

Г.Г. Кутелев, А.Б. Криворучко, А.Е. Трандина,  
А.М. Иванов, Д.В. Черкашин,  
А.А. Марченко, С.Л. Гришаев

## Новые подходы к отбору генетических маркеров, ассоциированных с многофакторными фенотипическими признаками

Военно-медицинская академия им. С.М. Кирова, Санкт-Петербург

**Резюме.** *Анализируются современные подходы к поиску ассоциаций между исследуемым фенотипом и структурными вариациями генома человека. Большинство сложных фенотипических признаков, в том числе и заболеваний, не подчиняются закономерностям менделевского наследования, а имеют многофакторную природу, то есть существенный вклад в их развитие вносит генетическая составляющая в сочетании с влиянием факторов внешней среды. В целом существует несколько подходов к дизайну ограниченного набора полиморфных маркеров для точечного генотипирования. Селекцию отдельных молекулярно-генетических маркеров осуществляют на основе либо их статистически значимой ассоциации с изучаемым многофакторным признаком, либо функциональной значимости для реализации этой особенности. Подход «ген-кандидат» позволяет сфокусироваться на одном или нескольких полиморфных вариантах в области гена (аллельный вариант), продукт которого, вероятно, вовлечен в развитие болезни или признака. Удешевление процедуры полногеномного скрининга с использованием микрочипов сверхвысокой плотности сделало доступным другой подход для поиска генетических предрасположенностей – полногеномный поиск ассоциаций.*

*Полагаем, что объединение обоих подходов в единый алгоритм выбора молекулярно-генетических маркеров для проведения точечного генотипирования позволит учитывать как маркеры, отобранные на основе априорного предположения о функциональной значимости гена-кандидата, так и поиск ассоциаций с изучаемым признаком на основании данных полногеномного поиска ассоциаций. Данный подход позволит оптимизировать диагностическую эффективность создаваемого тест-набора.*

**Ключевые слова:** *молекулярно-генетические маркеры, многофакторные фенотипические признаки, ген-кандидат, полногеномный поиск ассоциаций, точечное генотипирование, однонуклеотидный полиморфизм, алгоритм, секвенирование, критерии отбора, наследственная предрасположенность.*

Конец XX в. и начало нового тысячелетия ознаменовались бурным развитием молекулярной биологии, важнейшие достижения которой нашли практическую реализацию во многих областях клинической медицины.

Были созданы предпосылки для перехода к новой модели здравоохранения – так называемой 5П-медицине (предиктивная, персонализированная, превентивная, партисипаторная и позитивная) [5, 19]. Ее целью является выявление факторов риска, определение предрасположенности пациента к тем или иным болезням и их предотвращение. Поэтому модели, предсказывающие риск развития заболеваний, вызывают у исследователей все больший интерес [10].

Большинство сложных фенотипических признаков (особенностей), в том числе и заболеваний, не подчиняются закономерностям менделевского наследования, а имеют многофакторную природу, то есть существенный вклад в их развитие вносит генетическая составляющая в сочетании с влиянием факторов внешней среды. В популяции эти факторы распределяются и комбинируются случайным образом, действуя аддитивно: их суммарный эффект

равен сумме эффектов каждого из них в отдельности [1, 3, 4]. Заболевания с наследственной компонентой характеризуются большим числом клинических вариантов, образующих ряд переходных состояний: от минимальных стёртых форм до тяжёлых проявлений.

В формировании наследственной предрасположенности участвует не один, а множество генов, исторически получивших название генов-кандидатов или генов предрасположенности [3, 7, 15]. Число генов-кандидатов, взаимодействующих между собой и индивидуальным геномным «контекстом», формируют так называемую «генную сеть», тем самым влияют на формирование многофакторного признака и могут исчисляться десятками и сотнями. При этом на данный момент ни для одного мультифакторного признака (заболевания) не удалось определить исчерпывающий перечень вариантов генов. В большинстве современных исследований подчеркивается, что генов (вариантов), объясняющих значительное количество фенотипических отклонений, в популяции просто не существует [8, 14, 27, 28].

Более того, полиморфные варианты, имеющие достоверную ассоциацию с признаком, диффузно

располагаются в любых, в том числе и малоизученных участках генома, а не непосредственно в белок-кодирующих участках-экзонах, как считалось ранее. Например, В. Border et al. [11] показали, что приблизительно 1% полиморфных вариантов шизофрений находится в экзонах, остальные же наиболее ассоциированные варианты располагаются в некодирующих частях генома. Из этого следует, что секвенирование экзона не может быть использовано для проведения генетико-эпидемиологических исследований.

