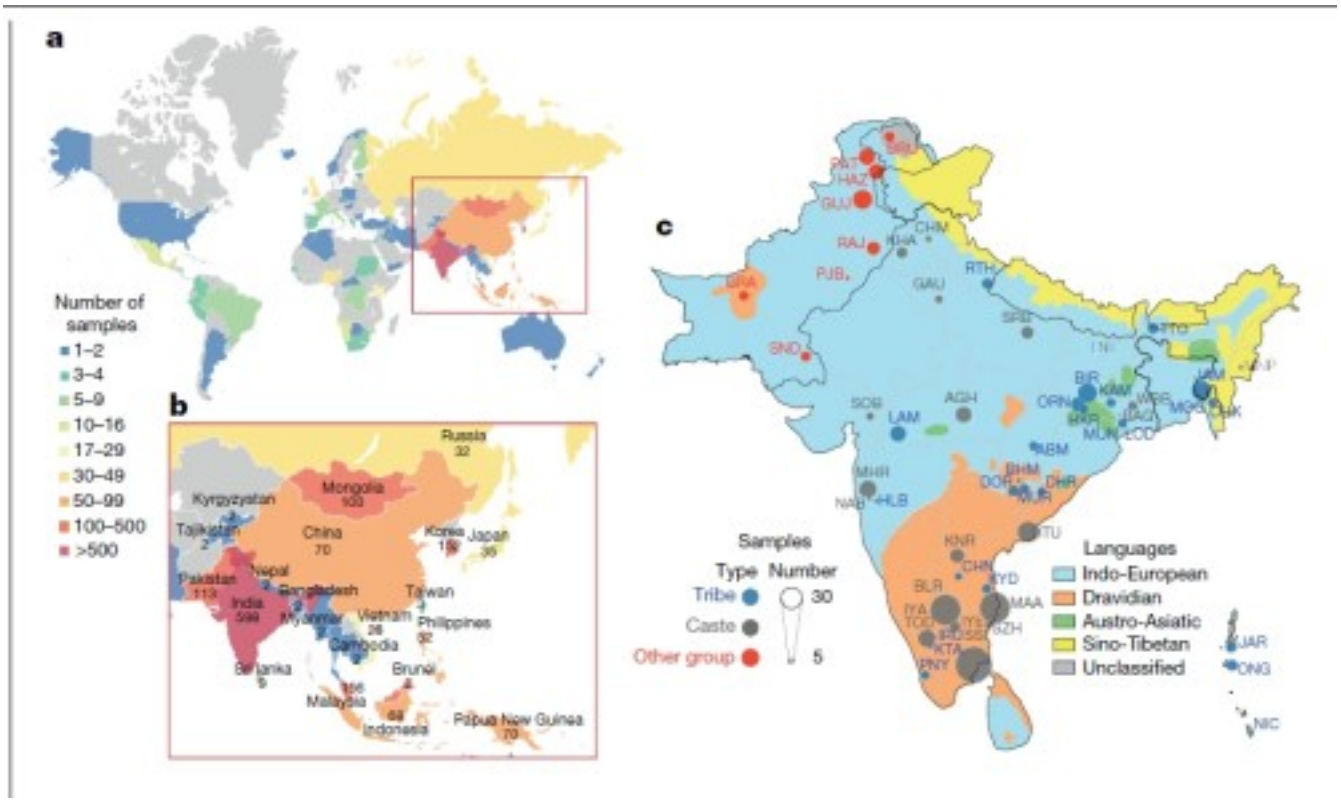
Азия включена в геномную картину мира

Опубликованы результаты пилотной стадии проекта GenomeAsia 100K. Проект направлен на расширение базы геномного разнообразия человечества за счет азиатских популяций. В его рамках секвенированы 1267 геномов из Южной, Юго-Восточной и Восточной Азии, составлен каталог геномных вариаций, описала популяционная структура и генетические ассоциации с заболеваниями.

