгоритма PLINK, основанного на частотах аллелей. График PCA показал, что генетический ландшафт Европы отражает ее географию.



A. Регионы Европы, из которых происходили индивиды в изученной выборке Биобанка Великобритании; размер кружка соответствует числу индивидов того или иного региона. B. Число индивидов с выборке, происходящих из разных стран; цвет соответствует региону Европы. C. Анализ главных колмпонент по геномным данных 550 индивидов, происходящих из разных регионов Европы (разные цвета); страны обозначены в квадратиках.

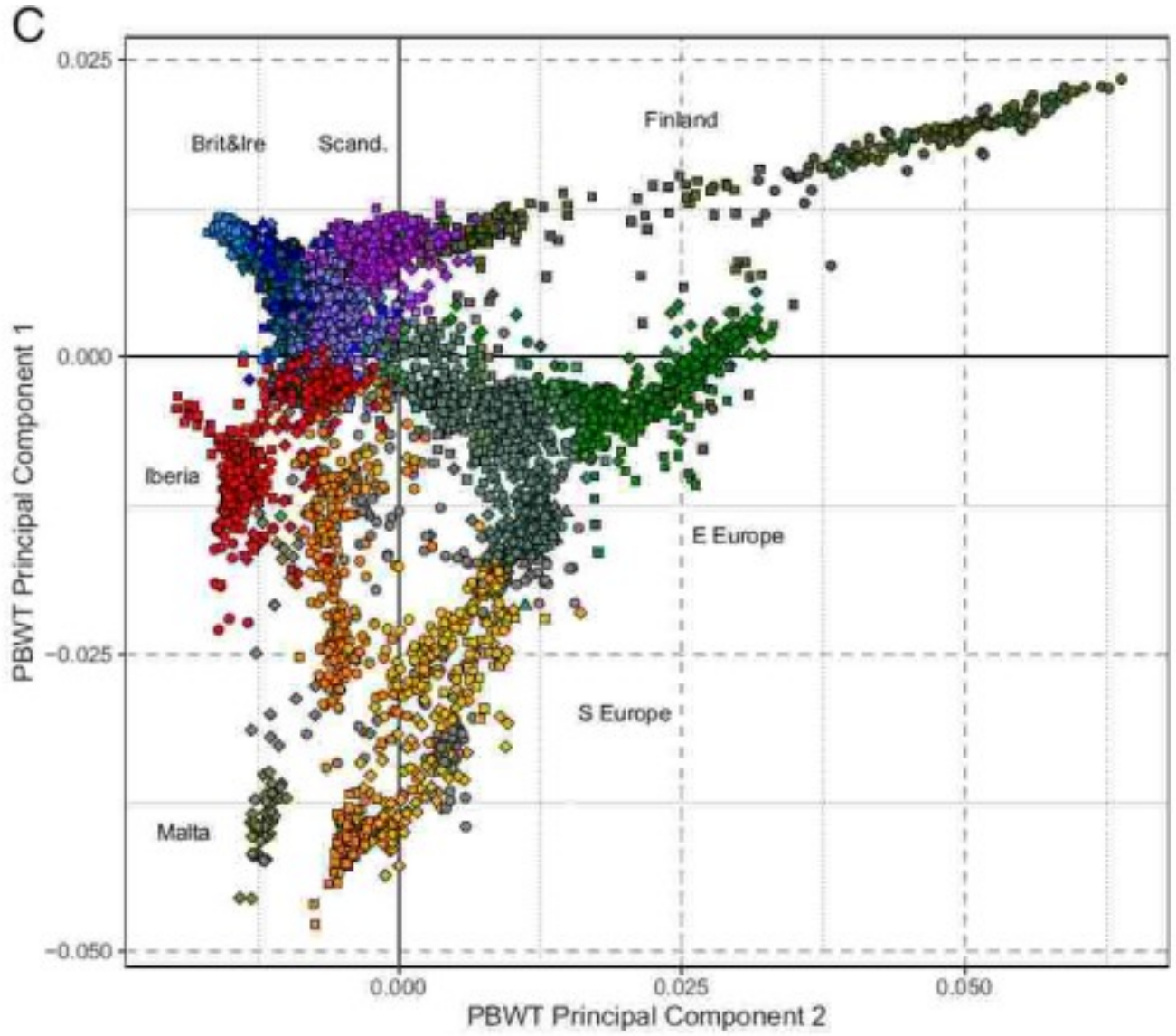
Используя алгоритм Leiden, авторы выделили 41 кластер, объединяющий индивидов по степени родства. На основе этих кластеров они построили филогенетическое дерево.



Дендрограмма, на которой показаны генетические кластеры, выявленные по алгоритму Leiden. Три основных ветви обозначены разными цветами, дополнительные субветви – вертикальными линиями.

Затем исследователи оценили долю общих гаплотипов между индивидами 41 кластера для обнаружения недавних генетических потоков и оценки родства между кластерами. Наконец, расстояние между кластерами оценили, используя программу ADMIXTOOLS2.

Анализ главных компонент (PCA) на основе анализа гаплотипов (pbwt paint coancestry matrix) более четко выделил кластеры, чем тот же PCA на основе частот аллелей. Образовались три группы кластеров, объединяющих индивидов, родившихся в Северо-Западной Европе, Центральной и Восточной Европе и Южной Европе.



Анализ главных компонент на основе pbwt paint coancestry matrix. Цвета соответствуют цвету кластеров по Leiden. Указаны регионы происхождения индивидов.