Большинство знаний о полиморфизме популяций и вариабельности генома были получены в результате реализации таких масштабных международных проектов, как «Геном человека», «НарМар» и «1000 геномов». Значительное удешевление технологии высокопроизводительного секвенирования (next generation sequencing) позволило многим странам создать собственные национальные проекты геномной медицины и проводить дополнительные популяционные генетические исследования на огромных выборках более 1 млн человек (Соединенные Штаты Америки, Великобритания, Китайская Народная Республика и Саудовская Аравия). Было установлено, что однонуклеотидный полиморфизм (single nucleotide polymorphism – SNP) является наиболее частой причиной существования нескольких вариантов (аллелей) одного гена. В общем контексте с SNP часто рассматривают и небольшие вставки или делеции участков геномной дезоксирибонуклеиновой кислоты (ДНК) [10]. Считается, что геномы двух неродственных индивидов различаются в среднем по одному нуклеотиду из каждой тысячи.

В текущей версии базы данных dbSNP (build 153), аккумулирующей информацию об однонуклеотидной вариабельности геномов, представлена информация о более чем 690 тысячах валидированных SNP в геноме человека. Открытие структуры индивидуальной вариабельности человеческого генома позволило использовать его в популяционной и эволюционной генетике, генетическом картировании болезней [9].

Значительные перспективы получили исследования, направленные на выявление ассоциации полиморфной генетической компоненты с индивидуальной предрасположенностью к заболеваниям, устойчивостью организма к воздействию факторов окружающей среды, определенным метаболическим профилем индивидуума, функциями высшей нервной деятельности и многими другими. Чаще всего это исследования, построенные на сравнении частот встречаемости SNP, предположительно ассоциированных с изучаемым признаком, в группе индивидуумов, демонстрирующих признак в сравнении с контрольной группой (ассоциативное исследование случай/контроль) [10].

Бурное развитие ассоциативных исследований в последние годы привело к накоплению значительного количества данных о полиморфных аллелях, показывающих различную степень доказательной связи с фенотипическими особенностями человека. Это связано с ускорением и удешевлением процедуры полно-

геномного скрининга индивидуальных генетических полиморфизмов. Данный факт диктует необходимость систематизации этих сведений.

Для решения большинства задач репликативных исследований с уникальной выборкой испытуемых на практике чаще всего используется точечное генотипирование SNP с ограниченным количеством маркеров в диагностической панели. Исследователю важно определить количество генетических маркеров, которое необходимо использовать при разработке тест-систем для определения индивидуальной предрасположенности к развитию многофакторной фенотипической особенности. Применение всех маркеров, ассоциированных с многофакторным заболеванием или признаком, неминуемо приведет к увеличению стоимости тест-системы и снижению ее специфичности. Напротив, исследование малого количества используемых маркеров отрицательно сказывается на чувствительности и уменьшает прогностические свойства набора.

Поэтому при планировании исследований рисков развития мультифакториальных заболеваний и признаков важное значение имеет разработка критериев отбора достоверных и информативных полиморфных локусов.

В настоящее время стратегия анализа генетической предрасположенности подразумевает существование нескольких подходов к отбору молекулярных маркеров.

Задолго до начала эры масштабных геномных исследований и изучения функциональных полиморфизмов в свете формирования наследственной предрасположенности генетические исследования основывались преимущественно на генеалогическом и эпидемиологическом подходах [19]. В дальнейшем эти подходы стали исходной точкой для таких методик генетического картирования сложных наследуемых признаков, как анализ сцепления, анализ ассоциаций в популяциях и семьях, методика идентичных по происхождению аллелей (параметрический метод), анализ неравновесия по сцеплению и анализ экспериментов по скрещиванию модельных организмов. Анализ сцепления, основанный на проверке конкретной модели наследования болезни в родословных, выявление генов, ассоциированных с определенными фенотипическими признаками, и создание мутантных генетических линий, проходивших параллельно с изучением структурной организации хромосом, явились основой для построения цитогенетических карт генов [26].

Несмотря на очевидные преимущества данных методик в выявлении и картировании признаков с традиционным менделевским типом наследования, они оказались малоэффективными при выявлении причинно-следственных связей для сложных многофакторных фенотипических особенностей [24].

