

НОВОСТИ НАУКИ

ГЕНОМИКА. МНОГООБРАЗИЕ ЧЕЛОВЕЧЕСКОЙ ПРИРОДЫ.

Когда-то английский писатель Сомерсет Моэм, будучи студентом медицинского института, не смог на занятии по анатомии найти нерв там, где он должен был находиться, его преподаватель показал ему этот "спрятанный" нерв, сказав, что нормальное расположение нерва встречается крайне редко. В октябрьском выпуске журнала "Nature" за 2005 год (т.437, стр. 1299-1320) опубликована статья **"A haplotype map of the human genome"** (гаплотипная карта генома человека), в которой международный консорциум по гаплотипированию карт приводит подробный анализ одной из важнейших причин отсутствия "нормального типа" человека.

Геном человека насчитывает около 10 миллионов полиморфизмов, называемых генетическими вариантами, в которых минорные формы генов представлены одним на 100 возможных генетических форм. У любых, не связанных друг с другом, двух индивидуумов встречаются миллионы генетических различий, благодаря которым люди выглядят и ведут себя по-разному. Такие различия человек унаследовал из своего эволюционного прошлого. Наряду с ними генетика значительно влияет на здоровье человека.

Современные генетики добились больших успехов в процессе выявления генетических аномалий, являющихся причиной многих наследственных заболеваний, таких, например, как кистозный фиброз или болезнь Тея-Сакса. Такие аномалии встречаются сравнительно редко. Генетические нарушения, такие как рак или сердечно-сосудистые и нейродегенеративные заболевания, в большинстве случаев многокомпонентны. Такие распространенные генетические заболевания, по всей видимости, неизбежны, поскольку некоторые из них поражают человеческий организм на раннем этапе жизни и поддаются лечению гораздо лучше, а количество других, проявляющихся позже, например, деменция, стремительно растёт.

Такие распространенные наследственные заболевания сложны по своей природе и до сегодняшнего времени были недоступны для генетического анализа. Нет сомнений в том, что генетика играет чрезвычайно важную роль. Судя по результатам многочисленных исследований, генетические мутации объясняют более 40% вариаций наиболее распространенных заболеваний в популяции и более 70% ряда нарушений, например, шизофрению. Вся сложность заключается в том, что в сочетании с другими генными вариантами и факторами окружающей среды такие вариации только предрасполагают человека к определенным заболеваниям, но не являются причиной их возникновения.

Каким образом ученые определяют, какой именно из 10 миллионов полиморфизмов влияет на заболевание? Поскольку полиморфизмы в геноме человека зачастую зависимы друг от друга, нет необходимости анализировать все 10 миллионов образцов и определять их взаимосвязь с самим заболеванием. Когда появляется мутация, это связывают с особыми вариантами, присутствующими на той же самой хромосоме (варианты, которые объединяются воедино, принято называть "гаплотипом"). По этой и ряду других причин между полиморфизмами зачастую возникает статистическая взаимосвязь - присутствие определенного варианта на одном участке хромосомы может предсказать или "маркировать" присутствие определенного варианта на другом участке.

Основная цель HarMap Consortium состояла в том, чтобы обнаружить эту взаимосвязь среди генных вариантов, и ученым удалось это сделать самым впечатляющим образом. В ходе первого этапа проекта, о котором сообщается в упомянутой выше статье, проведено компилирование данных относительно генетического состава, или генотипа, из групп индивидуумов-представителей четырех популяций, насчитывающих более миллиона однонуклеотидных полиморфизмов (single nucleotide polymorphisms или SNPs). (SNPs - это один из наиболее распространенных типов генных вариантов в нашем геноме, в отличие от включений и делеций). Необходимо было гарантировать, что один SNP прошел испытание для каждого из 5000 оснований последовательности. В течение второго (пока не завершённого) этапа проекта количество SNP значительно возрастет.

Используя концептуально простой и надежный способ отбора, группа HarMap предположила, что число маркеров, необходимых для представления обычных вариантов по всему геному, составляет менее одной десятой от общего количества на таких участках. Вероятно, более сложные методы отбора окажутся еще более эффективными.

Проект HarMap продвинулся вперед и в других областях. Три года назад ученым было известно менее 1,7 миллионов полиморфизмов. Сегодня благодаря проектным усилиям их количество составляет более 8 миллионов. Знание учеными большинства полиморфизмов может дать огромные преимущества. Появилась возможность применить прикладные методы биоинформатики к этим данным и уделить особое внимание полиморфизмам функциональной группы: например, сосредоточить усилия на исследовании таких полиморфизмов, которые приводят к изменениям аминокислотной последовательности белков, или они локализованы в функционально важных хромосомных областях (отобранных по степени сохранности такого хромосомного участка на всем протяжении биологического вида), либо особо значимых полиморфизмов с использованием других геномных критериев.