Недостаточная представленность индивидов неевропейского происхождения в базе геномного разнообразия человечества ограничивает применение ее результатов для большой доли населения мира. Это побудило генетиков к созданию консорциума GenomeAsia, который координирует и объединяет работу отдельных исследовательских групп в большом международном проекте GenomeAsia 100K. Результаты пилотной стадии проекта (GenomeAsia Pilot, GAsP) опубликованы в журнале Nature. GAsP включает данные по секвенированию геномов 1739 человек из 219 популяций 64 стран Азии. По этим данным составлен каталог геномных вариаций, описала популяционная структура и генетические ассоциации с заболеваниями.

В рамках проекта GAsP были секвенированы 1267 геномов из Южной, Юго-Восточной и Восточной Азии со средним покрытием 36x. Вместе с ранее опубликованными данными по 596 геномам общая база данных GAsP (1739 геномов) включила 598 из Индии, 156 из Малайзии, 152 из Южной Кореи, 113 из Пакистана, 100 из Монголии, 70 из Китая, 70 из Папуа Новая Гвинея, 68 из Индонезии, 52 из Филиппин, 35 из Японии и 32 из России. Ученые анализировали их в контексте геномных данных из Европы, Африки и Америки, всего охвачены 7 регионов мира.



Географическое положение образцов, входящих в базу GAsP. На картах a (весь мира) и b (Юго-Восточная, Восточная и

Южная Азия) разными цветами указаны популяции с соответствующим числом образцов. c – Расположение, язык и социально-иерархические группы в изученных популяциях Южной Азии.

Для изучения популяционной структуры, обнаружения признаков недавнего отбора и прослеживания истории различных популяций авторы применили набор стандартных подходов — анализ главных компонент, ADMIXTURE, F-статистику, анализ гаплогрупп Y-хромосомы и митохондриальной ДНК и другие методы. Анализ ADMIXTURE продемонстрировал сложную генетическую структуру популяций Южной и Юго-Восточной Азии. В частности, население Индии, Малайзии и Индонезии имеет множество генетических корней и по большей части сформировалось в процессе смешения разных популяций. Авторы показали, что самое древняя отделившаяся популяция в Юго-Восточной Азии — это меланезийцы, она прослеживается на глубину 40 тыс. лет и примерно 20-30 тыс. лет назад отделилась от остальных.

Показано, что группы людей, антропологически определяемые как негритосы, из разных регионов — Индии, Малайзии и Филиппин, генетически более близки к своим географическим соседям, чем друг к другу. Этот факт говорит о том, что темная пигментация кожи – это, скорее, адаптация к условиям внешней среды (высокая солнечная радиация), чем показатель генетического родства.