На смену пришел новый подход к анализу генетической предрасположенности, основанный на понимании роли белкового продукта в этиологии,

патогенезе заболевания и молекулярных механизмов формирования признака, получивший название «ген-кандидат». Гены-кандидаты в исследование отбирают исходя из априорного знания биологического функционального воздействия гена на признак.

Подход «ген-кандидат» позволяет сфокусироваться на одном или нескольких полиморфных вариантах в области гена (аллельный вариант), продукт которого, вероятно, вовлечен в развитие болезни или признака. Генетический маркер считается ассоциированным с болезнью, если его частота среди больных значимо выше, нежели в контрольной выборке. Стандартное исследование проводится на группах из нескольких сотен больных и здоровых индивидов. Выполняют анализ распределения аллелей и генотипов исследуемого генетического маркера в выборке из неродственных здоровых лиц (популяционный контроль) и в группе больных (группа «случай») с тем, чтобы выявить значимые различия в частоте генетического маркера. Генетический маркер считается ассоциированным с болезнью, если его частота среди больных значимо выше, нежели в контрольной выборке [4].

Например, одним из первых успешных исследований, использующих ген-кандидатный подход, стало обнаружение мутации в некодирующей области АРОС3 (ген апополипротеина С3), которая связана с более высоким риском гипертриглицеридемии и атеросклероза [29]. Для многих распространенных признаков и заболеваний с наследственной предрасположенностью, таких как индивидуальные физические и психологические особенности организма, метаболизм, транспорт лекарственных препаратов, гипертоническая болезнь, бронхиальная астма, ишемическая болезнь сердца, сахарный диабет, по функциональному признаку были определены включающие аллельные варианты «наборы генов», ассоциированных с риском развития многофакторных фенотипических особенностей.

Несмотря на то, что «генные панели» широко используются в современной клинической лабораторной диагностике, данный подход не лишен существенных методологических недостатков. Подход «ген-кандидат» подразумевает наличие небольшого количества генов, оказывающих сильное влияние на формирование признака, в то время как современные исследования убедительно доказывают несостоятельность этой гипотезы.

Например, несмотря на признание факта полигенной природы многофакторных заболеваний, ряд исследователей продолжали настаивать на предположении о том, что шизофрения предопределяется несколькими или даже 1 полиморфным геном [22].

Также часть ассоциаций, полученных в исследованиях с использованием «кандидатного» подхода, могут оказаться ложными из-за счет малочисленности выборки, этнической неоднородности группы, некорректности клинических критериев отбора при формировании групп «случай/контроль» или неправильном представлении об этиопатогенезе заболевания, то

есть выраженной гетерогенности по негенетическим факторам риска.

В итоге нередко возникали ситуации с «плавающими» оценками генетических маркеров, иными словами – ассоциации ранее выявленных «аллелей риска» пересматривались в результатах более поздних исследований. Например, недавнее исследование не смогло подтвердить достоверность ассоциативных связей с депрессией для 18 ранее хорошо изученных генов-кандидатов, несмотря на значительно больший объем выборки [11]. Данные факты заставляют многие научные коллективы пересматривать систематизированные ранее результаты исследований. Например, Национальный институт психического здоровья США, признавая актуальность изучения геномных аспектов поведения, в 2017 г. удалил ссылки на конкретные гены, подчеркнув при этом необходимость повышения надежности данных ассоциативных исследований.

Удешевление процедуры полногеномного скрининга с использованием ДНК-микрочипов сверхвысокой плотности сделало доступным другой подход для поиска генетических предрасположенностей – полногеномный поиск ассоциаций (Genome – wide Association Studies – GWAS). Уже сейчас данный подход использует недоступные ранее для генетического исследования выборки в несколько сотен тысяч или даже миллионов человек при отсутствии сведений о функции и расположении связанных с болезнью генов на хромосоме. В каждом индивидуальном геноме анализируются миллионы полиморфных генетических маркеров – SNP. Постоянно увеличивающиеся объемы выборки значительно повышают достоверность генетической ассоциации. Например, в исследовании диабета второго типа, проводимом Оксфордским университетом, участвовало более 800 тысяч человек [21].

Полногеномный поиск ассоциаций устанавливает корреляцию конкретного варианта генома, характеризующегося уникальным набором SNP, с наличием какого-то заболевания или признака. GWAS основывается на анализе частоты встречаемости SNP: если при генотипировании внутри выборки некоторые полиморфные варианты встречаются у людей с исследуемым фенотипом значимо чаще, чем другие, именно их можно условно признать «ответственными» за проявление этого фенотипа. Таким образом, главные критерии применимости GWAS – наличие репрезентативной выборки (как правило, с большим количеством участников) и, конечно, сама возможность выявить связь (ассоциацию) между генотипом и фенотипом.

