

Применение метода секвенирования экзонов для диагностики предрасположенности к комплексным заболеваниям в профилактической медицине

Щербакова Н. В., Мешков А. Н., Бойцов С. А.

ФГБУ “Государственный научно-исследовательский центр профилактической медицины” Минздрава России. Москва, Россия

Развитие профилактической медицины в XXI веке, по-видимому, невозможно без использования индивидуальных генетических данных пациента. Генодиагностика в настоящий момент служит золотым стандартом в диагностике моногенных заболеваний. В последние годы активно изучаются генетические факторы, влияющие на развитие комплексных (мультифакториальных) заболеваний, таких как ишемическая болезнь сердца, гипертоническая болезнь и т.д. Данный обзор посвящен анализу возможности выявления новых генетических факторов наследственности ком-

плексных заболеваний, полученных при секвенировании экзонов пациентов.

Ключевые слова: профилактическая медицина, генетическая диагностика, секвенирование экзона, комплексные или мультифакториальные заболевания.

Кардиоваскулярная терапия и профилактика, 2013; 12 (6): 24–28

Поступила 11/11–2013

Принята к публикации 14/11–2013

Exome sequencing and the diagnostics of complex disease predisposition in preventive medicine

Shcherbakova N. V., Meshkov A. N., Boytsov S. A.

State Research Centre for Preventive Medicine. Moscow, Russia

The further development of preventive medicine in the 21st century may be impossible without the assessment of individual patients' genetic data. At present, genetic methods are the gold standard in the diagnostics of monogenic diseases. Recently, genetic factors linked to complex (multifactorial) diseases, such as coronary heart disease and arterial hypertension, have been actively explored. This review focuses

on the possible identification of new genetic factors of complex disease heritability, using the exome sequencing approach.

Key words: preventive medicine, genetic diagnostics, exome sequencing, complex or multifactorial diseases.

Cardiovascular Therapy and Prevention, 2013; 12 (6): 24–28

Введение

Развитие профилактической медицины в XXI веке невозможно без использования индивидуальных данных о генетике пациента. В пользу этого соображения свидетельствует, с одной стороны, ограниченность современных систем стратификации риска в предсказании вероятности развития мультифакториальных заболеваний, а, с другой стороны, прогресс в области генетики человека в последние два десятилетия. Прогресс генетики связан, во-первых, с развитием технологий и снижением стоимости секвенирования и, во-вторых, с расширением возможностей биоинформационной обработки данных. Изучение наследственной составляющей заболеваний человека основывается на упрощенном разделении болезней на простые, т.е. моногенные, редко встречающиеся в популяции (с менделевским типом наследования), и мультифакториальные, обусловленные влиянием как факторов внешней среды, так целого ряда генетических факторов (с неменделевским типом наследования). Эти заболевания очень широко распространены в популяции [1]. Генодиагностика в настоящий момент является золо-

тым стандартом в диагностике моногенных заболеваний. В последние годы активно изучаются генетические факторы, влияющие на развитие мультифакториальных заболеваний, таких как ишемическая болезнь сердца (ИБС), гипертоническая болезнь (ГБ) и др. С развитием технологий секвенирования стало возможным быстрое прочтение геномов конкретных пациентов. Одновременно наметилась тенденция к секвенированию областей генома, содержащих только кодирующие участки генов, — экзоны (экзом-совокупность всех экзонов у одного организма). Такой подход еще более удешевляет исследования и позволяет упростить биоинформационный анализ. Данный обзор посвящен анализу возможности выявления новых генетических факторов наследственности мультифакториальных заболеваний, полученных при секвенировании экзонов пациентов.

Проблема “потерянной наследственности”

Для большинства мультифакториальных заболеваний вклад наследственности в риск их развития колеблется от 30% до 80%. Первые доказательства значимости роли генетических факторов в развитии мультифак-

*Автор, ответственный за переписку (Corresponding author):

Тел.: 8 (901) 512–12–71

e-mail: meshkov@lipidclinic.ru

[Щербакова Н. В. — лаборант исследователь лаборатории молекулярной генетики, Мешков А. Н.* — руководитель лаборатории, Бойцов С. А. — директор, руководитель отдела клинической кардиологии и молекулярной генетики].

