

НОВОСТИ НАУКИ

МЕЖДУНАРОДНЫЙ КОНСОРЦИУМ ГАПМЭП (HAPMAP PROJECT) НАИБОЛЕЕ ЗНАЧИМЫЕ МЕЖДУНАРОДНЫЕ ПРОЕКТЫ.

*После завершения проекта "Геном человека" наступило время нового
глобального проекта - Геп-Мэп.*

По сообщению журнала Nature (426, 2003), недавно консорциум, участвующий в работе по данному проекту, определил некоторые важные принципы для доступа к данным и финансирования работ по проекту.

ГепМэп - рассчитанный на три года международный проект, финансируемый правительственными и неправительственными (частными) организациями Японии, Великобритании, Канады, Китая, Нигерии и США. Его целью является картирование т.н. гаплотипов человека, которые позволят описать общие паттерны вариаций нуклеотидных последовательностей ДНК человека. По мнению организаторов, ГепМэп должен стать ключевым ресурсом для исследователей, который позволит находить гены, влияющие на здоровье, развитие болезни и ответы на лекарственные средства и факторы окружающей среды. Проект, стоимость которого оценивается в 100 миллионов долларов, официально стартовал в октябре 2002 г.

Это следующий шаг, предпринятый учеными после выяснения последовательности одного генома человека. Международное научное сообщество, занимающееся вопросами геномики человека полагает, что настало время "каталогизировать" различия между людьми и расшифровать их роль в норме и патологии. Поэтому ученые пытаются создать карту общих отличий в геноме человека, используя комбинации маркеров, которые дают мозаичные рисунки, называемые "гаплотипами" (см. также Nature 421, 409-412; 2003).

Международный проект "ГепМэп" (ГМ) (перевод англоязычной аббревиатуры HapMap - haploid mapping - гаплоидное картирование) представляет собой широкомасштабное сотрудничество государственных и частных научно-исследовательских учреждений, которые будут исследовать 270 геномов человека из 4-х популяций, чтобы создать общедоступный инструмент для использования всеми учеными, занимающимися биомедицинскими проблемами.

Учитывая, что роль каждого партнера достаточно четко очерчена, а также тот факт, что проект позволит обеспечить быстрое и полное получение важных данных, возникает естественный вопрос, что получают те, кто выполняет эту работу.

В январе 2003 г. представители государственного фонда ученых в области геномики, встретились в Форт Лодердейл (штат Флорида, США) чтобы разрядить напряжение между открытым доступом к широко-масштабным базам данных и вкладом в финансирование по проекту. Предлагаемое решение должно было состоять в определении ряда проектов как "государственно-финансируемых проектов", в которых различные правила должны применяться при распределении данных для пред-публикаций и публикаций. В качестве примеров приводятся

недавние государственно-профинансированные проекты секвенирования генома мыши и человека и международный проект ГапМэп. Для государственно-финансируемых проектов группа предложила новый тип публикаций, названный "описание проекта", целью которой является "информировать научное сообщество о новом научном проекте и обозначить источник получения данных".

Каждый, кто использует ГМ-данные в своих собственных публикациях - даже до того как получатели первичных данных опубликовали результаты их анализа - могут теперь должным образом цитировать планы авторов-создателей. Другой положительный момент состоит в том, что многие ученые на всех уровнях, которые уже сейчас вносят свой вклад в важный проект, могут это указать в своих CV.

Подобно многим другим широкомасштабным проектам, процесс получения результатов достаточно сложен и многие проблемы, в частности, проблема оптимального подхода к анализу огромных масс данных, все еще остаются нерешенными.

Журнал Nature публикует статью, представленную международным ГМ-консорциумом, в которой описана стратегия в отношении получения согласия пациентов на его участие в генетических исследованиях и в получении образцов, а также разрешение типичных в таких случаях конфликтов, возникающих при организации быстрого и свободного доступа к данным, в то же время избегая возможных проблем с патентованием.

По мнению журнала, такая публикация может начать дискуссию, которая окажется полезной исследователям данной проблемы.

Как пишут в своей статье, опубликованной в декабрьском выпуске журнала Nature за 2003 (том 426 стр. 789-796) участники этого проекта, целью Международного Проекта ГМ является определение общих паттернов вариации в последовательности ДНК генома человека и предоставление возможности свободного доступа к этой информации. Международный консорциум разрабатывает карту таких паттернов по всему геному - на основе определения генотипов более 1 млн вариантов нуклеотидных последовательностей индивидуальных геномов, частоты их встречаемости и степени ассоциации между ними, используя образцы ДНК, взятые у населения, живущего в отдельных районах Африки, Азии и Европы. ГМ даст возможность определить варианты нуклеотидной последовательности, которые свойственны наиболее часто встречающимся заболеваниям. Это может облегчить разработку орудий диагностики и расширит возможности медиков в плане выбора мишени для терапевтического вмешательства.