Первый в истории результативный GWAS был проведен сотрудниками лаборатории статистической генетики Рокфеллеровского университета в 2005 г. на выборке из 96 пациентов с возрастной макулярной дегенерацией и 50 здоровых людей контрольной группы. Сравнение их геномов выявило два ассоциативных SNP [18]. Эти SNP были локализованы в гене, кодирующем фактор комплемента H, что было

неожиданным открытием в исследовании возрастной макулярной дегенерации. Результаты этого анализа привели к дальнейшим функциональным исследованиям в направлении терапевтических манипуляций с системой комплемента [16].

Дальнейшие масштабные исследования были продолжены в рамках консорциума Wellcome trust case control consortium (WTCCC). Исследование включало геномный скрининг 14000 случаев таких распространенных многофакторных заболеваний, как ишемическая болезнь сердца, диабет 1-го и 2-го типов, ревматоидный артрит, болезнь Крона, биполярное расстройство и гипертоническая болезнь. В 2007 г. в журнале «Nature» были опубликованы результаты анализа молекулярных механизмов развития этих заболеваний, выявлен факт наличия общих полиморфных генов и метаболических путей [14].

Количество и масштаб выполненных GWAS каждый год прогрессивно увеличивается, поэтому удобно пользоваться представленными в открытом доступе агрегаторами данных, систематизирующими и каталогизирующими результаты большинства полногеномных исследований ассоциаций, такими как Phe GenI, GWAS Central, GWAS Catalog и др. В данных каталогах для каждого исследованного полиморфного генетического маркера отдельно рассчитываются статистические параметры, характеризующие степень достоверности его ассоциации с многофакторным признаком. В большинстве случаев это отношение шансов (ОШ), относительный риск (ОР), бета-коэффициент или частоты аллелей среди больных и здоровых [6]. Показатели ОШ и ОР имеют схожий смысл в исследованиях типа «случай/контроль», позволяя определить степень влияния полиморфного аллеля на изучаемый признак и показывают, во сколько раз увеличивается или уменьшается вероятность быть здоровым (больным) при носительстве вариантного аллеля. Часто дополнительно определяют чувстви-

тельность, специфичность и диагностическую валидность ассоциативных маркеров [2].

Параметром, определяющим достоверность полученных различий между исследуемыми группами людей с многофакторным признаком и без него, является р-уровень значимости (p-value). Достоверно ассоциированными в GWAS-исследованиях необходимо считать те генетические маркеры, ассоциация которых достигла полногеномной значимости ( $p < 5 \cdot 10^{-8}$ ) и была подтверждена на независимой выборке, то есть чем ниже данный показатель, тем достовернее ассоциация SNP с фенотипическим признаком (более мягкие критерии допускают использование значения р меньше, чем  $10^{-5}$ , для установления номинальной значимости) [25].

Для визуализации данных, полученных в GWAS, часто используют так называемые манхэттенские графики (рис. 1), где степень ассоциации отображается как отрицательный логарифм р-значения в зависимости от локуса, то есть геномных координат SNP [17].

Вместе с тем, несмотря на громкие заявления о завершении эры изучения «генов-кандидатов» [15], ряд исследователей подвергает GWAS-подход критике. Основной ошибкой признается тот факт, что обнаружение ассоциации еще не означает прямой каузальной связи с ожидаемым событием. Более того, даже если такого рода ассоциации и имеются, мы пока не можем объяснить, как индивидуальный геномный контекст и негенные факторы (образ жизни и факторы окружающей среды) изменяют соотношение рисков формирования сложных признаков.

Вклад конкретного генетического варианта часто является косвенным. Например, полиморфизм может влиять на работу регуляторного комплекса, связанного с сотней генов, а некоторые из них уже могут быть непосредственно связаны с заболеванием или признаком. Исходя из этого, E.A. Boyle et al. [12] предлагают признать, что в реальности все влияет

