**Опять актуален вопрос: что такое ген?**

Кандидат биологических наук В. В. Вельков

Чудес есть много, о Горацио, Которые не снились нашим мудрецам…

Шекспир

**Один ген — один признак**

Создатель генетики, монах августинского монастыря Георг Иоанн Мендель, слова „ген“ не знал. Но благодаря результатам своих чётко продуманных и безупречных опытов пришёл к выводу, что у организмов существуют какие-то дискретные наследственные факторы, определяющие дискретные морфологические признаки. Гениальный и дерзкий труд смиренного монаха, опубликованный в 1866 г., остался незамеченным.

В 1900 г законы Менделя были переоткрыты одновременно в трёх разных странах. Правоту Менделя. совершенно не зная о нём, подтвердили в Голландии Гуго де Фриз, Карл Эрих Корренс в Германии и Эрих фон Чермак в Австрии. Но термин „ген“ предложил датский исследователь Вильгельм Иоганнсен. В 1909 г он написал: „свойства организмов обуславливаются особыми, при известных обстоятельствах отделимыми друг от друга и в силу этого до известной степени самостоятельными единицами или элементами в половых клетках, которые мы называем генами. В настоящее время нельзя составить никакого определённого представления о природе генов; мы можем лишь довольствоваться тем, что подобные элементы действительно существуют. Не являются ли они химическими образованиями — об этом мы пока не знаем решительно ничего“. Что ж, в 1909 г Иоганнсен действительно не знал, что ещё в 1902 г английский врач А. Гаррод, исследуя родословные семей, пришёл к выводу, что алкаптонурия, болезнь, связанная с нарушением обмена веществ (нарушением „биохимии“, то есть), передаётся по наследству. К открытию английского врача судьба была не более благосклонна, чем к открытию католического монаха — оно было признано и оценено через 30 лет.

**Один ген — одна биохимическая реакция**

Когда мы говорим о том, что такое ген, будем различать его структуру и его функции. Структура — из чего и как организован ген, функция, — что и как он делает. Из работ Георга Менделя следует, что функция гена — определение морфологического признака. Сэр Арчибальд Гаррод, работая врачом в госпитале для больных детей и „демонстратором“ химической патологии в больнице Св. Варфоломея, обнаружил, что алкаптонурия вызывается повреждением одного рецессивного гена и что болезнь проявляется, согласно анализу родословных, когда мутантный аллель находится в гомозиготном состоянии. Отсюда был сделан вывод, что повреждение одного гена вызывает отсутствие одной биохимической реакции. А раз биохимические реакции катализируются ферментами то, предположил сэр Арчибальд, ген предопределяет наличие активного фермента. А отсюда рукой подать до вывода „один ген — один фермент“. Но он был сделан только через 30 лет. Детский врач А. Гаррод публиковал свои результаты в самом респектабельном, но медицинском журнале „Lancet“ — генетики его не читают.

**Один ген — один фермент**

В 1935 г Джордж Бидл и Борис Эфрусси изучали, как мутации в генах плодовых мушек дрозофил влияют на окраску их глаз и обнаружили, что различные мутации приводят к прекращению синтеза различных предшественников в пути биосинтеза глазного пигмента. Был сделан вывод: в норме гены обеспечивают наличие ферментов, осуществляющих биохимические реакции.

В 1940 г Дж. Бидл и Эдвард Татум использовали новый подход для изучения того, как гены обеспечивают метаболизм у более удобного объекта исследований — у микроскопического грибка Neurospora crassa. Ими были получены мутации, у которых отсутствовала активность того или иного фермента метаболизма. А это приводило к тому, что мутантный гриб был не способен сам синтезировать определённый метаболит (например, аминокислоту лейцин) и мог жить только тогда, когда лейцин был добавлен в питательную среду. Сформулированная Дж. Бидлом и Э. Татумом теория „один ген — один фермент“ — быстро получила широкое признание у генетиков, а сами они были награждены Нобелевской Премией.

Методы селекции так называемых „биохимических мутаций“, приводящих к нарушениям действия ферментов, обеспечивающих разные пути метаболизма, оказались очень плодотворными не только для науки, но и для практики. Сначала они привели к возникновению генетики и селекции промышленных микроорганизмов, а потом и к микробиологической промышленности, которая использует штаммы микроорганизмов, сверх продуцирующие такие стратегически важные вещества, как антибиотики, витамины, аминокислоты и др. . В основе принципов селекции и генной инженерии штаммов сверхпродуцентов лежит представление, что „один ген кодирует один фермент“. И хотя это представление отлично работает на практике, приносит многомиллионные прибыли и спасает миллионы жизней (антибиотики) — оно не является окончательным. Один ген — это не только один фермент.

Один ген — один полипептид

Фермент, как и любой другой белок, может состоять из нескольких субъединиц, объединённых в одно целое. Такие субъединицы, представляющие собой индивидуальные полипептидные цепи (полипептиды), могут объединяться за счёт т. н., не ковалентных белок-белковых взаимодействий, или за счёт ковалентных связей. Например, за счёт –S–S– связей между цистеинами, (аминокислотами, содержащими –SH группу), расположенными в разных полипептидах.