Социально значимые болезни, такие как сердечно-сосудистые заболевания, рак, ожирение, диабет, психические нарушения и др., обусловлены сочетанием множества генетических и природных факторов. Выяснение этих генетических факторов даст возможность лучше понять проблемы патогенеза, диагностики и лечения многих болезней человека. Поиск общих вариантов в хромосомных районах, которые (варианты) идентифицируются с помощью анализа сцепленности, оказался вполне успешным в случае многих редких "моногенных" заболеваний, однако подобного рода исследования оказались гораздо менее успешными при определении и выявлении генетических вариантов, вызывающих мультигенные заболевания, поскольку каждый вариант по отдельности, оказывает лишь незначительное влияние на риск возникновения данного заболевания. При идентификации этих генетических риск-факторов используется дополнительный подход, основанный на поиске связи между специфическим вариантом и самой болезнью - путем сравнения группы заболевших лиц с контрольной группой здоровых людей. При отсутствии ярко-выраженной естественной селекции, очевидно, можно предположить наличие широкого спектра частотности таких вариантов, многие из которых, возможно, являются обычными для населения.

Целый ряд так называемых ассоциативных исследований по выявлению

связи между специфическим вариантом и самой болезнью, направленных на поиск генов-кандидатов, сцепленных районов, ассоциированных с болезнью, уже привели к открытию генетических риск-факторов для социально значимых болезней, таких, например, как сахарный диабет 1-го типа (антиген лейкоцита человека (HLA), инсулин и CTLA4), болезнь Альцгеймера (APOE), глубокий тромбоз вен (фактор V), заболевания кишечного тракта (NOD2), а также 5q31, гипертриглицеридемия (APOAV), сахарный диабет 2-го типа (PPARG), шизофрения (нейрегулин), астма (ADAM33) и инфаркт миокарда (LTA).

Один из подходов в ассоциативных исследованиях включает тестирование каждого предполагаемого причинного варианта на его корреляцию с болезнью (прямой подход). Поиск ассоциаций с болезнью по всему геному приведет к существенным расходам: для идентификации кандидатов-вариантов придется провести секвенирование всего генома путем изучения многочисленных проб (ткани, сыворотки больных). В настоящее время этот подход ограничивается секвенированием функциональных частей генов-кандидатов (отобранных на основе предыдущей функциональной или генетической гипотезы) для потенциальных, ассоциируемых с болезнью кандидатов-вариантов. Поэтому был предложен альтернативный (непрямой) подход, согласно которому набор вариантов нуклеотидных последовательностей в геноме может быть использован в качестве генетических маркеров для выявления ассоциации между данным районом генома и болезнью, независимо от того, являлись ли сами маркеры носителями функциональных эффектов или нет. Поиск ассоциированных с болезнью вариантов может в этом случае быть ограничен геномными районами демонстрирующими связь с болезнями. Два подхода, сформировавшиеся в ходе исследований генетики человека, свидетельствуют о том, что непрямой подход способен охватить большую часть вариаций нуклеотидных последовательностей (если речь идет о геноме человека) и с большей эффективностью, чем прямой подход. Во-первых, приблизительно 90% таких вариаций среди отдельных лиц относится к общим (обычным) вариантам.

Во-вторых, большинство из них возникло в результате отдельных исторических мутационных событий и, следовательно, они связаны с близлежащими вариантами, которые присутствовали на родительской хромосоме в том месте, где происходила мутация. Эти ассоциации делают непрямой подход пригодным для изучения вариантов в генах-кандидатах, а также в районах и по всему геному. Предварительного знания возможных функциональных вариантов не требуется. Вместо этого, данный подход использует информацию, полученную из относительно небольшого количества вариантов, охватывающих большинство общих вариационных паттернов в геноме, в результате чего любой район или ген может быть протестирован на предмет его связи (ассоциации) с конкретной болезнью.

Целью международного ГМ-проекта является выявление общих паттернов вариаций нуклеотидных последовательностей ДНК в геноме человека - с помощью характеристики определенных вариантов нуклеотидных последовательностей, частоты их возникновения и корреляции между ними в образцах ДНК, взятых у населения отдельных районов Африки, Азии и Европы. Проект таким образом, предоставляет реальную возможность применить непрямой ассоциативный подход к любому функциональному ген-кандидату в геноме, к любому району, который является наиболее перспективным в плане анализа сцепленности, основанного на семейных связях и, наконец, ко всему геному - с целью сканирования его на предмет выявления факторов риска для данного заболевания.