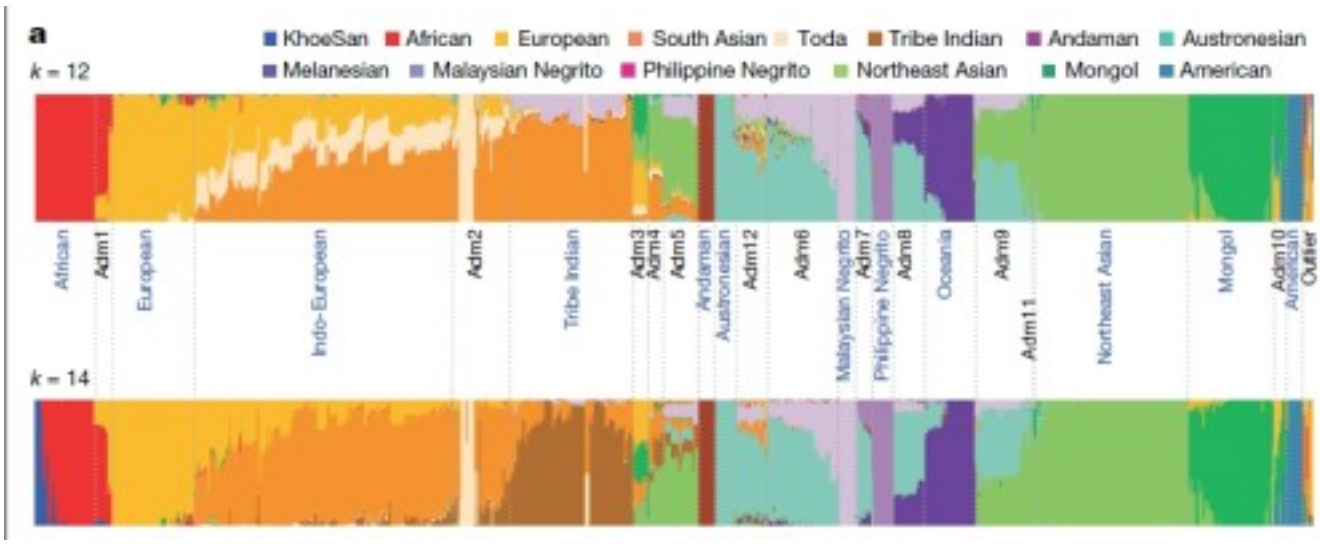
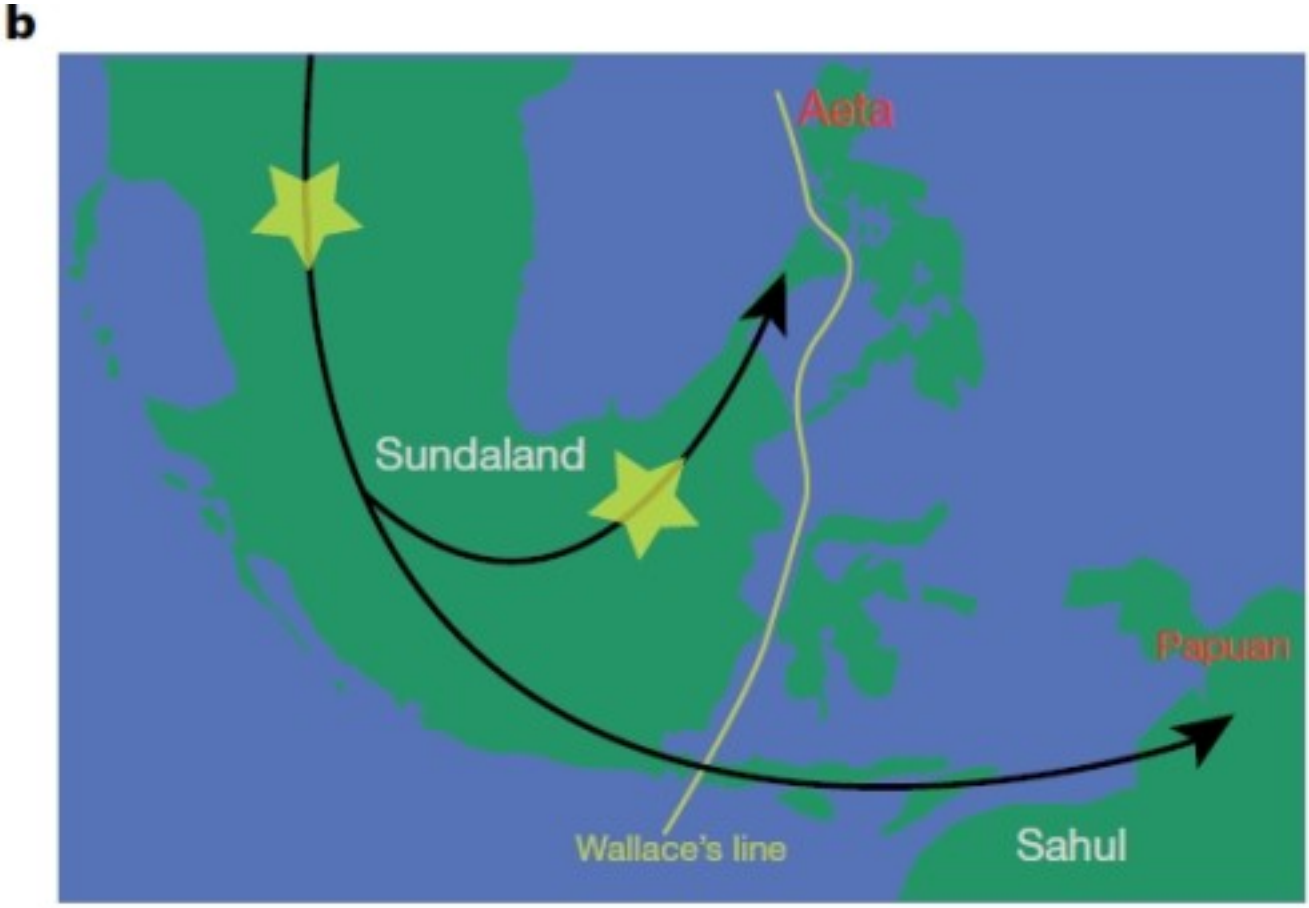


График анализа ADMIXTURE для k=12 и k=14 (заданное число предковых популяций). Генетические компоненты, обозначенные разными цветами, перечислены вверху.