В 1953 г американский генетик Сеймур Бензер, исследуя особенности рекомбинации между двумя различными мутантами бактериофага Т4, одновременно заражавших кишечную палочку Escherichia coli, установил, что ген — это линейная структура, кодирующая синтез одного полипептида, а функциональный белок может состоять из нескольких полипептидов., при этом полипептиды в индивидуальном виде свою функцию, например, ферментативную, выполнять не могут за исключением случаев, когда функциональный белок (фермент, например) состоит из одной полипептидной цепи.

После обнаружения, что наследственные признаки определяются именно ДНК, сделанного в 1928 г Ф. Гриффитом и О. Эвери, и подтверждённого Ч. Мак-Леодом и М. Мак-Карти в 1944 г, а так же после открытия Джеймсом Уотсоном и Френсисом Криком в 1953 г того, что ДНК — это линейная двунитевая молекула, стало ясным, что

ген — это участок ДНК, в котором последовательность нуклеотидов определяет последовательность аминокислот в полипептидной цепи.

Участок ДНК, в котором расположена нуклеотидная последовательность, кодирующая единый полипетид иногда называется цистроном (от наименования т. н., „цис-транс теста“, с помощью которого С. Бензер сделал своё открытие). Кроме этого названия употребляется так же термин „структурный ген“, ибо ген кодирует структуру белковой молекулы.

Расшифровка генетического кода, сделанная в 1961 г Маршаллом Ниренбергом, блестяще подтвердила это положение. Но означает ли оно, что все гены кодируют полипептидные цепи? Оказалось, что нет. Во-первых, есть гены, которые кодируют не информационные РНК, а так называемые стабильные РНК, которые функциональны как таковые и матрицами для синтеза белков не являются. Это, прежде всего, гены рибосомальных ДНК и гены транспортных РНК. Рибосомальные РНК являются компонентами рибосом, а транспортные РНК принимают участие в трансляции, в синтезе полипептидных цепей, кодируемых структурными генами. Но есть ещё один класс генов, которые не кодируют ни полипептидных цепей, ни стабильных РНК.

Регуляторный ген — участок ДНК,

к которому присоединяется регуляторный белок

Именно фенотип мутации указывает на то, что может определять ген. Георг Мендель обнаружил, что мутации изменяют морфологию („морфологические мутации“), Арчибальд Гаррод, — что биохимическую реакцию („биохимические мутации“), Сеймур Бензер, — что структуру полипептида (мутации в структурных генах). Новый тип мутаций и новые функции генов в пятидесятые годы прошлого века обнаружили французские генетики Франсуа Жакоб, Жак Моно и Андрэ Львов. Оказалось, что у кишечной палочки одна мутация может приводить к исчезновению активности сразу нескольких генов. Для того, чтобы использовать в качестве пищи молочный сахар — лактозу E.coli применяет сразу три фермента. Была обнаружена мутация, которая находилась вне этих трёх генов, но приводила к тому, что активности всех трёх ферментов отсутствовали и такие мутантные клетки не могли расти на среде с лактозой. Выяснилось, что эти три гена транскрибируются ДНК зависимой РНК полимеразой без остановок. (ДНК зависимая РНК полимераза — фермент. осуществляющий синтез РНК на матрице ДНК, далее для краткости — РНК полимераза). В результате образуется единая длинная матричная РНК, которая кодирует все три соответствующих фермента. Поскольку одна мРНК кодирует несколько полипептидов (цистронов), она была названа полицистронной. Стало ясным, что мутации в той части ДНК, которая находится перед тремя структурными генами, приводит к невозможности транскрипции и. как результат, одна мутация приводила к отсутствию трёх ферментов.

Участки ДНК, необходимые для транскрипции структурных генов, к которым присоединяются либо РНК-полимераза (либо специальные белки, регулирующие, включающие или выключающие, транскрипцию) были названы регуляторными генами. В этих генах расположены особые последовательности нуклеотидов, с которыми связываются регуляторные белки, которые, в свою очередь, имеют для этого в составе своей молекулы специфические последовательности аминокислот. „узнающие особые последовательности“ нуклеотидов. Понимание молекулярных механизмов такого „белок-нуклеинового“ узнавания необходимо, в частности, для генной инженерии при конструировании организмов с заранее заданными механизмами регуляции генов. Наиболее распространённые типы регуляторных генов — это промоторы (к ним присоединяется РНК полимераза, чтобы начать транскрипцию), терминаторы (на таких участках РНК полимераза кончает транскрипцию), операторы (к ним присоединяются белки — репрессоры. выключающие работу РНК полимеразы), энхансеры и сайленсеры — участки ДНК, к которым присоединяются особые белки. изменяющие скорость транскрипции и тем самым, скорость синтеза соответствующих белков.

