

## Исследованы структурные вариации генома в 54 популяциях

Структурные вариации генома представляют важную часть глобального генетического разнообразия. Исследователи составили каталог структурных вариаций по геномам 911 человек 54 популяций, описали 126 тысяч вариаций 78% которых не были известны ранее. Найдены вариации, специфичные для отдельных популяций, которые находятся под действием локального отбора. Среди структурных вариаций обнаружены новые последовательности, не содержащиеся в референсном геноме. Некоторые структурные вариации, имеющие функциональное значение, в том числе влияющие на здоровье, были заимствованы современным человеком от неандертальцев и денисовцев.

Специалисты из Института Сенгера (Wellcome Sanger Institute) в Великобритании опубликовали каталог структурных вариаций генома по данным 911 человек из 54 популяций мира, статья с результатами этой работы [опубликована в журнале Cell](#), ведущие авторы статьи – Крис Тайлер-Смит и Яли Сю. Для этой работы секвенировали 911 геномов из коллекции Human Genome Diversity Project (HGDP), в которой хранятся клеточные линии от 1000 с лишним человек из более чем 50 популяций мира. Геномы секвенировали со средним покрытием 36x и картировали на референсный геном GRCh38.

В ходе нескольких глобальных проектов по секвенированию геномов специалисты в большей степени концентрировались на однонуклеотидном полиморфизме (SNP). Авторы данного исследования сосредоточились на структурных вариациях, которые затрагивают 50 и более пар оснований. К ним относятся делеции (выпадения участка), дубликации (удвоения участка), вариации число копий, инсерции (вставка участка), инверсии (поворот участка на 180 градусов). Наряду с однонуклеотидным полиморфизмом структурные вариации вносят немалый вклад в геномное разнообразие, а многие из них ассоциированы с теми или иными заболеваниями.



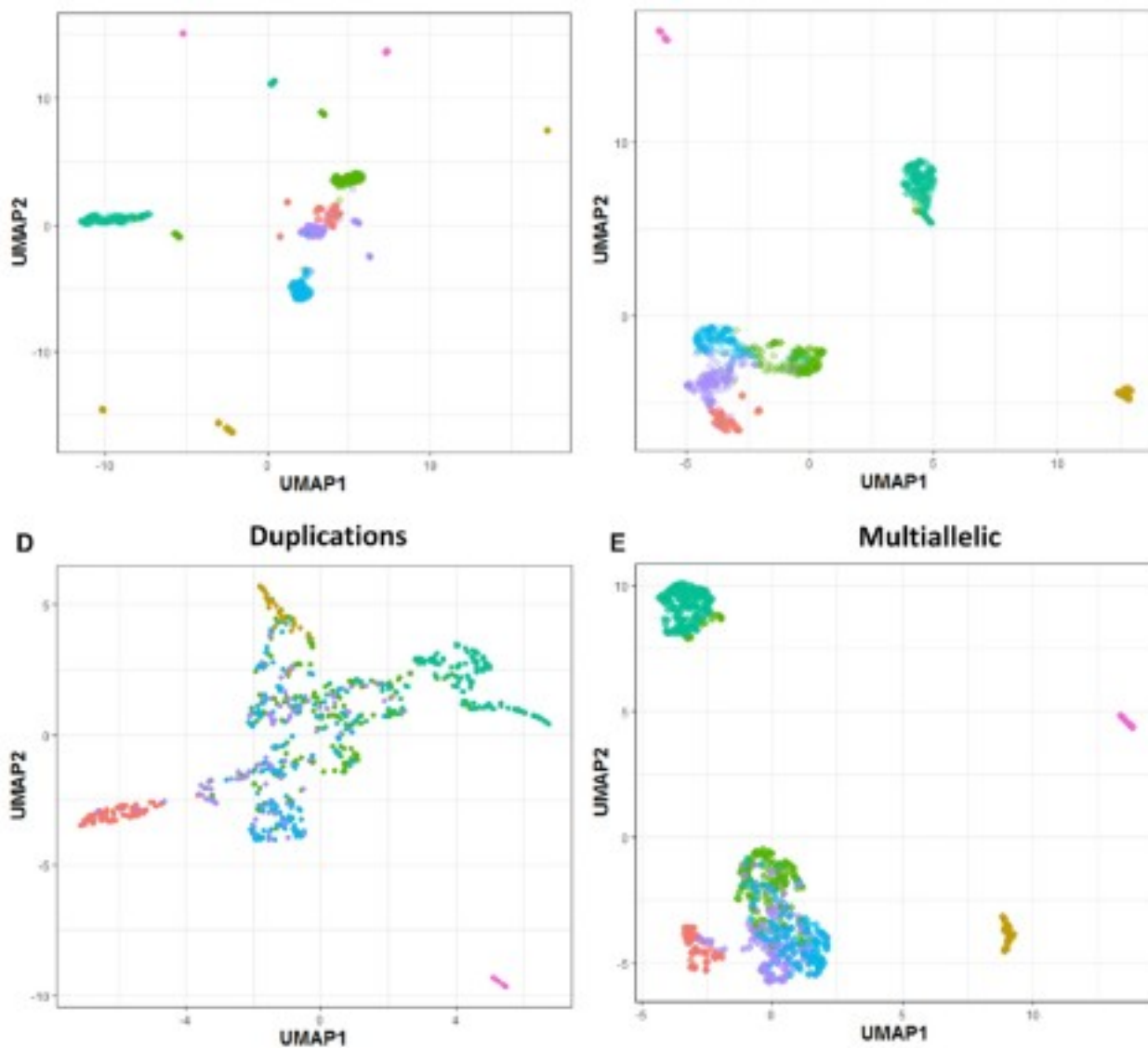
На карте показаны популяции, образцы которых из коллекции HGDP вошли в работу. Разными цветами обозначены региональные группы.