С помощью этой стратегии будут наиболее полным образом охвачены общие варианты, ответственные за риск для данного заболевания, но не все предрасполагающие варианты являются общими. Следует отметить, однако, что даже относительно необычный, ассоциируемый с болезнью вариант может потенциально быть обнаружен с помощью этого подхода. Размышляя о его

исторических корнях, можно предположить, что необычный вариант способен "путешествовать" по хромосоме, которая несет на себе характерный паттерн близлежащих вариантов нуклеотидной последовательности. В группе людей, пораженных болезнью, редкий вариант обычно повторяется с большей частотой, чем в контрольной группе здоровых людей. Это наблюдение, в частности, очень помогло при идентификации генов, ответственных за фиброзно-кистозную дегенерацию (муковисцидоз) и дистрофическую дисплазию, после того как анализ сцепленности указал на общий хромосомный район.

Что такое варианты нуклеотидных последовательностей человека?

Любые 2 экземпляра человеческого генома отличаются друг от друга 0,1% нуклеотидных участков (т.е. в среднем 1 вариант на 1 тыс. оснований). Наиболее обычный тип варианта, известный как полиморфизм точечных мутаций (замен нуклеотидных оснований) или SNP (от англ. single nucleotide polymorphism), заключается в различии между хромосомами по одному азотистому основанию (рис.1). Например, некоторые хромосомы в определенной группе населения могут иметь на этом участке цитозин (С)

("С - аллель"), в то время как другие содержат тимин (Т) ("Т-аллель"). Среди всего населения Земли, около 10 млн участков (т.е. в среднем 1 вариант на 300 азотистых оснований) варьируют таким образом, что оба аллеля наблюдаются с частотой $\geq 1\%$, эти 10 млн общих SNP составляют около 90% вариаций населения Земли. Оставшиеся 10% обусловлены редко встречающимися вариантами. Наличие данных SNP-аллелей у индивидуума определяется генотипированием геномного образца ДНК.

Едва ли не каждый вариабельный участок возникает в результате единичного исторического мутационного события: скорость мутации очень низка (порядка 10^{-8} на участок для одного поколения) относительно количества поколений, прошедших с момента появления самого недавнего общего предка для любой пары людей (порядка 10^4 поколений). По этой причине, каждый новый аллель изначально ассоциируется с другими аллелями, оказавшимися в наличии на том конкретном хромосомном фоне, на котором он возник. Специфический набор аллелей, наблюдаемых на единичной хромосоме, или части хромосомы, называется гаплотип (рис.1 б). Новые гаплотипы формируются путем дополнительных мутаций или путем рекомбинации, когда материнская и отцовская хромосомы обмениваются соответствующими сегментами ДНК, в результате чего хромосома становится мозаичным соединением двух родительских гаплотипов.

Со-наследование SNP-аллелей на этих гаплотипах ведет к ассоциациям между этими аллелями у населения (так называемое неравновесное сцепление, linkage disequilibrium, LD). Поскольку вероятность рекомбинации между двумя SNP возрастает по мере увеличения расстояния между ними, в среднем, количество таких ассоциаций между SNP уменьшается при увеличении расстояния. Многими эмпирическими исследованиями выявлены значимые уровни LD и часто - прочные ассоциации между соседними SNP генома человека. Наличие таких прочных ассоциаций означает, что во многих хромосомных районах имеется всего несколько гаплотипов и эти гаплотипы составляют большинство вариаций среди людей в этих регионах.

Прочные ассоциации между SNP в конкретном регионе имеют практическое значение: генотипирование лишь нескольких тщательно отобранных SNP в регионе дает достаточную информацию для предсказания информационных данных относительно оставшихся обычных SNP в этом регионе. В результате, для идентификации каждого из обычных гаплотипов в регионе требуются только несколько из этих "ярлыковых" (таговых от англ. tag - ярлык, метка).

Поскольку степень ассоциации между соседними маркерами сильно различается в пределах генома, использовать SNP, выбранные наугад, или равноудаленные друг от друга в геномной последовательности, нецелесообразно.