ториальных заболеваний основывались на демонстрации семейной агрегации и оценке наследственности и стимулировали исследования по поиску аллелей риска [2, 3]. Исследования аллельных ассоциаций в генах-кандидатах включали в себя большое количество вариантов последовательности дезоксирибонуклеиновой кислоты (ДНК), которые подозревались как связанные с риском развития заболевания. Однако результаты были противоречивыми и стали объектом большого числа ложных ассоциаций. Появление исследований геномных ассоциаций (GWAS — genome-wide association study) возвратило надежду на то, что объективный и современный метод выявит генетические детерминанты предрасположенности к мультифакториальным заболеваниям. GWAS стал методом генетических исследований сложных заболеваний и применялся для выявления аллелей риска практически для всех мультифакториальных заболеваний [4].

GWAS фокусировались на известных и распространенных однонуклеотидных заменах (SNP — single nucleotide polymorphism) с минимальной аллельной частотой $> 5\%$. За последнее десятилетие данные исследования выявили сотни распространенных аллелей, увеличивающих риск развития частых заболеваний [5–10]. Эти исследования стали возможны, благодаря наличию больших, хорошо обозначенных групп пациентов, усовершенствованию технологий генотипирования и методов анализа полученных данных [11]. Исследования геномных ассоциаций пролили свет на связь генов с возрастной макулодистрофией [12], болезнью Крона [13] и других заболеваний. Однако в большинстве случаев интерпретация полученных данных затруднительна, т.к. описанные генетические варианты имеют очень слабые последствия на функцию белков. Связано это с тем, что ассоциированные SNP в основном расположены в некодирующей части ДНК (40%) или в интронах и не сцеплены с несинонимичными заменами, которые могли бы повлиять на функцию геномного продукта [11]. 9.11.2012г Национальный Институт исследования человеческого генома опубликовал список 1416 исследований, выявивших 7688 SNP, ассоциированных с одним из 900 различных фенотипов, включающих в себя заболевания, физиологические показатели, ответы на лекарственную терапию и поведенческие особенности [14]. Было выделено >500 независимых, сильно ассоциированных с заболеваниями SNP с $p < 1 \cdot 10^{-8}$. Обнаруженные в результате проведения GWAS ассоциации позволили выявить гены, которые раньше не рассматривались как гены-кандидаты для тех или иных заболеваний, что позволило по-новому взглянуть на патологический механизм, лежащий в основе этих заболеваний. В то время как идентификация такого количества значимых вариантов была несомненным триумфом GWAS, опубликованные варианты имели малое значение для предсказания индивидуального риска развития конкретного заболевания, или для применения их в клинической практике [14]. Для подавляющего большинства выявленных с помощью GWAS вариантов, размер эффекта был достаточно мал, с отношением шансов не превышающим 1,1–1,4. Несмотря на то, что в развитие большинства распространенных заболеваний, которые рассматривались в исследованиях геномных ассоциаций, достаточный вклад вносит наследственность, удивительно малая часть этой наследственности была объяснена полученными результатами,

возможно, в т.ч. и потому, что с помощью GWAS невозможно оценить участие в наследственности ранее неизвестных SNP, редких мутаций, tandemных повторов и структурных изменений [15].

Когда исследования геномных ассоциаций начинались, в научном сообществе доминировала гипотеза об участии распространенных в популяции генетических вариантов в развитии частых комплексных заболеваний. В настоящий момент эта модель опровергается в свете так называемой проблемы “потерянной наследственности” — почти все, за несколькими исключениями, локусы, выявленные в исследованиях геномных ассоциаций, объясняют лишь малую часть наследственности [16]. Несколько десятков локусов со средним эффектом и средней частотой каждый не могут объяснить имеющиеся проценты риска заболевания в популяции.

Современные модели наследственности комплексных заболеваний

Аллели, часто встречающиеся в популяции, как показали исследования GWAS, объясняют малую часть наследственности сложных фенотипов. Технологический прорыв в полногеномном и экзомном секвенировании вызвал широкий интерес научного сообщества к попыткам объяснения феномена “потерянной наследственности” [4]. При этом было предложено несколько объяснений проблемы “потерянной наследственности” и способов решений по ее обнаружению.