Каким образом достигнутые результаты повлияют на изучение болезней человека и различной ответной реакции на предложенное лечение? Ещё до запуска проекта HarMap для ответа на вопрос, влияет ли общее генетическое изменение в молекулярной мишени лекарственного средства на ответную реакцию пациента, пришлось бы заново секвенировать ген в группе индивидуумов - представителей исследуемой популяции для идентификации необходимого количества полиморфизмов. Например, такого рода исследование было проведено группой ученых из Института геномных наук (Северная Каролина, США) для гена *SCN1A*, кодирующего белок-мишень антиэпилептических лекарственных препаратов. Учёным потребовалось два года для идентификации обычных полиморфизмов и соответствующих маркеров. Сегодня такой же анализ можно провести с помощью простых компьютерных алгоритмов за считанные минуты, используя данные HarMap.

Как повлияют результаты проводимых исследований на здоровье человека? Ученые знают, что генетика влияет на возникновение и протекание определенных заболеваний, в то же время доподлинно не известно, какие именно генные варианты являются ответственными за развитие болезни. Если они представляют собой обычные варианты (то есть один на 10 миллионов), новые разработанные методы значительно ускорят идентификацию генетического варианта, связанного с заболеванием. В том случае, если "ответственные" варианты встречаются реже, обнаружить их будет гораздо сложнее.

Существует мало прямых доказательств того, что идентификация факторов риска для ряда заболеваний человека способна улучшить общее состояние здоровья населения. В тех редких случаях, когда обнаруживали "очевидные" факторы риска, они, как правило, оказывались бесполезными в лечении или предотвращении того или иного заболевания. Например, вариант гена *APOE* - характерный признак позднего проявления болезни Альцгеймера, однако пока не обнаружено никаких характерных особенностей образа жизни или диеты,

изменение которых сопровождают болезнь. Остается надеяться на то, что благодаря факторам риска будут разработаны новые перспективные методы лечения подобных заболеваний, хотя и в этом случае успешных примеров мало.

В ближайшем будущем будет проведена клиническая идентификация генетических показателей ответной реакции пациентов на лечение. Полиморфизмы могут существенно повлиять на такие реакции, в результате чего ученые смогут предложить альтернативные методы лечения. Существует небольшая разница в общей эффективности различных нейролептических препаратов (лекарств нового поколения по сравнению с лекарствами, применявшимися десять лет тому назад). Однако следует отметить огромную разницу между ответной реакцией на различные препараты у разных пациентов. Если бы ученые смогли идентифицировать варианты гена, предсказывающие, например, вызывают ли новые препараты нежелательную прибавку в весе, а старые препараты - нарушения двигательной функции, тогда способы лечения можно было бы подбирать индивидуально. Проект НарМар значительно облегчит поиск таких вариантов.

В настоящее время геномная наука представляет собой своего рода "неуклюжего подростка". Возможности современной геномики поистине поразительны. За несколько лет ученые прошли путь от полного незнания того, что может быть охарактеризовано "геномным" (речь идет скорее о целых геномах, нежели об отдельных генах), к наличию законченных последовательностей генома для многих организмов, а теперь уже располагают полным каталогом общих генетических различий у человека. Тем не менее, технические достижения - это еще не показатель зрелости в науке. На следующем этапе геномных исследований необходимо прежде всего сосредоточить внимание на биологическом понимании и применении полученных результатов в клинической практике. Специалистам в области геномики пора применить технологические достижения на практике.

ПРОНИКНОВЕНИЕ В ГЕНОМ

Международный проект НарМар Consortium систематизировал более 1 миллиона однонуклеотидных полиморфизмов (SNP) в геномных последовательностях 269 человек - представителей четырех различных популяций. Большинство этих SNP непосредственно на функцию генов не влияют, однако дают ценную информацию об общей модели хромосомной организации и новые методы обнаружения генов, вызывающих то или иное заболевание у человека.

Еще один, наиболее прямой способ обнаружения таких генов - "медицинское повторное секвенирование" (MRS). Ключевые фрагменты предполагаемых генов секвенируют и сравнивают между пациентами и контрольной группой для идентификации вероятных генетических вариаций, относящихся к заболеванию. Такой подход весьма популярен и результативен, но в его основе лежит подбор правильного гена-кандидата в самом начале. Популярность MRS возрастает, и несколько крупных центров по секвенированию уже анализируют сотни индивидуальных генов. Сейчас активно обсуждается использование MRS для анализа всех известных на сегодняшний день генов человека (более 20000) в отобранных группах пациентов и контрольных группах. Это значительно увеличит вероятность успешных поисков гена, вызывающего определенное заболевание.