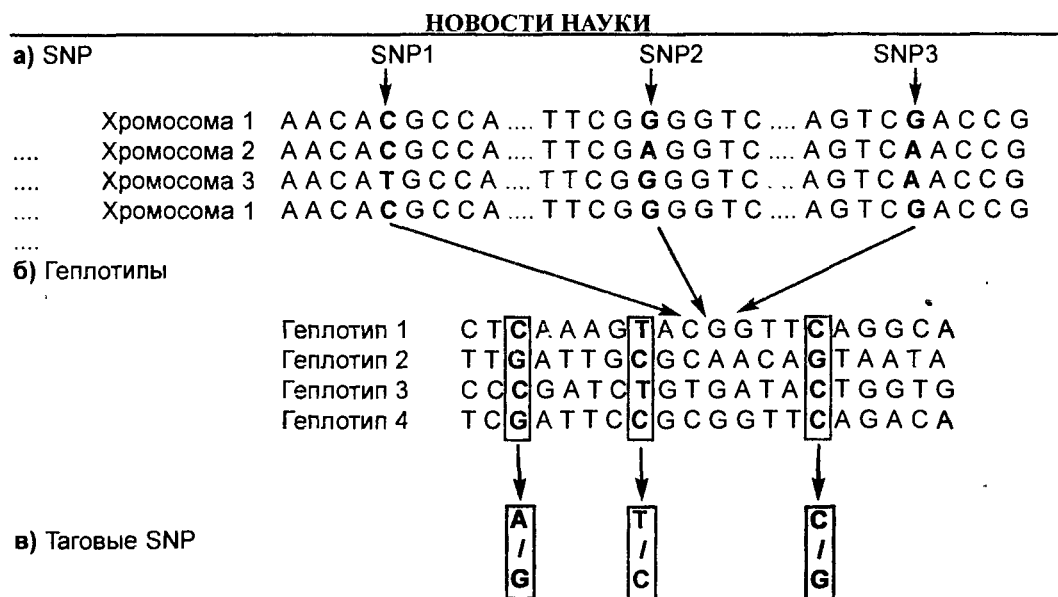


Рисунок 1.

- SNP, гаплотипы таговые SNP (приводятся по Nature (2003), 426, 789-796 с изменениями).
- а) SNP. Представлен короткий отрезок ДНК 4-х вариантов одного и того же хромосомного района у различных людей. Большая часть нуклеотидной последовательности ДНК в этих хромосомах идентична, но имеются 3 основания, где происходят вариации. Каждый SNP имеет 2 возможных аллели; 1-й SNP содержит аллели С и Т.
- б) Гаплотипы. Гаплотип состоит из конкретной комбинации аллелей из близлежащих SNP. Здесь показаны генотипы для 20 SNP расположенных на 6.000 оснований ДНК. Показаны только изменяющиеся основания, включая 3 SNP, которые находятся на панели а. Для этого района большинство хромосом при обследовании популяции имеют гаплотипы 1-4.
- в) Таговые SNP. Генотипирование этих 3-х из 20-ти SNP достаточно для идентификации 4-х гаплотипов. Например, если конкретная хромосома имеет паттерн А-Т-С на этих 3-х таговых SNP, то этот паттерн совпадает с паттерном определяемым для гаплотипа 1. Следует отметить, что многие хромосомы несут на себе общие для всей популяции гаплотипы.

Вместо этого, ассоциированные паттерны для эффективного выбора таговых SNP должны определяться эмпирически. На основе эмпирических исследований было установлено, что большая часть информации о генетических вариациях, представленной 10 млн обычных SNP, у населения может быть получена путем генотипирования 200 тыс. - 1 млн таговых SNP. Таким образом, существенное уменьшение работы по генотипированию может быть достигнуто с малой потерей информации - путем использования данных по LD, имеющихся в геноме. Для общих SNP, которые, как правило, старше чем редкие SNP, LD-паттерны в основном, отражают историческую рекомбинацию и демографические события. Некоторые рекомбинационные события происходят снова и снова в горячих точках. В результате, нынешние хромосомы представляют собой мозаики прародительских хромосомных районов. Это объясняет почему гаплотипы и LD-паттерны повторяются в очевидно несвязанных хромосомах внутри группы населения и обычно среди разных групп населения. Рассмотренные данные составляют концептуальное и эмпирическое основание для разработки гаплотипных карт генома человека, ГМ. Такая карта должна описывать обычные паттерны вариаций, включая ассоциации между SNP, а также включать таговые SNP, отобранные для наиболее эффективного и исчерпывающего охвата этой информации.

Создание консорциума

Первая встреча для обсуждения научных и этических проблем, связанных с разработкой гаплотипной карты генома человека, состоялась в Вашингтоне в июле 2001 г. на ней были сформированы группы для обсуждения этических проблем,

НОВОСТИ НАУКИ

для разработки научного плана и для выбора населения, подлежащего изучению. Международный ГМ проект был затем формально утвержден; встреча была назначена в Вашингтоне на октябрь 2002 г. Группы участников и источники финансирования представлены в таблице.