Одна из популярных теорий предполагает, что, т.к. наличие заболевания обусловлено взаимодействием очень многих различных факторов (генетических, эпигенетических, внешней среды), влияние их всех невозможно учесть в универсальных диагностических моделях. При этом предполагается, что промежуточные фенотипы, связанные с заболеванием, станут более легкими для исследования объектами геномных ассоциаций, чем заболевания сами по себе.

Согласно другой гипотезе предполагается, что варианты, вносящие вклад в распространенные заболевания, имеют небольшой эффект и контекст-зависимы, т.е. специфические комбинации распространенных генетических вариантов вкупе с факторами внешней среды могут привести к более существенному увеличению эффекта, чем те же эффекты, измеряемые для каждого из генетических вариантов по отдельности. Несмотря на то, что эта теория вполне вероятна, существует лишь небольшое число очевидных примеров распространенных генетических вариантов, выявленных в GWAS, для которых роль взаимодействия между собой или с факторами окружающей среды была доказана в развитии заболеваний. При этом необходимо отметить, что статистические методы, способные описать взаимодействие между распространенными вариантами генетических полиморфизмов в человеческом геноме, развиты слабо [17]. Однако можно ожидать, что новые исследования выявят новые профили риска, которые будут объяснены множественным взаимным влиянием [18].

Третья версия предполагает, что большая часть наследственности для некоторых распространенных заболеваний определяется редкими вариантами, с частотой аллелей $< 1\%$, но обладающими относительно высокой пенетрантностью. Согласно этой модели, экспрессия генов может быть модифицирована мутацией, влиянием друго-

го локуса или факторов окружающей среды [16], но главное заключается в том, что редкое изменение экспрессии разных генов приводит к формированию сходных фенотипов.

Таким образом, малую на настоящий момент роль влияния генетических факторов в развитии распространенных в популяции заболеваний, объясняют тремя следующими причинами:

- взаимодействие генотипических, эпигенетических и факторов внешней среды (модель наследственности “в широком смысле”) [19, 20];
- большое количество распространенных генетических вариантов с малым эффектом на всем аллельном частотном спектре (модель бесконечно малых величин) [21];
- большое количество редких генетических вариантов с большим эффектом (модель редких аллелей) [22].

Модель наследственности “в широком смысле”

Модель наследственности “в широком смысле” предполагает, что аддитивный эффект широко распространенных вариантов и редких аллелей с большим эффектом не объясняет всей генетической детерминанты многих заболеваний. Предпосылками к этой теории явились множественные наблюдения взаимодействия гена и ген-окружающая среда в модельных организмах качественных генетических исследований [23, 24], а также исследования эпигенетических эффектов [25, 26]. Основная идея заключается в том, что исследования геномных ассоциаций измеряют лишь средний эффект аллели во всей популяции, не имея возможности оценить истинный размер эффекта на индивидуальном и семейном уровнях [16].

Модель бесконечно малых величин

Модель бесконечно малых величин предполагает, что варианты, часто встречающиеся в популяции, объясняют всю наследственную предрасположенность к сложным заболеваниям и количественным признакам, где каждый из сотен и тысяч локусов вносит свой вклад в развитие заболевания. Пациенты с заболеванием будут иметь несколько большее количество аллелей, связанных с риском развития болезни, чем пациенты, не страдающие рассматриваемым заболеванием. Согласно данной модели, в конечном счете, каждый ген будет вносить определенный вклад в любое мультифакториальное заболевание, но с таким бесконечно малым размером эффекта, что доступные в настоящее время объемы популяционных исследований не могут его выявить [16]. С помощью мета-анализов результатов исследований геномных ассоциаций для количественно описываемых признаков (индекс массы тела и рост), каждое из которых включало до нескольких сотен тысяч человек, была показана малая вероятность того, что для каждого из показателей будет возможным подтвердить ассоциацию с более чем сотней локусов, которые даже в сумме не объяснят и половины факта наследственности [27, 28]. Название этой модели заимствовано из начальной формулировки количественной генетической теории Фишера [16]. Это означает, что наследственность не столько потеряна, сколько скрыта за пределами статистической значимости, используемыми для определения аллелей риска с высокой достоверностью [29].