В общей сложности в работе было описано 126 018 структурных вариаций генома, 78% которых не были известны ранее. Среди них обнаружили как редкие варианты, так и достигающие высокой частоты в определенных популяциях.

В распределении структурных вариаций прослеживается популяционная структура: выделяются континентальные кластеры вариаций, а также группы внутри континентов. В наибольшей степени популяционная структура проявилась в делециях. Так, по делециям произошло глубокое разделение в пределах Африки: популяции мбути, бушмены и биака отделились от остальных народов. Смешанные группы, такие как хазарейцы и уйгуры образовали кластер, отдельный от прочих популяций

Центральной, Южной и Восточной Азии. Другие классы геномных вариаций — инсерции, дубликации, вариации числа копий и инверсии, также продемонстрировали популяционную структуру, хотя и не в такой степени, как делеции.

На рисунке представлены результаты анализа UMAP, это алгоритм машинного обучения, который был применен к базе структурных вариаций разных континентальных групп (они обозначены теми же цветами, что на предыдущем рисунке). Представлены делеции, дубликации, инсерции и мультиаллельные вариации.



Результаты анализа UMAP по делециям, дубликациям, инсерциям и мультиаллельным вариациям. Региональные группы показаны разными цветами, как на предыдущем рисунке.

Сравнивая частоту вариаций в популяциях, ученые определяли, какие из них находятся под отбором. Один из наиболее ярких примеров связан с двумя популяциями папуасов, живущих в горах и на равнине. Делеция в гене *HBA2* достигает частоты 86% у папуасов, живущих на равнине, но полностью отсутствует у высокогорных папуасов. Эта делеция связана с устойчивостью к малярии, которая не встречается в горах, но распространена в низинах. С другой стороны, у папуасов высокогорья с частотой 93% встречается небольшая инсерция близ экзона гена *VGLL4*, а у равнинных папуасов она гораздо менее часта.

Некоторые популяционно специфические структурные вариации достигают высокой частоты не из-за положительного отбора, а из-за других факторов. Например, в популяции южноамериканских индейцев каритиана с частотой 40% встречается делеция размером 14 kb. Эта делеция инактивирует ген, кодирующий фермент, участвующий в расщеплении крахмала. Известно, что данный ген был мишенью положительного отбора в древних южноамериканских популяциях при их переходе к земледелию. Так что отбор действовал против данной мутации. Но в последнее время популяция индейцев каритиана сокращается в численности, и эта неблагоприятная мутация повышает свою частоту под действием дрейфа генов. В африканской популяции мбути с частотой 54% встречается специфическая делеция, которая приводит к гиперреактивности иммунной системы и к

аутоиммунным заболеваниям. Как ни странно, эта делеция по всем признакам находится под положительным отбором, и причина этого неясна.