Таблица. Группы, принимающие участие в международном проекте HarMap

Страна	Научные учреждения	Роль	Процент генома	Хромосомы	Финансирующая организация
Япония	Университет Рикен, Токио	Генотипирование (Third Wave)	25,1%	5, 11, 14, 15, 16, 17, 19	MEXT
	Универ. Наук о здоровье, Хоккайдо Институт этики, Университет Шиншю	Общественные консультации, образцы			
Великобритания	Институт Сангера	Генотипирование (Illumina)	24,0%	1, 6, 10, 13, 20	Wellcome Trust
	Университет Оксфорда	Анализ			TSC, НИЗ, США
	Университет Оксфорда	Анализ			Wellcome Trust TSC, НИЗ, США
Канада	Университет МакГилла, Геномный инновационный центр, Квебек	Генотипирование: (Illumina)	10,0%	2, 4p	Геном, Канада Геном, Квебек
Китай	Пекинский Институт Геномики; Китайский Нац. центр генома человека, Пекин	Генотипирование (PerkinElmer, Sequenom)	4,8%	10% 3, 8p, 21	MOST, Китайский Университет, Фонд естественных наук, Комиссия по инновациям и технологиям, Гонконг Фонд по предоставлению университетских грантов, Гонконг
	Китайский Национальный Центр Генома Человека, Шанхай	Генотипирование (Illumina)	3,2%		
	Университет Гонконга, Гон-Конгский научно-технологический Университет, Китайский Университет, Гонконг	Генотипирование: (Sequenom)	2,0%		
	Пекинский Университет	Работа с населением			MOST
	Пекинский Институт Геномики	Образцы			
США	Illumina	Генотипирование: Illumina	15,5%	30,9%	НИЗ, США 8q, 9, 18q, 22, X 4q, 7q, 18p, Y 12 7p
	Институт Уайтхед	Генотипирование: (Sequenom)	9,1%		
	Медицинский колледж Бэйлор, ParAllel	Генотипирование: (ParAllel)	4,4%		
	UCSF, Вашингтонский Университет	Генотипирование: PerkinElmer	1,9%		
	Университет Джона Хопкинса	Анализ			
	Университет Юта	Работа с населением, образцы			Фонд В.М. Кека, Фонд Д.Д. Экле, НИЗ, США
Нигерия	Университет Говарда, Ибаданский Университет	Работа с населением, образцы			НИЗ, США
	Лаборатория Колдспринг Харбора, Нью-Йорк	Центр по координации данных			TSC

Образцы ДНК и человеческие популяции

Человеческие популяции являются продуктами многочисленных социальных, исторических и демографических процессов, поэтому нет оснований говорить о типичных, особых или строго отграниченных популяциях. Поскольку самые обычные вариационные паттерны могут быть обнаружены в любой популяции, ни одна из них не может рассматриваться как особая при включении в ГМ. Тем не менее, участники проекта решили ввести в ГМ несколько популяций из различных родовых географических мест проживания, с тем, чтобы в эту карту могло войти не только большинство общих вариаций, но и некоторые из ранее встречающихся общих вариантов, а также рассмотреть различные гипотезы, касающиеся LD-паттернов.

Изучение частотности распределений аллелей свидетельствует о том, что родовая география является вполне приемлемой основой для характеристики человеческих популяций. Предварительные исследования с использованием образцов, взятых у жителей народности Йоруба, а также у японцев, китайцев и у людей, происходящих из Северной и Западной Европы, показали наличие сходства в их гаплоидных паттернах, хотя по частотности гаплотипы порою различаются. Учитывая полученные данные, а также имея в виду проблемы этического, социального и культурного характера, участники проекта обратились к этим группам населения с просьбой о включении их в ГМ после соответствующих консультаций.

Гаплоидное картирование, разработанное на основе образцов, взятых у этих 4-х больших групп населения, будет включать значительное количество генетических вариаций, имеющихся у всех популяций, населяющих Землю. Поскольку проект ориентирован на медицину, то обычные вариационные паттерны, обнаруженные в ходе его реализации, могут оказаться полезными при идентификации генов, влияющих на развитие болезней и на лекарственные реакции во многих других популяциях. Образцы, взятые у некоторых других популяций, в настоящее время отбираются для исследований, которые должны показать, насколько гаплотипные паттерны в этих группах сходны с паттернами, включенными в ГМ. Если в этих паттернах будут обнаружены большие различия, то образцы, взятые у некоторых из этих популяций, могут быть генотипированы в широком масштабе, чтобы сделать ГМ более применимым к ним. Вполне возможно, что учеными разных стран будут предприняты дальнейшие мониторинговые исследования в других популяциях, малых и больших для обнаружения генов общих болезней.

