

# №10 1999 год

## ГЕНЕТИЧЕСКИЙ АНАЛИЗ: ЭТАПЫ РАЗВИТИЯ, ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

Генетическим анализом называют систему мероприятий, направленных на изучение механизмов генетической детерминации признаков. Среди них видное место отводится задачам формальной генетики, а именно формулированию и формализации моделей наследования и проверке генетических гипотез на конкретном эмпирическом материале. В мировой литературе этот раздел генетического анализа известен под названием *segregation and linkage analysis*. Именно о нем пойдет речь.

Как известно, генетика начиналась с генетического анализа, который был чрезвычайно популярен в первой половине века, являясь основным инструментом исследований. Затем популярность его падает, и новый подъем интереса к этому разделу генетики наблюдается лишь в последние годы. Как это ни парадоксально, но и потеря и новое приобретение популярности генетического анализа вызваны развитием молекулярно-генетических исследований. Причина первого понятна. Формальная генетика оперирует абстрактным понятием генов как неких дискретных факторов, линейно расположенных в группах сцепления и прямо или опосредованно влияющих на проявление признака. Формальная генетика не рассматривает проблем, связанных со структурой гена, с особенностями его функционирования, процессов, происходящих на пути от гена до признака. Именно поэтому в формальной генетике любое решение задачи имеет вероятностный, а не абсолютный характер. Результатом решения является формулировка модели наследования признака и указание степени ее правдоподобия. Молекулярная генетика, напротив, дает вполне конкретные ответы: она позволяет шаг за шагом проследить весь путь от гена до признака, выяснить, каким образом изменение нуклеотидного состава какого-либо участка ДНК приводит к полиморфизму признака. Очевидно, что знания, полученные методами молекулярной генетики, более однозначны и, следовательно, более ценны, чем результаты генетического анализа. Именно поэтому в 1960–1970-е годы постепенно формируется и закрепляется представление о том, что генетический анализ изжил себя и его следует рассматривать как один из архаизмов генетики.

Это утверждение было бы справедливо, если бы не существовало широкого класса признаков, называемых сейчас комплексными (*complex traits*). Эти признаки проявляются на уровне целого организма, в их контроле задействовано большое число генетических и средовых факторов, часто взаимодействующих друг с другом. К комплексным признакам относится подавляющее большинство количественных признаков, характеризующих хозяйственно ценные свойства растений и животных, сюда относят многие распространенные наследственные болезни человека, такие, как гипертония, диабет, ишемическая болезнь сердца, некоторые формы онкологических заболеваний, многие психические расстройства. Очевидно, что, пользуясь только методами молекулярной генетики, нельзя разобраться в природе этих признаков. Прогресс может быть достигнут лишь при объединении усилий молекулярной и формальной генетики. Поскольку темпы развития генетического анализа существенно отставали от темпов развития молекулярной генетики, сейчас генетический анализ является лимитирующим звеном на пути прогресса генетики комплексных признаков. Именно поэтому в последние годы наблюдается заметное повышение интереса к формальной генетике: усиливается финансирование, организуются конференции, семинары и школы. К сожалению, в нашей стране пока не ощущается больших перемен, и для многих генетиков термин «генетический анализ» ассоциируется с исследованиями, проводимыми в начале века. Чтобы продемонстрировать прогресс в этом направлении генетики, я попыталась проследить логику развития генетического анализа, эволюцию объектов исследования и предположений, в рамках которых велся анализ, а также указать те проблемы, которые необходимо решить, чтобы достичь прогресса в понимании природы комплексных признаков.

В истории развития генетического анализа можно выделить три основных этапа, отличающихся друг от друга объектом исследования, эмпирическим материалом, тестируемыми генетическими гипотезами и общими предположениями, в рамках которых ведется анализ. Объектом исследования этого раздела генетики является признак, который может быть формально представлен как бинарная, качественная или количественная характеристика.