Модель редких аллелей

Хорошо известно, что редкие, практически полностью пенетрантные генетические варианты приводят к развитию моногенных заболеваний с менделевским типом наследования. В связи с выраженностью влияния подобных мутаций на фенотип, отрицательным характером влияния в естественном отборе, аллельная частота таких генетических вариантов в человеческой популяции никогда не возрастала до значимого уровня, при этом доминантные варианты возникали как мутации *de novo*, а рецессивные аллели сохранялись с минимальной частотой. Для широко распространенных в популяции заболеваний, подобные аллельные варианты не являются отрицательным фактором в естественном отборе, т.к. значимый эффект на здоровье человека оказывается во второй половине жизни (например, сахарный диабет, сердечно-сосудистые заболевания, болезнь Альцгеймера) [14]. Данное предположение было протестировано, и создана модель вариабельности степени предрасположенности к заболеванию с учетом генетического дрейфа нейтрального или слабого отрицательного характера влияния на отбор и достоверным разбросом мутационного процесса [22]. В мультилокусной модели, симулирующей многофакторное заболевание, было обнаружено, что наибольшей вариабельностью в проявлении фенотипа обладают те локусы, которые подвержены выраженному мутационному процессу и слабому отрицательному отбору, что приводит к сохранению этих аллельных вариантов с малой или средней частотой в популяции. Высокий уровень мутационного процесса принципиален для формирования множественных независимых генетических вариантов в одном гене, приводящих к вариабельности фенотипических проявлений. Таким образом, даже если частота аллельных вариантов в одном гене, связанном с развитием заболевания, достаточно велика, то аллельная частота каждого отдельно взятого генетического варианта будет слишком мала для обнаружения в исследованиях подобных GWAS [14]. Также при анализе результатов исследований геномных ассоциаций в отношении болезней, распространенных в популяции, было показано, что большее отношение шансов имели более редкие мутации, чем генетические варианты с большей аллельной частотой [30]. Таким образом, стало возможным предположить, что большая часть генетической основы заболеваний, распространенных в популяции, обусловлена редкими и преимущественно функционально значимыми генетическими вариантами, вносящими большой вклад в риск развития заболевания у конкретных пациентов. Очевидно, что с наибольшей величиной эффекта будут те варианты, которые несут очевидные функциональные последствия [18]. Это вовсе не означает, что частые варианты не играют роли в развитии распространенных заболеваний. Наоборот, изучение некоторых болезней с менделевским типом наследования показывает, что частые варианты могут играть ключевую роль в модификации эффекта более редких и пенетрантных генов, влияющих на степень риска заболевания. Разумно предположить, что это также может оказаться правдоподобным для заболеваний, распространенных в популяции [18].

Как бы то ни было, особенности подобных ассоциаций подробно не изучались до последних работ Dickson SP, et al. [31]. Они предложили теоретическую модель, в которой один или несколько редких генетических вари-

антов были единственными причинами сложных заболеваний. Далее авторов заинтересовал вопрос, насколько часто сигнал риска, обусловленный такими вариантами, будет выявлен и соотнесен с частыми вариантами в исследованиях геномных ассоциаций. Например, каузальные варианты могут быть отделены миллионами нуклеотидов в геноме от часто встречаемых генетических вариантов, которые воспринимаются как сигнальные для риска развития заболеваний, и реальный эффект обусловленного ими риска может быть в несколько раз выше, чем тот, который приписывается частым вариантам. К тому же, вероятность определения значимых геном-ассоциированных сигналов, приписываемых частым вариантам, может возрасти с увеличением числа редких каузальных вариантов в исследуемом регионе (т.к. множественные редкие варианты могут “нагружаться” в один гаплотип, определенный общим частым вариантом). Подобным ассоциациям было дано название “синтетические ассоциации”, подчеркивая их значение как особый вид непрямых связей.