Проект предусматривает изучение 270 образцов ДНК: 90 образцов, взятых у населения Шт. Юта, США - людей северо- и западно- европейского происхождения (образцы, собранные Центром по изучению человеческого полиморфизма - CEPH - и используемые для других генетических карт: 30 "троек", включающих двух родителей и взрослого ребенка, а также новые образцы, взятые у 90 людей народности Йоруба в Ибадане, Нигерия (30 "троек"), 45 родственно несвязанных японцев, живущих в Токио и 45 родственно несвязанных китайцев, живущих в Пекине. Все доноры дали согласие на их включение в проект. Состав популяций определялся в соответствии с культурными традициями каждой группы: у людей народности Йоруба спрашивали у донора все ли 4 деда были йоруба; у китайского донора спрашивали являются ли по, крайней мере, 3 из 4-х дедов китайцами; у японцев все сведения основывались на "самоидентификации". CEPH-образцы были получены от некоммерческого медицинского института Корнелла; клеточные линии и ДНК из новых образцов будут представлены для последующих исследований институтом Корнелла в начале 2004 г. вместе с протоколами, одобренными соответствующими комитетами по этике. Предполагается, что в будущем другие исследователи проведут генотипирование дополнительных SNP в этих образцах и что полученные данные будут способствовать постоянному совершенствованию ГМ.

Эти образцы будут снабжены анонимными идентификационными сведениями о популяционной и половой принадлежности. Поскольку целью проекта является исключительно идентификация генетических паттернов, не будет дано никакой медицинской и фенотипической информации. Количество собранных образцов на 50% превышает требуемое, что позволяет избежать включение в ГМ образца, отождествляемого с конкретным донором, который должен оставаться анонимным.

Образцов от 45 несвязанных между собой лиц будет достаточно, чтобы выявить 99% гаплотипов с частотой 5% и более в данной популяции. При изучении LD можно использовать взятые "наугад" индивидуальные образцы, а также "тройки" или более обширные группы родственников. Каждый подход имеет свои преимущества (легкость взятия образца) и свои недостатки: уменьшение эффективности при увеличении количества связанных родственными узлами лиц. Анализ существующих данных и компьютерное моделирование свидетельствуют о том, что несвязанные индивиды и тройки обладают существенными преимуществами, когда дело касается оценки локальных LD-паттернов. Тройки дают полезную информацию о точности генотипированных платформ, используемых в проекте.

Выбор SNP

SNP должны иметь высокую плотность, необходимую для адекватного описания генетических вариаций по всему геному. В самом начале проекта средняя плотность маркеров в общественной базе данных dbSNP составляла 1 SNP на 1000 пар оснований, т.е. приблизительно 2,8 млн SNP, но, учитывая неравномерность их распределения, многие районы имели более низкую плотность SNP.

Последующие SNP были получены при случайном (выборочном) секвенировании из целльно-геномных или целльно-хромосомных библиотек с использованием методов, разработанных для первоначальной карты SNP человека. Полезным результатом этого поиска было подтверждение SNP, полученных ранее. SNP, для которых каждая аллель наблюдалась независимо в 2-х или более пробах (SNP "двойного попадания"), имеют более высокую среднюю частоту минорной аллели, чем SNP "единичного попадания". Это позволяет сократить расходы при разработке анализа. По состоянию на 4-ое ноября 2003 г. в базе dbSNP количество SNP (с уникальной геномной позицией) составляло 5 - 7 млн, а количество SNP двойного попадания превышало 2 млн. К февралю 2004 г. ожидается пополнение базы данных SNP до 6,8 млн, включая 2,7 млн SNP двойного попадания.

Так как уровень вариаций LD и гаплотипов в геноме составляет два порядка, была принята иерархическая стратегия генотипирования. В начальном периоде генотипирования в проекте предполагается успешно проводить генотипирование 600.000 SNP, расположенных в интервале 5 тысяч пар оснований, причем каждый из SNP будет иметь минимальную частоту встречаемости аллеля равную, по меньшей мере, 5%. Приоритет отдан ранее подтвержденным SNP двойного попадания, а также SNP, вызывающим изменения в аминокислотной последовательности. К середине 2004 г. планируется получить данные по генотипированию, которые будут проанализированы на предмет ассоциации между соседними SNP. Дополнительные SNP будут затем генотипированы в тех же образцах ДНК при более высокой плотности там, где ассоциации проявляются слабо. Последующие стадии анализа и генотипирования будут проведены согласно существующим требованиям. Ожидается, что будет генотипировано более 1 млн SNP. Иерархическая стратегия позволит районам генома с наименьшими LD быть охарактеризованными при плотности до 1 SNP на 1000 оснований, что даст возможность максимизировать характеристику районов с ассоциациями, распространяющимися только на очень короткие дистанции.