Северо-Западная Европа

Ветвь Северо-Западной Европы включает индивидов происходящих из Скандинавии, Нидерландов, Бельгии, Франции, Швейцарии, Великобритании и Ирландии. Она отличается низким уровнем генетической дифференциации, как следует из величины генетических расстояний (Fst) внутри нее (0,0006). Северо-западная ветвь делится на две субветви, отделяющие британских и ирландских индивидов от континентальной Европы. Британско-ирландская ветвь разделяется на восточную (Англия и Уэльс) и северо-западную (Ирландия и Шотландия) субветви. Французский кластер группируется с британским, что отражает генетический поток вдоль Ла-Манша и согласуется с генетической близостью между северо-западной Францией и соседней Британией. Среди прочих индивидов Скандинавия (включая Данию) образует свою собственную ветвь, а норвежская подветвь внутри нее включает все образцы исландцев. Кластеры Нидерландов и Дании близки по общим гаплотипам, что согласуется с их географической близостью.

Центральная и Восточная Европа

Ветвь Центральной и Восточной Европы включает три субветви: Северо-Восточная Европа (страны Балтики, Польша и Россия), Центрально-Восточная Европа (север Балкан, центр и восток Европы) и отдельно Финляндия. В соответствии с предыдущими исследованиями, Финляндия демонстрирует признаки генетической изоляции от остальных европейских регионов, что подтверждается по PCA и по генетическим расстояниям. Большинство индивидов из Эстонии вошли в смешанный скандинавский кластер, который также объединил шведов, норвежцев и часть финнов.

Индивиды из балтийских стран, включая Латвию и Литву, группируются вместе. Эстонцы расположились в пространстве главных компонент между финнами и балто-русским кластером, что предполагает генетический поток от Финляндии к

Балтике. Близки к латвийско-литовскому два кластера, преимущественно поляков и русских, образующие западно-восточную клину.

Ветвь Центрально-Восточной Европы в пространстве главных компонент соединяет Северо-Восточную и Западную Европу. Она состоит из двух кластеров — Центрально-Восточная Европа 1 и 2. Последняя группирует индивидов из более западных стран, а первая – из более восточных, что отражает клину, тянущуюся с востока на запад.

Два кластера на севере Балкан – северо-запад и северо-восток Балкан – демонстрируют географическую клину с востока на запад; первый объединяет индивидов из Хорватии, Боснии и Герцеговины и Сербии, второй – из Румынии и Болгарии. В северо-восточном кластере больше прослеживается южный генетический компонент, чем в северо-западном.

Южная Европа

В Южной Европе анализ с использованием алгоритма Лейдена выделяет три группы кластеров. Одна группа объединяет индивидов, родившихся в Восточном Средиземноморье (Италия, Греция, Турция и Кипр) и на Иберийском полуострове, наконец, два кластера считаются аутгруппами – «смешанная Европа» и Мальта. Большинство индивидов из Греции образует единый кластер, который моделируется как смешение гаплотипов из соседних южных кластеров, таких как Италия и Турция, а также гаплотипов с северо-востока Балкан. В Италии кластеризация не разделяет индивидов с севера и юга, как отмечалось ранее. В иберийской субветви индивиды из Испании и Португалии группируются в свои собственные кластеры, в соответствии с прежними наблюдениями, хотя португальский кластер включает четверть индивидов из Испании. В исследовании представлено наибольшее на сегодняшний день количество генотипов с Мальты (200 индивидов). Они образуют три кластера, что предполагает наличие генетического потока на Мальту с запада.

Чтобы реконструировать демографическую историю Европы, авторы применили анализ в геномах сегментов общего происхождения (IBD сегментов). Они поставили задачу вычислить возраст возникновения и размер популяций в каждом кластере. По анализу IBD сегментов оценили коэффициент инбридинга; длинные IBD сегменты в популяциях говорят о недавней изоляции или о практике эндогамии.

Данные анализа подтвердили градиент с юга на север, по которому увеличивается доля общих гаплотипов в европейских геномах. Индивиды, родившиеся в Финляндии, демонстрировали самый высокий уровень внутрипопуляционных общих IBD сегментов и высокий показатель гомозиготности (ROH), из чего следует высокий коэффициент инбридинга. Эти результаты согласуются с ранее проведенными исследованиями и объясняются исторически зафиксированным бутылочным горлышком. Результаты подтвердили предыдущие наблюдения о повышенной доле общих гаплотипов в островных популяциях (на примере Оркнейских и некоторых других островов Британского архипелага). Для Южной Европы характерен низкий уровень общих гаплотипов, хотя есть исключения, например, Мальта, где он почти так же высок, как в Финляндии. В Юго-Восточной Европе повышенный уровень общих гаплотипов демонстрируют Турция, Греция, Кипр и Албания.

Итак, исследователи впервые продемонстрировали возможности применения методов анализа гаплотипов в масштабах всей Европы для выявления различий в демографической истории разных ее регионов. Так, IBD анализ позволил установить степень популяционной изоляции и эффективный размер популяций в разных частях Европы. Биобанк Великобритании оказался ценным ресурсом для исследования геномного ландшафта Европы вне Великобритании.

текст Надежды Маркиной

Источник:

Edmund Gilbert , Ashwini Shanmugam , and Gianpiero L. Cavalleria. Revealing the recent demographic history of Europe via haplotype sharing in the UK Biobank // PNAS, 2022, Vol. 119, No. 25 e2119281119 https://doi.org/10.1073/pnas.2119281119

https://www.pnas.org/doi/10.1073/pnas.2119281119