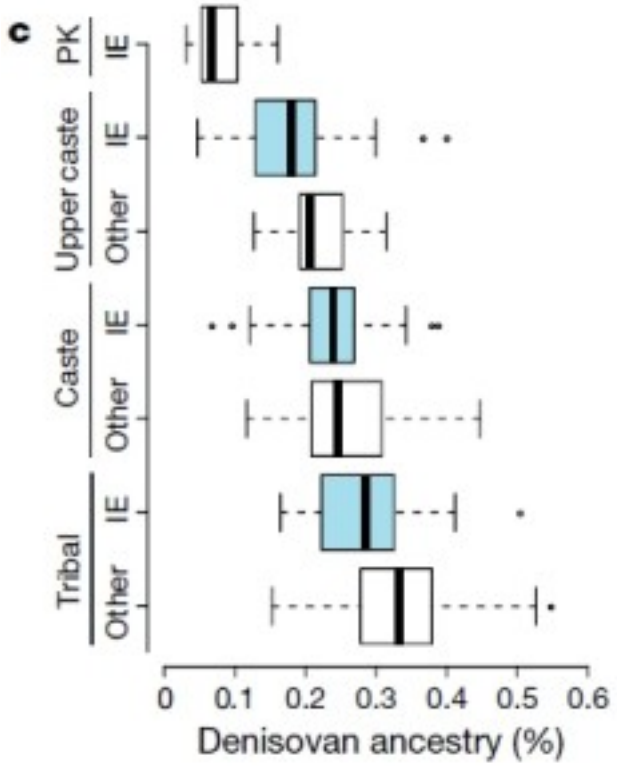
В рамках работы исследователи оценили величину генетического наследия денисовцев в разных популяциях. Оказалось, что он максимален у меланезийцев и народа аэта (проживающего на Филиппинских островах), несколько меньше у ати (также жители Филиппинских островов) и групп с индонезийского острова Флорес. В остальных популяциях Южной, Юго-Восточной и Восточной Азии денисовский генетический вклад меньше, хотя и не нулевой. У негритосов с Филиппин денисовский компонент выше, чем у малайских негритосов и негритосов с Андаманских островов. Популяция аэта, как показал мультифакторный анализ, несмотря на высокий денисовский вклад, генетически ближе к популяциям без такового (например, к малайским негритосам), чем к меланезийцам. Авторы объясняют это недавним потоком генов к аэта из популяций без денисовского вклада. С другой стороны, они предполагают, что популяция аэта получила два отдельных включения денисовских генов, второй – после разделения аэта и меланезийцев.



Предполагаемые пути современного человека в Юго-Восточную Азию в период Последнего ледникового максимума. Желтые звезды обозначают места предположительной метисации с денисовцами. Зеленым цветом обозначены участки, находившиеся в то время над водой, белая линия обозначает современную линию берега.

Для Южной Азии величина денисовского вклада оказалась несколько выше, чем предполагали ранее. Авторы проверили, не коррелирует ли она с языком, который используется населением, или социальным статусом (кастой). Они разделили популяции Южной Азии на группы, говорящие на индоевропейских и неиндоевропейских языках, а также проанализировали разные социально-культурные группы (трибы – этнические группы), и касты (высшие и низшие). Результаты показали, что степень денисовского генетического наследия различается в зависимости от языковой и социо-культурной принадлежности индивида. Во всех группах носители индоевропейских языков имели более низкий денисовский вклад, чем носители иных языков. По мнению авторов, это согласуется со сценарием, в котором группы индоевропейских мигрантов прибыли в Индию с северо-запада и смешались с коренными жителями.