Генотипирование

Каждый генотирующий центр отвечает за генотипирование всех образцов для всех отобранных СНИПов в отобранных хромосомных районах (см. табл.). В центрах используется всего пять наиболее отработанных методов генотипирования, что даст возможность сравнить их точность, степень успеха, выработку и цену. Доступ к нескольким платформам является преимуществом проекта: анализ SNP, не удавшийся на 1-ой платформе, может быть успешно проведен с использованием другого метода (для заполнения пробела в ГМ). Все платформы будут оценены на основе общих критериев успешности - чтобы убедиться в одинаково высоком стандарте поставляемых для проекта данных.

Качество генотипирования оценивается 3-мя путями. Вначале, при открытии проекта, всем центрам был дан один и тот же случайным образом выбранный набор из 1.500 SNP для проведения анализа и генотипирования в 90 CEPH образцах ДНК, используемых в проекте. Генотирующие центры выдали данные с полнотой 99,2% и с точностью 99,5% (при сравнении согласующихся данных по крайней мере, с 2-мя другими платформами). Во-вторых, каждый из экспериментов по генотипированию включает образцы для внутренних проверок качества; при этом каждый планшет, состоящий из 96 лунок, содержал дубликаты 5 разных образцов и 1 бланк. Вдобавок, данные, полученные от троек, обеспечивают проверку на постоянное по Менделю наследование аллелей SNP. Для всех популяций данные полученные от несвязанных образцов, удостоверяют, что все SNP находятся в равновесии Hardy-Weinberg (тест на генетически сочетающиеся паттерны). Хотя небольшое количество SNP может не пройти эту проверку по биологическим причинам, более типичным является непрохождение проверки в тех случаях, когда генотирующая платформа делает постоянные ошибки такие например, как "недовыявление" гетерозигот. В-третьих, образцы SNP-генотипов, представленные каждым центром, будут выборочно отбираться для регенотипирования другими центрами. Эти точные и строгие трехсторонние оценки качества должны обеспечить полноту и надежность данных, полученных при выполнении проекта.

Передача данных

Проект рассчитан на быструю и полную передачу данных научной общественности и свободный (бесплатный) доступ к ним.

Все данные по новым SNP, условиям анализа, аллельным и генотипным частотам будут быстро передаваться всем заинтересованным лицам по Интернету и концентрироваться в ГМ - Центре по Координации Данных (ЦКД); они будут включены в dbSNP. Данные по отдельным генотипам и гаплотипам первоначально станут доступны в ЦКД на основе краткосрочного (click-what) лицензионного соглашения. Это сделано с той целью, чтобы данные проекта не были включены в ограничительные патенты, и чтобы они оставались доступными в будущем.

Единственное условие при получении доступа к данным состоит в том, что пользователи не должны ограничивать доступ к данным других лиц и делиться данными только с теми, кто соглашается получать их на таких условиях. Только после того как будут определены гаплотипы в данном районе последует передача индивидуальных генотипов, гаплотипов и таговых SNP этого района в db SNP, где не существует никаких лицензионных ограничений. Участники проекта пришли к согласию, что их собственные лаборатории получат доступ к данным через ЦКД и по "click - what" - лицензии, что обеспечит равный доступ для всех исследователей.

Как полагают участники консорциума, данные по SNP, генотипам и гаплотипам при отсутствии конкретной (специфической) полезности не несут патентоохранной информации. Конкретная польза будет заключаться, например, в установлении ассоциации SNP или гаплотипа с важным с медицинской точки зрения фенотипом - таким как риск по какому-либо заболеванию или реакция на лекарство. Проект не включает в себя какие-либо исследования по фенотипной

(фенотипической) ассоциации. Однако, проводимая участниками проекта политика в области получения данных не запрещает пользователям получать сведения, представляющие соответствующую интеллектуальную собственность, по таким ассоциациям, до тех пор, пока любой вышедший патент не будет использован для ограничения доступа других исследователей к работе с данными, полученными в ходе реализации ГМ-проекта.