Часто, когда ассоциированный SNP находится далеко от ближайшего гена, предполагается, что задействованы регуляторные участки генома. Так, например, SNP в 8q24, который ассоциирован с колоректальным раком [32], находится на расстоянии более сотни тысяч нуклеотидов от онкогена MYC. Но, несмотря на то, что есть данные о регуляторной функции этого региона, достоверной связи между регуляциями и обнаруженными генетическими вариантами так и не было продемонстрировано [33].

Для изучения связи редких аллелей с мультифакториальными заболеваниями на основе методов полногеномного или экзомного секвенирования применяются два типа дизайна исследований. Первый основан на анализе сегрегации заболевания в семьях, а второй — на изучении экстремальных фенотипов. Первый наиболее оптимально подходит для работы с заболеваниями с менделевским типом наследования. Второй дизайн больше подходит для работы с признаками, представленными в популяции с различной степенью пенетрантности, в картине мультифакториальных заболеваний. Если заболевания, распространенные в популяции, наследуются по похожим принципам, что и заболевания с менделевским принципом наследования, то стоит ожидать, что генетические варианты должны быть представлены миссенс-мутациями, мутациями с изменением стоп-кодонов или очевидно функционально значимыми макромутациями. При сравнении частоты различных классов генетических вариантов в десяти случаях контроля и десяти случаях заболевания было выявлено 383913 варианта (SNP и макромутации), присутствовавших в двух случаях заболевания и отсутствовавших в случаях контроля. Исследования такого большого количества вариантов потребовало бы значения $p=1,3 \cdot 10^{-7}$ для того, чтобы говорить о достоверно значимой ассоциации. Однако если исследование ограничить только вариантами, содержащимися в кодирующей части, это число упадет до 2354, что потребует значения $p=2,2 \cdot 10^{-5}$. Если же среди оставшихся вариантов выбрать только те, которые оказывают значимое влияние на функцию белка, требуемое значение p падает до $3,3 \cdot 10^{-4}$ [18].

При использовании дизайна экстремальных фенотипов для секвенирования отбираются пациенты с двух

концов распределения фенотипа. Так, например, проект The National Heart, Lung, and Blood Institute (NHLBI) — GO Exome Sequencing Project (ESP), выполняемый совместно “Broad Institute” и рядом университетских центров США, предполагает секвенирование экзонов >7 тыс. пациентов для выявления генов, лежащих в основе распространенных сердечно-сосудистых заболеваний (ССЗ): ИБС, артериальная гипертензия (АГ), инсульты (МИ) и заболеваний легких: бронхиальная астма и хроническая обструктивная болезнь легких (ХОБЛ). В связи с тем, что частоты аллелей, связанных с заболеванием, обогащены на концах распределения фенотипа, секвенирование даже небольшого числа пациентов может потенциально выявить кандидатные аллели. Гораздо реже используется дизайн исследований случай-контроль, хотя дешевизна полноэкзомного секвенирования и все возрастающий публичный доступ к экзомам контрольных выборок, должны увеличить популярность исследований этого типа [1]. Несмотря на то, что зачастую требуется большой размер выборки для выявления редких вариантов с малым эффектом, существует много примеров вариантов с достаточно большим эффектом в отдельных популяциях. Например, пациенты гомозиготные по делеции C—C в рецепторе хемокина-5 дельта 32, которые составляют 1% от европейской популяции, имеют иммунитет к ВИЧ [34]. Такой протективный вариант привел к тому, что в популяции с высоким риском заражения ВИЧ, здоровые пациенты, гомозиготные по данной мутации, встречаются в 20 раз чаще, чем в других популяциях [13].