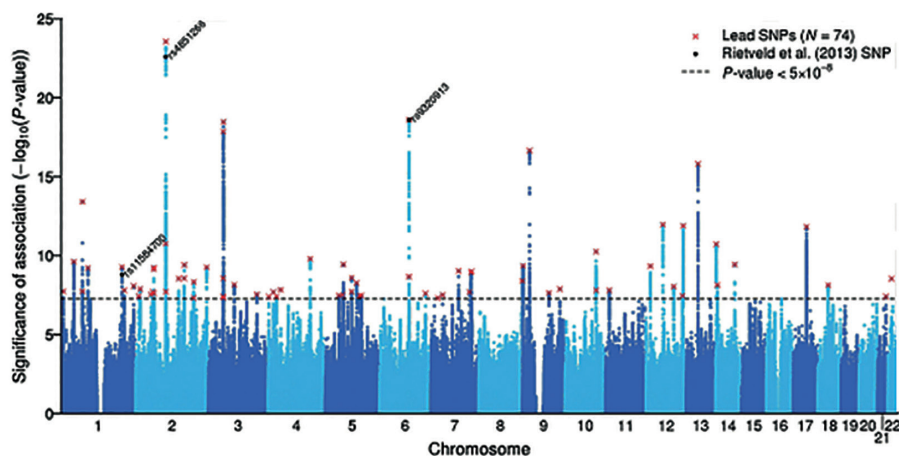


Рис. 1. Манхэттенский график EY, где каждая из точек представляет собой SNP. Ось X — расположение SNP в геноме, ось Y — уровень ассоциации каждого SNP (чем сильнее ассоциация с признаком, тем ниже ее р-значение, а значит, выше отрицательный логарифм этого значения и соответствующий «столбец»). На графике представлены SNP из GWAS 2013 г. и 74 недавно обнаруженных локуса (помечены красными крестиками), связанных с обучением [23]



на все, и перейти от концепции «мультигенных» признаков к «всегенной» концепции для интерпретации уже собранных данных. Необходимо сосредоточиться на анализе внутриклеточных регуляторных сетей, а не проводить новые, еще более масштабные исследования. Только так, по их мнению, можно понять, почему вариант X ассоциирован с признаком Y, и сделать практические выводы.

К некорректному поиску ассоциаций могут приводить не недостатки методики, а такие типичные ошибки планирования исследования, как отсутствие однородной выборки или четких критериев для определения изучаемой фенотипической особенности.

С другой стороны, именно активное внедрение полногеномного скрининга ассоциаций в последние годы позволило убедительно подчеркнуть полигенность многофакторных фенотипических особенностей, диффузность распределения полиморфных генетических локусов по всему геному и практически полное отсутствие SNP, оказывающих сильное влияние на развитие признака. Основная концепция традиционного «ген-кандидатного» подхода также вынужденно корректируется, поскольку данные GWAS указывают на то, что в белок-кодирующих участках генома-экзонах локализуется минимальное количество полиморфных вариантов, а чаще всего их расположение фиксируют в межгенных пространствах, интронах, untranslated region, некопирующих рибонуклеиновых кислотах и прочих относительно плохо изученных областях генома.

Эти факты в совокупности поставили под сомнение результаты многих ранее проведенных ассоциативных исследований, а результаты некоторых из них были признаны ложноположительными [11, 15]. Это объясняет причину изменения подхода к оценке новых и ранее полученных данных, а также пересмотр результатов многих экспериментов по поиску ассоциаций.

Учитывая вышеописанные особенности GWAS-исследований, можно рекомендовать к использованию следующие критерии отбора генетических маркеров [30]:

- необходимо учитывать только генетические маркеры с достоверностью ассоциации  $p < 5 \cdot 10^{-8}$  и ниже;
- обнаруженные ассоциации должны быть подтверждены минимум в 2 независимых GWAS-исследованиях, либо дополнительных метаанализах высокого уровня;
- при исследовании одного маркера в нескольких GWAS предпочтение следует отдавать более поздним анализам ассоциаций, включавшим большее количество участников;
- размеры выборки участников исследования должны быть достаточно велики для возможности обнаружения ассоциации генетических маркеров с определенными частотами встречаемости;
- однородная выборка должна иметь четкие фенотипические критерии.

Постоянное увеличение объемов выборки и открытие новых молекулярно-генетических маркеров с

различной локализацией в геноме позволит сделать GWAS более надежными и улучшить прогностическую ценность оценки риска развития заболевания с генетической предрасположенностью.

В целом следует признать, что, несмотря на наличие некоторых недостатков, модель полногеномного поиска ассоциаций на данный момент признана многими авторами как наиболее достоверно обосновывающая молекулярные механизмы формирования мультифакторных фенотипических особенностей и перспективна для выявления новых ассоциативных генетических маркеров.