В таких широкомасштабных проектах предпочтительнее действовать централизованно, по аналогии с проектом расшифровки генома человека. MRS-активность должна увеличиваться, тем не менее, основной целью остаётся обнаружение генетической вариации в общей популяции (не только у пациентов и

в контрольной группе), непосредственно влияющей на генную функцию. Во время этого этапа секвенирования нет необходимости знать статус заболевания анализируемого образца. Этот подробный каталог генетических изменений, названный "функциональные варианты", включает SNP, изменяющие аминокислотные последовательности в белках, а также, возможно, сплайсинг или уровни экспрессии. Большинство из вышеназванных мишеней более редкие, по сравнению с теми, которые изучались в ходе реализации НарМар. Такая база данных по "функциональным вариантам" станет немедленно доступна для всех учёных и окажет огромное влияние на будущие генные исследования подобного рода.

Некоторые функциональные варианты, вызывающие заболевания, уже каталогизированы благодаря исследованиям менделевских болезней - состояния человека с простым наследованием (один ген). Полимеразная цепная реакция (ПЦР), флуоресцентное секвенирование ДНК и другие технологии позволили обнаружить около 1700 генов менделевских заболеваний, большинство из которых имеют множественные функциональные варианты. Такие заболевания достаточно редкие, примерно 1 человек на 10000 и даже на 100000. Частота менделевских функциональных вариантов в обычной популяции составляет менее 0,05%.

Данные НарМар Project помогут обнаружить функциональные варианты, вызывающие развитие таких заболеваний, как сердечно-сосудистые заболевания, онкологические и шизофрения. Такие нарушения в большинстве случаев имеют сложную полигенную природу, и технологии НарМар лучше всего проявляют себя, когда задействованные функциональные варианты встречаются в большой популяции с частотой более 5%.

Процесс выявления определенного SNP называется "генотипированием", и проект НарМар идентифицировал набор предполагаемых 3 миллионов SNP, которые могут быть идентифицированы в процессе генотипирования. Проект также идентифицировал еще один набор "меченых" одонуклеотидных полиморфизмов, наиболее полезных в генетических ассоциативных исследованиях, так как они связывают обычные гаплотипы - крупные фрагменты ДНК, наследуемые вместе. С применением этих "меток" скрининг целых геномов в образцах пациентов становится более реальной задачей, чем, скажем, три года назад.

Когда функциональные варианты, вызывающие заболевание, "снижаются" до 5% в пораженной популяции, методы, предложенные НарМар, не действуют и НарМар оказывается бессилен. Такие ограничения НарМар вполне ожидаемы, поскольку форма распределения SNP в популяциях человека подчиняется закону распределения: реже встречаются SNP с наиболее полной вариацией. Редкие функциональные варианты, относящиеся к менделевским нарушениям, попадают в эту категорию. Обычные SNPs включают несколько идентифицированных функциональных вариантов, вызывающих заболевание.

Маркеры в пределах 0,05-5% менее изучены, прежде всего потому, что их обнаружению препятствуют технические трудности. Тем не менее, именно эти маркеры влияют на заболевания человека - теоретическое моделирование подтверждает их значимость, и возникает прямое тому доказательство на примере заболеваний. Cohen и его коллеги (J.Cohen, A.Pertsemlidis, I.K.Kotowski, R. Graham, C.K.Garcia, H.H.Hobbs (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nature Genetics 37(2),161-165) провели исследование отдельных представителей, генетически подверженных отложениям холестерина и, следовательно, сердечно-сосудистым заболеваниям, в крови которых в дальнейшем проанализировали уровень липопротеинов низкой плотности (LDL). Ученые пришли к выводу, что функциональные варианты в главных генах легче всего обнаружить у пациентов с самым высоким и самым низким уровнем LDL. При помощи MRS Cohen обнаружил у обследуемых с низким уровнем LDL пару функциональных вариантов в главном гене, PCSK9. Интересно, что найденные функциональные варианты с низким уровнем LDL соответствовали категории "очень редких мутаций", однако встречались

достаточно часто в общей популяции, рассматриваемой как "низкочастотная вариация". Эти функциональные варианты невозможно идентифицировать при помощи генотипирования HarMap.

Глобальный поиск предполагаемых функциональных вариантов предусматривает логическое расширение возможностей HarMap проекта. Ученые из медицинского колледжа Baylor (Houston, Texas) предлагают помимо HarMap проекта создать другой крупномасштабный геномный проект по каталогизированию редких генетических вариаций. В этом направлении уже работают несколько общественных и частных групп, однако необходима более мощная программа по изучению различных популяций человека, независимо от заболевания и фенотипа, которая ускорила бы открытие низкочастотных функциональных вариантов.