Анализ данных

В проекте предусматривается применение существующих и новых методов анализа и демонстрации данных. LD между парами маркеров будет рассчитываться, используя стандартные методы, такие как D' , r^2 и др. Апробируются различные методы для районов с высоким содержанием LD и гаплотипов на хромосоме. Среди существующих методов используются такие, как LD-профили "скользящего окна", LD-карты, гаплотипные блоки и оценка скоростей мейозных рекомбинаций на хромосоме. После анализа LD на 1-ой стадии проекта, районы, содержащие небольшие количества LD или вообще не содержащие LD, будут идентифицированы и ранжированы для дальнейшей селекции SNP и генотипирования. Будут разработаны и оценены методы отбора оптимальных коллекций таговых SNP. В проекте, таким образом, будут даны оценки полученных данных и изложены взгляды на таговых SNP, которые могут оказаться полезными для научного сообщества. Поскольку все данные и все аналитические методы будут доступны для исследователей, другие ученые смогут также проанализировать полученные данные и улучшить аналитические методы.

Для оптимального отбора SNP и анализа LD и гаплотипов проводится пилотное исследование, целью которого является определение плотного набора генотипов в больших геномных районах. Десять районов генома (500 тысяч пар оснований в каждом) планируется секвенировать в 48-ми несвязанных образцах ДНК из ГМ (16 CEPH - в настоящее время они секвенируются): 16 Йоруба, 8 японцев и 8 китайцев. Все идентифицированные SNP в государственных базах данных будут генотипированы во всех 270-ти ГМ - ДНК образцах, и генотипированные данные будут переданы для изучения, согласно вышеописанным принципам. Это исследование предоставит в распоряжение исследователей данные по плотным генотипам для разработки методов отбора SNP и для оценки полноты полученной информации, что позволит направленно проводить последующие стадии генотипирования.

Когда ГМ будет использован для обследования больших геномных районов, возникнет проблема множественных сравнений, т.к. придется протестировать десятки тысяч и даже сотни тысяч SNP и гаплотипов на их связь с болезнями. Здесь ожидаются трудности отделения истинных результатов от ложноположительных. Таким образом, новые статистические методы, репликационные исследования и функциональные анализы вариантов окажутся чрезвычайно важными для подтверждения новых данных и для идентификации функционально-значимых SNP.

Итак, целью Международного ГМ-проекта, начатого более года назад, является разработка научной методики, которая позволит ученым всего мира обнаруживать генетические факторы, влияющие на восприимчивость к заболеваниям, защитные механизмы и на реакцию организма на лекарственные средства. Участники проекта верят, что ГМ обеспечит новый стимул к проведению исследований по генам-кандидатам, сцепленным и ассоциативным вариантам всего генома. По своему масштабу и потенциальным последствиям международный ГМ проект сопоставим с проектом "Геном человека". Оба проекта амбициозны в научном плане и сложны в технологическом исполнении. Основываясь на тесном международном сотрудничестве, тот и другой предполагают быструю передачу данных в распоряжение научной общественности. Оба проекта призваны существенно углубить наше понимание

биологии человека и его здоровья. Однако, если секвенирование охватывает весь геном, включая 99,9% генома, в пределах которого все люди одинаковы, ГМ должен охарактеризовать общие паттерны в пределах 0,1%, где мы отличаемся друг от друга.

Для полной реализации ГМ потенциала должно произойти несколько событий. Технология генотипирования должна стать более рентабельной, а методы анализа нуждаются в улучшении. Пилотные исследования, проводимые с другими популяциями, необходимо завершить, чтобы подтвердить повсеместную применимость ГМ, учитывая возможность его расширения, с тем, чтобы все основные мировые популяции могли извлечь из него наибольшую пользу. Для использования технологий, созданных при реализации ГМ, последующие проекты должны провести тщательное фенотипирование групп больных и здоровых людей для многих социально значимых заболеваний. Массовые исследования сотен тысяч людей также будут неоценимы в плане оценки генетических и природных причин возникновения заболевания.

Участники проекта уделяют самое пристальное внимание этическим проблемам, возникающим при выполнении ГМ-проекта. Консультации с представителями донорных популяций по проблемам получения согласия и другим деликатным вопросам, возникающим перед сбором образцов, позволили существенно улучшить этические стандарты, соблюдаемые при проведении международных исследований в области популяционной генетики. В будущих проектах в области популяционной генетики эти подходы будут непрерывно совершенствоваться.

Ученые, использующие ГМ-методы, должны представлять свои открытия таким образом, чтобы избежать навешивания ярлыков на отдельные популяции в духе генетического детерминизма, или придавать слишком большое биологическое значение таким, в основном, социальным конструкциям, как расы.

ГМ имеет большие перспективы, являясь новым мощным орудием исследования, углубляющим наше понимание генетических факторов, действующих в норме и при патологических состояниях. Осуществление всех полезных возможностей этого проекта потребует тесного сотрудничества ученых, работающих в области фундаментальных исследований, а также популяционных генетиков, эпидемиологов, клиницистов, исследователей в области социальных и этических проблем всего мирового сообщества.