Для анализа ассоциативной связи с многофакторными ассоциативными особенностями необходимо применить комплексный подход к выбору ограниченного набора полиморфных генетических локусов. Это позволит учитывать функциональную роль полиморфизмов, которая важна для проработки потенциальной терапевтической стратегии и перспективности новых маркеров в сочетании с высоким уровнем статистической значимости предоставленных GWAS-подходом.

При отборе полиморфных локусов для проведения ассоциативных исследований предлагается сфокусироваться на изучении молекулярных механизмов формирования многофакторной фенотипической особенности (рис. 2).

На первом этапе поиска необходимо проанализировать биохимические процессы, происходящие на различных уровнях, включая внутриклеточные и межклеточные сигнальные каскады. Эту информацию можно получить в современных литературных источниках, руководствах и открытых онлайн-базах данных, таких как KEGG PATHWAY или Reactome Pathway и др. В результате анализа необходимо сформировать список генов, вовлеченных в процесс формирования признака. Одновременно с изучением молекулярных механизмов рекомендуется провести анализ фенотипических особенностей изучаемого признака, что позволит провести поиск в каталогах признаков, ранее изученных в различных GWAS. Дополнительное направление поиска дает возможность формировать запросы вне функциональной/генной составляющей.

На втором этапе гены и фенотипические признаки из списка необходимо проанализировать в нескольких независимых онлайн-агрегаторах GWAS-исследований. После получения уникальных идентификационных номеров маркеров ( $r_s$  в dbSNP) и данных по достоверности ассоциации и силе эффекта необходимо дополнительно актуализировать эти сведения по ссылкам на литературные источники. Если ассоциация полиморфизма с фенотипическим признаком достоверно достигла уровня полногеномной значимости  $p < 5 \cdot 10^{-8}$ , то такой маркер можно признать перспективным к включению в список приоритетных для исследования вариантов. В случае, если достоверность ассоциативной связи полиморфного локуса превысила уровень номинальной значимости  $p < 5 \cdot 10^{-5}$ , но не достигла уровня  $p < 5 \cdot 10^{-8}$ , возможен альтерна-

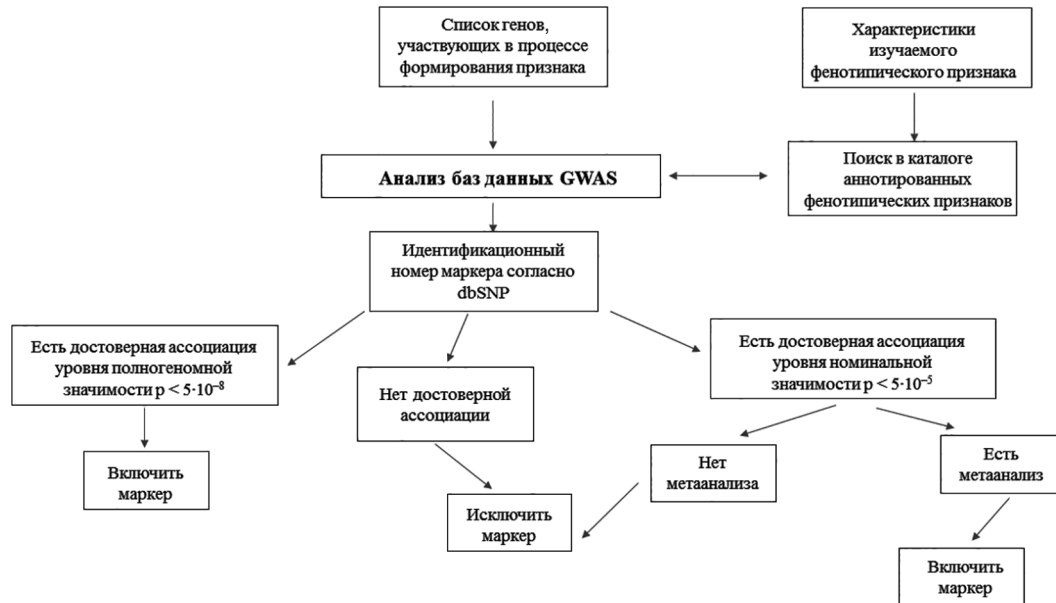


Рис. 2. Алгоритм изучения молекулярного механизма формирования многофакторного признака

тивный поиск взаимосвязи в метаанализах высокого уровня. Согласно результатам метаанализа, генетический маркер возможно признать перспективным вариантом для исследования. При отсутствии ассоциации полиморфизма с многофакторной фенотипической особенностью по данным GWAS-исследований и метаанализам маркер не следует использовать при проведении репликативных исследований.