Выявленные структурные вариации исследователи сравнили с геномами неандертальца и денисовца, чтобы определить, есть ли среди них те, которые наши предки заимствовали от древних видов людей. Они нашли сотни вариаций, имеющих в древних геномах и специфичных для африканских популяций. Вероятно, эти вариации присутствовали у наших общих предков, но были потеряны, когда современные люди вышли из Африки и прошли через бутылочное горлышко. Наряду с ними найдены вариации, которые имеются в древних геномах, но отсутствуют у африканцев. Их можно объяснить метисацией современного человека с древними видами людей. Исследование показало, что неандертальцы и денисовцы внесли некоторые функционально значимые структурные вариации в геном современного человека. Поскольку многие из таких вариаций вовлечены в работу иммунной системы, вероятно, они были связаны с адаптацией к патогенам, с которыми человек столкнулся после выхода из Африки.

Больше всего структурных вариаций, заимствованных от денисовцев, оказалось в популяциях Океании. К ним относится дупликация на хромосоме 16p12.2. Очевидно, она была получена при метисации предков с денисовцами и затем оказалась под положительным отбором. Авторы детально исследовали эту мутацию и выяснили, что включенные в нее гены ассоциированы с 1% случаев аутизма. Почему эта мутация оказалась под отбором, непонятно. В популяциях Океании найдены и другие часто встречающиеся варианты, общие с денисовцами, некоторые из них связаны с системой адаптивного иммунитета. По-видимому, предки населения Океании претерпели отдельную метисацию с денисовцами, после чего популяция оказалась в длительной изоляции. Делеция, заимствованная от неандертальцев, обнаружена в популяциях индейцев суруи и пима. Она подавляет ген *MS4A1*, который играет важную роль в реакции на некоторые моноклональные антитела. Как пишут авторы, этот пример показывает, что терапия, разработанная для одних этнических групп, может быть неэффективной для других.

В следующем классе структурных вариаций — вариациях числа копий (copy number variants, CNV), авторы выделяют «gynaway» дупликации («сбежавшие из-под контроля», бесконтрольно размножившиеся). Это гены, которые в каких-то популяциях вдруг размножаются и приобретают много копий. Скорее всего, это происходит под действием локального отбора. Больше всего вариаций числа копий найдено в популяциях Африки. К ним относится ген *HPR*, который в популяциях Центральной и Западной Африки достигает 9 копий. Этот ген имеет ярко выраженное адаптивное значение, он кодирует белок, ассоциированный с защитой от возбудителя сонной болезни, трипаносомы, передающейся мухой це-це. В популяциях Восточной Африки участок гена олфакторного рецептора *OR7D2* может достигать 19 копий; множество копий этого же участка найдено у китайцев хань. Увеличение копий гена *HCAR2* обнаружено у азиатской популяции калашей, этот ген кодирует рецептор, который экспрессируется в иммунных клетках и является терапевтической мишенью, так как опосредует противовоспалительный эффект лекарства. В популяциях Океании большого числа копий достигает ген *SULT1A1*, который вовлечен в метаболизм лекарств и гормонов.

Исследователи отдельно секвенировали 25 геномов из разных популяций с покрытием 50x. В них они нашли совершенно новые последовательности – инсерции, которые отсутствуют в референсном геноме. Всего было найдено 1 643 таких новых инсерций, в общей сложности они составили 1.9 Мб. Чемпионами по числу новых вставок в геном оказались бушмены. Гены, в состав которых вошли эти инсерции, вовлечены в разнообразные процессы, включая иммунитет, метаболизм глюкозы и супрессию опухолей.

Полученные в работе результаты подтверждают концепцию, что адаптация человека к различным условиям среды включала потерю или приобретение генов, полностью или частично. Созданный каталог структурных вариаций генома может служить базой для будущих работ. Такой каталог важен в том числе для медико-генетических исследований, так как многие из описанных структурных вариаций оказывают влияние на здоровье и чувствительность к лекарствам. Вместе с тем авторы подчеркивают ограниченность использования одного референсного генома для полного исследования генетического разнообразия. Они считают, что эталонные геномы, секвенированные с высоким качеством, должны быть в каждой популяции.

*текст Надежды Маркиной*

#### **Источник:**

Mohamed A. Almarri et al. Population Structure, Stratification, and Introgression of Human Structural Variation // Cell. 2020, 182, 1–11, July 9 <https://doi.org/10.1016/j.cell.2020.05.024>

[Статья в открытом доступе](#)