Первостепенная задача этого проекта - создание базы данных функциональных вариантов при помощи MRS исследований, направленных на кодирующие области всех 20000 известных генов человека, а также дополнительного участка предполагаемого промотора для каждого из них. Исследование 2000 индивидуальных образцов ДНК (включая 4000 хромосом) поможет обнаружить генетическую вариацию, представленную на уровне 0,05% в общей популяции. Такой смелый шаг позволит зафиксировать большинство предполагаемых функциональных вариантов (от 50000 до 100000), обнаруженных у человека.

Отобранные популяции должны включить представителей Африки, Азии и Кавказа, а также латиноамериканские и другие родственные группы. Оптимально представить различные популяции достаточно сложно.

Для этого крупномасштабного проекта потребуется модернизация технологий MRS, чем уже занимаются специалисты. С технической точки зрения, MRS проще, чем секвенирование *de novo*, поскольку всегда можно применить для сравнения "последовательность-мишень". Это означает, что стоимость MRS работ будет снижаться быстрее, чем предполагалось для других геномных проектов.

Важно, что "метод функциональных вариантов" позволит избежать многочисленных исследований заболеваний, используя возможности крупных геномных центров в области секвенирования для выстраивания первоначальной базы данных. Всеобщей целью станет идентификация генетических вариаций, которые могут быть впоследствии протестированы при помощи аналогичных методов генотипирования, использованных для HarMap. Предполагается, что частные компании запустят производство так называемых "ДНК зондов" для проведения испытаний генетических маркеров в новых популяциях, что уже произошло с существующей базой данных 20000 SNPs, доступных научной общественности. Зонды могут быть распределены в отдельные исследовательские лаборатории как часть научного проекта по изучению каждого текущего заболевания.

Проект функциональных вариантов противопоставлен крупномасштабному MRS-исследованию, изучающему популяции пациентов, как, например, Cancer Genome Project, который стремится каталогизировать все основные мутации, встречающиеся при наиболее распространенных онкологических заболеваниях.

Исследование, проведенное в 2004 году, показало, что для идентификации анонимного пациента могут быть использованы всего 75 SNPs. Поскольку интенсивный генетический анализ образцов заболевания требует одновременного секвенирования нескольких генов у каждого пациента, идентификация не представляет сложности.

Проект функциональных вариантов имеет и другие преимущества по сравнению с крупномасштабным MRS-проектом относительно популяций пациентов. Хотя существуют тысячи хороших тканевых образцов в "семействах заболеваний", большинство из них рассредоточены по лабораториям различных ученых. В вопросе фенотипирования и хранения практически нет согласованности, протоколы по сбору данных и этические нормы не стандартизованы. Централизации таких коллекций, выполнению MRS и выпуску

данных в общественный доступ, препятствуют множество материально-технических проблем и правил Институтского наблюдательного совета (США). Клинические врачи, назначающие кредит на проведение поиска и сбора образцов, который может продлиться годами, чрезвычайно обеспокоены тем, что все внимание сосредоточено на новых исследованиях в области секвенирования. База данных функциональных вариантов избегает большинство из названных проблем путем разделения первоначального этапа секвенирования и лабораторных исследований определенного заболевания.

Один недостаток Cancer Genome Project состоит в том, что каждая новая обнаруженная мутация может присутствовать только у небольшой части клеток в мельчайшем тканевом образце. С целью подтверждения мутации к раку у каждого пациента должен быть проверен образец здоровой ткани. Это позволит исключить возможность редкой наследственной вариации. Такие сложности никоим образом не соотносятся с базой данных функциональных вариантов.

По сегодняшним расценкам, с учетом флуоресцентного секвенирования ДНК (PCR Sanger), стоимость исследования полного набора кодирующих генов для 2000 пациентов составит около 750 миллионов долларов США. Можно ожидать постепенного закономерного снижения стоимости MRS-исследований, и если проект продлится еще пять лет, затраты составят менее 500 миллионов долларов. Эта цифра в пять раз превышает средства, выделенные на проект HarMap, однако она ниже стоимости Cancer Genome Project (1, 35 миллиардов долларов).

Любое предложение для проекта MRS ставит перед геномикой знакомую дилемму - как сбалансировать высокоэффективную производительную способность секвенирования крупных геномных центров и действенные научные методы, предлагаемые мировым сообществом? Среди всех возможных будущих проектов исследование функциональных вариантов значительно улучшит и укрупнит взаимное сотрудничество биомедицинских исследователей так же, как в своё время учёных объединил Проект генома человека.

По материалам журнала "Nature" при участии Рыженковой О.Н.