Таким образом, предлагаемый алгоритм позволяет учитывать как маркеры, отобранные на основе априорного предположения о функциональной значимости «гена-кандидата», так и поиск ассоциаций с изучаемым признаком по всему геному на основании GWAS-исследований.

В целом существует несколько подходов к дизайну ограниченного набора полиморфных маркеров для точечного генотипирования. Селекцию отдельных ДНК-маркеров осуществляют на основе либо их статистически значимой ассоциации с изучаемым многофакторным признаком, либо функциональной значимости для реализации этой особенности.

Учитывая описанные преимущества и недостатки обеих стратегий, предлагаем объединить их в комплексный алгоритм. Представленная последовательность действий позволит учитывать как маркеры, отобранные на основе априорного предположения о функциональной значимости гена-кандидата, так и поиск ассоциаций с изучаемым признаком на основании данных GWAS. Данный подход позволит оптимизировать диагностическую эффективность создаваемого тест-набора.

Современные молекулярно-генетические исследования характеризуют значительные массивы полученных данных с возрастающей сложностью применяемых к ним аналитических подходов. Появление новых сведений о биологических процессах, лежащих

в основе формирования многофакторных фенотипических особенностей человека, позволит провести корректировку представленной методики либо выполнить построение новой модели, включающей уточняющие критерии отбора генетических маркеров.

### Литература

1. Баранов, В.С. Генетический паспорт – основа индивидуальной и предиктивной медицины / В.С. Баранов. – СПб.: Изд-во Н-Л, 2009. – 528 с.
2. Гинтер, Е.К. Наследственные болезни: национальное руководство: краткое издание / Е.К. Гинтер, В.П. Пузырев. – М.: ГЭОТАР-Медиа, 2017. – 464 с.
3. Горбунова, В.Н. Медицинская генетика: учебник для студентов мед. вузов и слушателей последипломного образования / В.Н. Горбунова. – СПб., 2012. – 357 с.
4. Иванов, В.И. Геномика – медицине / В.И. Иванов. – М.: Академкнига, 2005. – 392 с.
5. Лебедев, А.А. Превентивная медицина – медицина XXI века / А.А. Лебедев, М.В. Гончарова // Нац. проекты. – 2008. – № 12 (31). – С. 40–43.
6. Низамутдинов, И.И. Критерии отбора генетических маркеров для анализа предрасположенности к многофакторным фенотипическим особенностям / И.И. Низамутдинов [и др.] // Вестн. РГМУ. – 2016. – № 6. – С. 25–32.
7. Ньюсбаум, Р.Л. Медицинская генетика: учебное пособие. Пер. с англ. А.Ш. Латыпова / Р.Л. Ньюсбаум, Р.Р. Мак-Иннес, Х.Ф. Виллард. – М.: ГЭОТАР-Медиа, 2010. – 624 с.
8. Пузырев, В.П. Генетические основы коморбидности у человека / В.П. Пузырев // Генетика. – 2015. – № 51 (4). – С. 491–502.
9. Пузырев, В.П. Патологическая анатомия генома человека / В.П. Пузырев, В.А. Степанов. – Новосибирск: Наука. Сиб. предприятие РАН. – 1997. – 224 с.
10. Ребриков, Д.В. NGS: высокопроизводительное секвенирование / Д.В. Ребриков [и др.]. – М.: БИНОМ., 2015. – 232 с.
11. Border, R. No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples / R. Border [et al.] // Am. J. Psychiatry. – 2019. – № 176 (5). – P. 376–387.