Первый этап генетического анализа называют обычно классическим генетическим анализом, он базируется на гибридологическом эксперименте и строится на двух предположениях:

- исходные родительские линии гомозиготны по локусам, принимающим участие в контроле признака;
- главное место в контроле признака отводится майоргенам, полигенная и средовая компоненты пренебрежимо малы.

Из первого предположения следует, что обе родительские формы и  $F_1$  являются генетически однородными группами, в которых все особи абсолютно идентичны по генам, принимающим участие в формировании изучаемого признака. Следовательно, потомки всех однотипных скрещиваний между представителями этих трех групп ( $P_1$ ,  $P_2$  и  $F_1$ ) могут быть объединены без какой-либо потери информации.

Из второго предположения следует, что все особи с одинаковым генотипом фенотипически неразличимы. Следовательно, сложность формализованного представления признака ограничена числом генотипов. Именно поэтому объектом исследования на этом этапе являются бинарные признаки.

Таким образом, в рамках сделанных предположений эмпирический материал может быть представлен небольшим числом групп большой численности. Задача заключается в том, чтобы проверить, соответствуют ли эмпирические данные сформулированной модели наследования этого признака. На первом этапе модели наследования максимально просты – это рецессивная или доминантная модели. Единственным параметром интереса является сегрегационная частота, определенная как вероятность родителей с определенными генотипами произвести потомка с заданным фенотипом. Ожидаемая сегрегационная частота

определяется типом скрещивания и моделью наследования. Генетические гипотезы проверяются одним из двух статистических методов: либо подсчитываются ожидаемые численности различных фенотипических классов и сравниваются с наблюдаемыми, либо оценивается сегрегационная частота и сравнивается с ожидаемой.

Предельно прост и анализ сцепления – коэффициент рекомбинации оценивается долей рекомбинантных потомков, поскольку фаза сцепления генов у дигетерозиготного родителя, полученного в гибридологическом эксперименте, всегда известна.

Эволюция метода на первом этапе заключается главным образом в усложнении моделей за счет увеличения числа локусов, принимающих участие в контроле признака, и за счет рассмотрения различных типов взаимодействия генов.

Второй этап развития сегрегационного анализа характеризуется сменой объекта исследования, переходом на анализ признаков человека. У человека, как известно, существуют проблемы с получением чистых линий и принудительными скрещиваниями. Поэтому здесь нельзя использовать методы классического генетического анализа. К счастью, человек обладает одним замечательным свойством, не присущим животным и тем более растениям, – большинство людей помнят своих предков, а это значит, что можно строить и реконструировать родословные и выбирать из популяции те из них, которые генетику хотелось бы получить в эксперименте. На втором этапе материал максимально приближен к гибридологическому эксперименту, он представлен потомками родителей с известными генотипами. По своей идеологии этот этап очень близок первому. По-прежнему анализируют бинарные признаки, по-прежнему тестируют доминантные или рецессивные модели. Предположения почти идентичны тем, что были на первом этапе: предполагается, что генотипы родителей известны и признак контролируется главным образом майоргенами. Разница заключается в том, что эмпирический материал берется из популяции, а не из гибридологического эксперимента. Каким образом выбираются из популяции скрещивания с известными генотипами? При доминантном наследовании наиболее информативными являются скрещивания особей с нормальным гомозиготным и с гетерозиготным генотипами. Их легко выбрать из популяции, если мутантный аллель встречается с низкой частотой – практически все родительские пары, в которых один из супругов болен, а другой здоров, будут иметь требуемые генотипы. Потомки всех таких скрещиваний объединяются, и проводится анализ точно так же, как на первом этапе. Дополнительно используется проверка того, что распределение семей по числу больных потомков подчиняется биномиальному распределению.

При рецессивном контроле признака информативным является скрещивание двух гетерозигот. Популяционным аналогом этого скрещивания являются семьи, в которых оба родителя здоровы, а хотя бы один ребенок болен. Очевидно, что при таком условии в выборку никогда не попадут те семьи, где у гетерозиготных родителей чисто случайно не появилось ни одного больного ребенка. Поэтому полученные обычным образом оценки сегрегационных частот оказываются завышенными. Для коррекции этого смещения необходимо отказаться от объединения потомков аналогичных скрещиваний и рассматривать индивидуальные семьи.