Одной из основных трудностей полноэкзомного секвенирования является невозможность точного определения объема экзонов генов, которые формируют экзом. Сохраняется достаточная неясность в определении тех последовательностей человеческого генома, которые действительно кодируют белки. Коммерческие киты теперь покрывают как минимум все собрание из базы RefSeq (The Reference Sequence database) и все возрастающее число гипотетических белков [1]. Другой сложностью в использовании полноэкзомного секвенирования для определения новых патогенных генов как для менделевских, так и для мультифакториальных заболеваний, остается трудность в выявлении аллелей, связанных с заболеванием среди множества непатогенных полиморфизмов и ошибок секвенирования. В среднем полноэкзомное секвенирование выявляет порядка 24 тыс. вариантов SNP у афроамериканцев и порядка 20 тыс. у американцев европейского происхождения. Из этих вариантов 95% уже определены как полиморфизмы для человеческой популяции. Стратегии определения каузальных аллелей среди остального множества могут различаться в зависимости от множества факторов (типа наследования, структуры популяции и др.). Эти же факторы определяют объем необходимой выборки для достижения адекватной мощности исследования и выбора наиболее удачной схемы анализа [1].

В заключение можно сказать, что для ответа на вопрос относительно участия генетических факторов в развитии заболеваний и объяснении феномена “потерянной наследственности” следует говорить не о том, какая из теорий наиболее правильна, а о том как эти теории между собой сочетаются.

Работа поддержана за счет гранта РФФИ № 13–04–01259.

Литература

1. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet.* 2011; 12 (11): 745–55.
2. Marenberg ME, Risch N, Berkman LF, et al. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med.* 1994; 330 (15): 1041–6.
3. Mayer B, Erdmann J and Schunkert H. Genetics and heritability of coronary artery disease and myocardial infarction. *Clin Res Cardiol.* 2007; 96 (1): 1–7.
4. Marian AJ. The enigma of genetics etiology of atherosclerosis in the post-GWAS era. *Curr Atheroscler Rep.* 2012; 14 (4): 295–9.
5. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012; 44 (9): 981–90.
6. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011; 43 (4): 333–8.
7. The International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005; 437 (7063): 1299–320.
8. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467 (7319): 1061–73.
9. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010; 466 (7307): 707–13.
10. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science.* 2008; 322 (5903): 881–8.
11. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum Mol Genet.* 2012; 21 (R1): R1–9.
12. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005; 308 (5720): 385–9.
13. Huang Y, Paxton WA, Wolinsky SM, et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med.* 1996; 2 (11): 1240–3.
14. Wagner MJ. Rare-variant genome-wide association studies: a new frontier in genetic analysis of complex traits. *Pharmacogenomics.* 2013; 14 (4): 413–24.
15. Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008; 456 (7218): 18–21.
16. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet.* 2011; 13 (2): 135–45.
17. Bhattacharjee S, Wang Z, Ciampa J, et al. Using principal components of genetic variation for robust and powerful detection of gene-gene interactions in case-control and case-only studies. *Am J Hum Genet.* 2010; 86 (3): 331–42.
18. Cirulli ET and Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet.* 2010; 11 (6): 415–25.
19. Feldman MW and Lewontin RC. The heritability hang-up. *Science.* 1975; 190 (4220): 1163–8.
20. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010; 11 (6): 446–50.
21. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. *Nat Rev Genet.* 2008; 9 (4): 255–66.
22. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet.* 2001; 69 (1): 124–37.
23. Mackay TF. The genetic architecture of quantitative traits. *Annu Rev Genet.* 2001; 35: 303–39.
24. Mackay TF, Stone EA and Ayroles JF. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet.* 2009; 10 (8): 565–77.
25. Jablonka E, Raz G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol.* 2009; 84 (2): 131–76.
26. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* 2007; 447 (7143): 433–40.
27. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010; 42 (11): 937–48.
28. Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet.* 2010; 42 (7): 570–5.
29. Gibson G. Hints of hidden heritability in GWAS. *Nat Genet.* 2010; 42 (7): 558–60.
30. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet.* 2008; 40 (6): 695–701.
31. Dickson SP, Wang K, Krantz J, et al. Rare variants create synthetic genome-wide associations. *PLoS Biol.* 2010; 8 (1): e1000294.
32. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet.* 2007; 39 (8): 984–8.
33. Prokunina-Olsson L, Hall JL. No effect of cancer-associated SNP rs6983267 in the 8q24 region on co-expression of MYC and TCF7L2 in normal colon tissue. *Mol Cancer.* 2009; 8: 96.
34. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CCR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science.* 1996; 273 (5283): 1856–62.