12. Boyle, E.A. An Expanded View of Complex Traits: From Polygenic to Omnigenic / E.A. Boyle [et al.] // Cell. – 2017. – № 169 (7). – P. 1177–1186.
13. Brookes, A.J. The essence of SNPs / A.J. Brookes // Gene. – 1999. – № 234. – P. 177–186.
14. Burton, P.R. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls / P.R. Burton [et al.] // Nature. – 2007. – № 447 (7145). – P. 661–678.
15. Duncan, L.E. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete / L.E. Duncan, M. Ostacher, J. Ballon // Neuropsychopharmacology. – 2019. – P. 1–6.
16. Fridkis-Hareli, M. Design and development of TT30, a novel C3d-targeted C3/C5 convertase inhibitor for treatment of human complement alternative pathway-mediated diseases / M. Fridkis-Hareli [et al.] // Blood. – 2011. – № 118 (17). – P. 4705–4713.
17. Gibson, G. Hints of hidden heritability in GWAS. / G. Gibson // Nature Genetics. – 2010. – № 42 (7). – P. 558–560.
18. Haines, J.L. Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration / J.L. Haines // Science. – 2005. – № 308 (5720). – P. 419–421.
19. Hood, L. Systems biology and p4 medicine: past, present, and future / L. Hood // Rambam Maimonides Med. J. – 2013. – № 4 (2). – P. 12.
20. Kety, S.S. Mental Illness in the Biological and Adoptive Relatives of Schizophrenic Adoptees / S.S. Kety // Archives of General Psychiatry. – 1994. – № 51 (6). – P. 442.
21. Mahajan, A. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps / A. Mahajan [et al.] // Nature Genetics. – 2018. – № 50 (11). – P. 1505–1513.
22. Middeldorp, C.M. The value of polygenic analyses in psychiatry / C.M. Middeldorp, N.R. Wray // World Psychiatry. – 2008. – № 17 (1). – P. 26–28.
23. Okbay, A. Genome-wide association study identifies 74 loci associated with educational attainment / A. Okbay [et al.] // Nature. – 2016. – № 533 (7604). – P. 539–542.
24. Panoutsopoulou, K. Finding common susceptibility variants for complex disease: past, present and future / K. Panoutsopoulou, E. Zeggini // Briefings in Functional Genomics and Proteomics. – 2009. – № 8 (5). – P. 345–352.
25. Pearson, T.A. How to interpret a genome-wide association study / T.A. Pearson, T.A. Manolio // JAMA. – 2008. – № 299 (11). – P. 1335–1344.
26. Petty, E.M. Handbook of human genetic linkage / E.M. Petty // Trends in Endocrinology & Metabolism. – 1995. – № 6 (1). – P. 30–31.
27. Pickrell, J.K. Detection and interpretation of shared genetic influences on 42 human traits / J.K. Pickrell [et al.] // Nature Genetics. – 2016. – № 48 (7). – P. 709–717.
28. Polderman, T.J.C. Meta-analysis of the heritability of human traits based on fifty years of twin studies / T.J.C. Polderman [et al.] // Nature Genetics. – 2015. – № 47 (7). – P. 702–709.
29. Rees, A. Dna polymorphism adjacent to human apoprotein a-1 gene: relation to hypertriglyceridaemia / A. Rees [et al.] // The Lancet. – 1983. – № 321 (8322). – P. 444–446.
30. Wang, W.Y.S. Genome-wide association studies: theoretical and practical concerns / W.Y.S. Wang [et al.] // Nature Reviews Genetics. – 2005. – № 6 (2). – P. 109–118.

G.G. Kutelev, A.B. Krivoruchko, A.E. Trandina, A.M. Ivanov,  
D.V. Cherkashin, A.A. Marchenko, S.L. Grishaev

### **New approaches to the selection of genetic markers associated with multifactorial phenotypic traits**

**Abstract.** Modern approaches to searching for associations between the studied phenotype and structural variations of the human genome are analyzed. Most complex phenotypic traits, including diseases, do not follow the laws of Mendelian inheritance, but have a multi-factor nature, that is, a significant contribution to their development is made by the genetic component in combination with the influence of environmental factors. In General, there are several approaches to the design of a limited set of polymorphic markers for point genotyping. Selection of individual molecular genetic markers is carried out based on either their statistically significant Association with the studied multivariate feature, or their functional significance for the implementation of this feature. The «candidate gene» approach allows you to focus on one or more polymorphic variants in the region of a gene (allelic variant), the product of which is likely involved in the development of a disease or trait. The cheaper procedure for full-genome screening using ultra-high-density microchips has made available another approach for searching for genetic predispositions – full – genome Association search. We believe that the unification of both approaches into a single algorithm for the choice of molecular genetic markers to conduct point genotyping will allow for both markers selected based on a priori assumptions about the functional significance of candidate genes, and Association with the studied trait on the basis of genome-wide associations search. This approach will optimize the diagnostic efficiency of the created test suite.

**Key words:** molecular genetic markers, multi-factor phenotypic traits, candidate gene, genome – wide Association Studies, point genotyping, single-nucleotide polymorphism, algorithm, sequencing, selection criteria, hereditary predisposition.

Контактный телефон: 8-921-639-89-54; e-mail: vmeda-nio@mil.ru