Так же, как на первом этапе, здесь тестируются наиболее простые модели доминантного и рецессивного контроля признака. Развитие метода идет по трем направлениям. Во-первых, так же, как на первом этапе, усложняются модели, но эти усложнения, как правило, лежат в рамках моногенных диаллельных моделей. Во-вторых, происходит мощнейшая разработка так называемой ascertainment problem – проблемы коррекции смещения, вызванного неслучайностью выбора: формулируется понятие пробанда и разрабатываются основные принципы решения проблемы. И в-третьих, отказ от объединения потомков и затем введение дополнительного параметра (вероятности выбора пробанда) приводят к заметному усложнению оценки сегрегационной частоты. Возникает необходимость разработки специальных алгоритмов получения оценок, разработки специальных таблиц и формул, ведь это еще докомпьютерная эра и все вычисления делаются вручную.

Значительно усложняется анализ сцепления. Это уже не гибридологический эксперимент, и фаза сцепления генов у дигетерозиготного родителя, как правило, неизвестна. При этом уменьшается информативность данных и накладываются более жесткие требования к точности анализа. Именно в эти годы появляются две принципиально важные для анализа сцепления разработки: предлагается непараметрический метод тестирования сцепления, основанный на анализе идентичности генов по происхождению, и предлагается специальный критерий Lod score, благодаря которому был достигнут огромный прогресс в локализации генов, ответственных за менделевские признаки.

Тем не менее, второй этап развития генетического анализа – это анализ простых признаков и простых моделей. Лимитирующим здесь оказывается предположение об известных генотипах родителей. В самом деле, только при рецессивной или доминантной модели можно выбрать из популяции семьи, удовлетворяющие этому условию. Получается замкнутый круг – высказывается предположение о модели наследования, при ее справедливости устанавливаются генотипы родителей, а затем по фенотипам потомков этих родителей тестируется выбранная модель.

Выход из этой ситуации намечился лишь в 1970 г., когда было снято предположение о том, что генотипы родителей известны. Сначала было предложено рассматривать распределение фенотипов не только потомков, но и родителей, проводя суммирование по всем возможным генотипам родителей. А буквально на следующий год, в 1971 г., были сформулированы основы того, что сейчас называется комплексным сегрегационным анализом. Были введены три математические компоненты, описывающие распределение генотипов в популяции, правило передачи аллелей из поколения в поколение и правило построения фенотипа на заданном генотипе. Было показано, что в терминах этих трех распределений можно записать вероятность совместного распределения признаков у членов родословной. Родословная эта может иметь произвольную структуру, признак может быть формализован любым способом, в том числе быть количественным, и модели наследования могут не ограничиваться простыми рецессивными или доминантными отношениями. Кроме того, вычленение в отдельную компоненту вероятностей передачи генов от родителей потомкам позволяет ввести специальные критерии для проверки менделевской сегрегации генов.

Развитие метода продолжается главным образом по двум направлениям:

- Создаются новые более сложные модели наследования. Прежде всего в модель в явном виде вводится майоргенная, полигенная и средовая компоненты. Появляется возможность учесть общие внутрисемейные эффекты, фенотипические корреляции между родственниками, наличие факторов риска развития болезни, разный возраст проявления болезни и т.д.

— Анализ родословных произвольной структуры возможен лишь с помощью вычислительной техники, поэтому отдельное направление посвящено созданию алгоритмов и пакетов программ для генетического анализа.

Как видно, на третьем этапе генетический анализ выходит за рамки ограничений, характерных для классического генетического анализа, доступными для анализа становятся количественные признаки, модели наследования включают в себя различные генетические и средовые факторы. Теоретически проблема анализа комплексных признаков может считаться решенной. К сожалению, на практике все не так просто.