(Хотя надо сказать, что на рисунке ниже разница величины денисовского вклада в разных группах населения не выглядит достоверной из-за большого доверительного интервала.)



Оценка денисовского компонента в группах населения Южной Азии, стратифицированных по социо-культурным группам и по языку.

Ученые оценили перспективы использования базы данных GAsP в медико-генетических исследованиях. Они сравнили GAsP с базами данных ExAC (Exome Aggregation Consortium), gnomAD29, 1000 Genomes Project, ESP (Exome Sequencing Project) и dbSNP, сфокусировавшись на кодирующих белки геномных вариациях. Оказалось, что более 23% вариаций из базы GAsP не были включены в выше перечисленные ресурсы. Можно было ожидать, что большинство новых кодирующих геномных вариаций будут редкими. Но оказалось, что абсолютное число новых вариаций в пределах собранной паназиатской базы велико (n=194 585), и они достаточно часты для того, чтобы учитываться в масштабных исследованиях геномных ассоциаций. Среди них есть вариации, представленные с низкой частотой в общей базе данных, но со значительно более высокой частотой среди отдельных групп популяций. Авторы нашли 144 329 новых вариаций с частотой >1% в популяциях Южной Азии, Юго-Восточной Азии, Северо-Восточной Азии и Океании.

В базе GAsP насчитывались 732 геномные вариации, не содержащиеся в базе Human Gene Mutation Database (HGMD), это 686 SNP и 46 инсерций/делеций (инделов) в 514 генах. Один из примеров новых вариаций – это HBB на 11 хромосоме, мутация, ассоциированная с β- талассамией, встречается почти исключительно в Южной Азии и с низкой частотой в Юго-Восточной Азии. Было найдено 13 геномных вариаций в 6 генах, ассоциированных с риском рака. Они вызывали сдвиг рамки считывания, остановку транскрипции и нарушение сплайсинга в нескольких ассоциированных с раком генах: *BRCA2*, *BRCA1*, *ATM*, *BLM*,*NBN* и *PMS2*. Описали также несколько геномных вариаций, вовлеченных в ответ на действие лекарств и фармакокинетику.

В этой части работы по своей базе данных авторы вычислили частоты аллелей основных генов фармакогеномики для определения внутрипопуляционных различий, потенциально важных для клинических испытаний лекарств и для терапии. Так, около 400 млн человек, принадлежащих к австронезийским популяциям, имеют генетически высокий риск применения карбамазепина из-за повышенной чувствительности к нему, в их числе жители Индонезии, Малайзии и Филиппин.

Полученные в проекте данные показали, что во многих популяциях Юго-Восточной, Южной и Восточной Азии велик «эффект основателя» (радикальное снижение генетического разнообразия), возникший из-за прохождения популяции через «бутылочное горлышко» с резким снижением ее эффективного размера. Это сопровождается увеличением частоты рецессивных заболеваний. Большинство геномных вариаций, нарушающих синтез белка, в том числе с полной потерей функции гена, содержатся в популяциях в гетерозиготном состоянии. Но встречаются и гомозиготные мутации с потерей функции гена, это позволяет оценить фенотипический эффект потери гена и дает полезную информацию для терапии заболеваний. В дополнение к гомозиготным мутациям с потерей функции гена, включенным в базы gnomAD, dbSNP и 1000 Genomes Project (n = 856), авторы нашли еще 121 ранее не описанную геномную вариацию такого типа.

Увеличение числа и масштабов исследований геномных ассоциаций, выполненных на азиатских популяциях, расширит понимание генетической основы заболеваний человека, пишут ученые. Пилотная стадия проекта GenomeAsia, помимо непосредственно полученных в ней результатов, важна для стратегии будущих исследований. Эта стратегия включает создание популяционно-специфических геномных баз данных и популяционно-специфических SNP-чипов. Поставлен также широкий круг вопросов относительно генетической истории отдельных популяций Азии.

текст Надежды Маркиной

Источник:

GenomeAsia100K Consortium. The GenomeAsia 100K Project enables

genetic discoveries across Asia // Nature https://doi.org/10.1038/s41586-019-1793-z