Как отмечалось выше, был предложен специальный критерий для проверки менделевской сегрегации генов. В основе его лежит сравнение трех гипотез, отличающихся друг от друга предположениями о переходных вероятностях. Менделевская гипотеза фиксирует те значения этих вероятностей, которые соответствуют менделевским правилам. Общая, или неограниченная, модель не накладывает никаких ограничений на эти параметры, и средовая модель предполагает их равными. Менделевская сегрегация генов считается установленной, когда первая и вторая гипотезы не отличаются друг от друга, а вторая и третья значимо отличимы. В специальном генетико-стохастическом эксперименте было показано, что использование этого комплексного критерия препятствует ошибочному установлению менделизма. Однако, если майоргенный контроль не установлен, не обязательно, что его нет на самом деле. Причиной ложного отвержения менделизма чаще всего бывает генетическая гетерогенность, ошибки диагностики или некорректная формализация количественного признака, а также неадекватность модели наследования. В настоящее время существует множество подходов к решению этой проблемы: создаются критерии гетерогенности выборки, методы трансформации количественных данных, расширяются и усложняются модели наследования признаков. Однако основная проблема комплексных признаков заключается в том, что они настолько сложны и многообразны, что очень трудно, если не невозможно, описать их наследование единой функцией так, чтобы, варьируя значения параметров, можно было формализовать любую ситуацию. Даже если бы это удалось, число оцениваемых параметров было бы настолько велико, что вряд ли удалось бы собрать столько эмпирических данных, чтобы оценить эти параметры. Поэтому при анализе комплексных признаков приходится вести большую предварительную работу, изучая половые и возрастные особенности их проявления, выделяя модифицирующие факторы и формируя модель наследования, наиболее полно отражающую специфику анализируемого признака. Пока нет и речи о том, чтобы сделать анализ комплексных признаков автоматическим.

Та же ситуация наблюдается и для анализа сцепления. Прекрасно зарекомендовавший себя и обеспечивший огромный прогресс в картировании менделевских признаков критерий *Lod score* предполагает, что единственным оцениваемым параметром является коэффициент рекомбинации, а все параметры модели наследования известны. Пока есть проблемы с установлением модели наследования, анализ сцепления также не может быть переведен на автоматический уровень. Обычно используются два методических подхода: анализ генов-кандидатов и сканирование генома. Первый подход предполагает, что известны те ключевые звенья в цепи биохимических реакций, которые в принципе могут повлиять на проявление признака. Тестирование генов, ответственных за эти звенья, на причастность к контролю анализируемого признака осуществляется обычно непараметрическими методами анализа. Они хороши тем, что не требуют задания модели наследования, однако недостатком их является пониженная мощность. Другой путь – сканирование генома – заключается в том, что все хромосомы насыщаются анонимными маркерами, и предполагаемый ген, ответственный за развитие признака, «зажимается» между двумя соседними маркерами. Из-за невысокой информативности данных установить можно только тесное сцепление. Но для этого надо, чтобы маркеров было много и они плотно покрывали хромосому. Однако чем больше маркеров, тем больше повторных сравнений и тем выше уровень значимости критериев. Возможно, что именно из-за этих недостатков используемых подходов так редко удается добиться успешного картирования генов, ответственных за развитие комплексных признаков.

Сейчас развитие методов генетического анализа идет по пути совершенствования анализа сцепления. Большие надежды возлагаются на многоточечное картирование и на новые подходы, объединяющие параметрический и непараметрический методы, сегрегационный и рекомбинационный анализ.

Таким образом, несмотря на большой прогресс генетического анализа, мы еще весьма далеки от того, чтобы довести анализ комплексных признаков до автоматизма, который в свое время обеспечил прорыв в картировании и идентификации генов, ответственных за развитие менделевских признаков. К сожалению, ощущается явная нехватка специалистов в этой области, поскольку в мире насчитывается от силы два десятка лабораторий, где все эти годы продолжались разработки методов генетического анализа. Наш сектор является единственным в России центром, где ведутся такие исследования.

*Т.И.Аксенович*, д.б.н.,  
зав. сектором рекомбинационного и сегрегационного анализа,  
ИЦиГ СО РАН, Новосибирск