Оперон — несколько генов,

транскрибируемых как единая полистронная РНК

Если несколько ферментов участвуют в выполнении какой-то одной определённой задачи, например, последовательно катализируют цепь биохимических реакций, расщепляющих, например, лактозу или синтезирующих, например, лейцин или триптофан, то очевидно, что синтез каждого из этих ферментов должен быть скоординирован с синтезом других ферментов этого метаболического, иначе единый метаболический путь не будет работать нормально. У прокариот такая координация достигается тем, что гены таких ферментов расположены рядом (без „пробелов“, останавливающих транскрипцию) и транскрибируются они с единой регуляторной зоны (в которой расположены промотор и оператор) в виде полицистронной мРНК. Такая организация регуляторных и структурных генов названа опероном.

Чтобы оперон заработал РНК полимераза должна присоединиться к промотору, а репрессор, под действием определённого регуляторного сигнала, отсоединиться от оператора и, тем самым, открыть РНК полимеразе путь для транскрипции структурных генов. В геноме бактерий расположены тысячи оперонов, в которых, в свою очередь содержатся структурные гены, кодирующих белки (или стабильные РНК), участвующие в выполнении какой либо единой функции. Например, в геноме кишечной палочки содержится 2584 оперона, среди них 73% оперонов содержат только один цистрон, 16% — 2 цистрона, 4,6% — три цистрона, 6% — 4 и более цистронов. В 1965 г Ф. Жакоб, Ж. Моно и А. Львов за открытие оперонов были удостоены Нобелевской премии.

Наверное, предполагали генетики, гены и геномы эукариот устроены так же, как и гены прокариот. Но то что, обнаружилось — было ошеломляющим.

Мозаичная структура эукариотных генов

Оказалось, что внутри генов эукариот всегда есть участки, которые информационного смысла не имеют и не кодируют ни полипептидов, ни стабильных РНК. Эти участки назвали интронами. Термин интрон образован из английских слов — intervening zone — зона, „перемежающая, смысловую последовательность гена“. А те участки, которые смысл имеют, т. е кодируют, были названы экзонами. Экзон — от англ. expressing zone — экспрессируемая зона гена. Экзоны, как оказалось, кодируют так называемые части белковых молекул — модули или домены (компактные структуры), играющие важную роль в функционировании белковых молекул. Белковые молекулы состоят из нескольких модулей. Как правило, экзон кодирует участок полипептидной цепи длиной 30–40 аминокислот. А большинство интронов имеет длину от 40 до 125 нуклеотидных пар. Открытие мозаичной структуры эукариотных генов было сделано в 1977 г группами учёных, возглавляемых американскими исследователями Ричардом Робертсом и Филиппом Шарпом.

Но как работает такой ген? Состоящий из мозаики экзонов и интронов? А вот как:

Спеолдрласитьцйсиитбцнгорлю

Бессмысленное слово, не так ли? Но если из него удалить все интроны? Тогда получиться сплайсинг. Именно этот процесс и необходим для реализации функции эукариотного гена — все не смысловые (не кодирующие) участки должны быть удалены. Но не из гена, а из комплементарной ему РНК, которая должна быть синтезирована РНК полимеразой. Термин „сплайсинг“ в буквальном переводе с английского означает „соединение“. После вырезания интронов экзоны должны быть соединены.

Итак, чтобы эукариотный ген заработал необходимо создать (путём транскрипции) комплементарную РНК — копию мозаичного гена, состоящую из экзонов и интронов, из РНК интроны удалить, а экзоны объединить. Полученный окончательный транскрипт (теперь приобретший смысл) уже может быть использован для реализации его функции, для трансляции, например. Что важно, интроны из первичного транскрипта удаляются по очереди, а не все сразу. Вот так:

Спеолдрласитьцйсиитбцнгорлю

Спласитьцйсиитбцнгорлю + еолдр

Сплайсиитбцнгорлю + еолдр + ситьц

Сплайсингорлю + еолдр + ситьц + итбц

Сплайсинг + еолдр + ситьц + итбц + орлю

Таким образом, если в гене есть N интронов, то для сплайсинга необходимо N стадий вырезания интронов и сшивания экзонов. И если в какой-нибудь стадии произойдёт ошибка, например, при вырезании интронов будет вырезан один нуклеотид из экзона — это приведёт к тому, что ген свою функцию не выполнит и „наказанием“ за такую неточность будет или смерть, или тяжёлое нарушение жизнеспособности, что в ряду поколений кончится тем же летальным исходом.

Для чего же такая умопомрачительная, весьма дорогостоящая и опасная, в случае ошибок, сложность? А для

Аищалцуюофеьолтжиуекеруюнабюутхаипровбюуньцфыйооопс

В этом „гене“ 12 экзонов и 12 интронов. Если в 12 стадий удалить поочерёдно все интроны, то получится название особого типа сплайсинга:

Альтернативный + ища + цуюофе + ол + жиу + ке + ую + бюу + ха + про + бюу + ьцф + ооопс

И вот в чём смысл альтернативного сплайсинга: некоторые, чётко определённые экзоны вырезаются вместе с интронами. И тогда из

Аищалцуюофеьолтжиуекеруюнабюутхаипровбюуньцфыйооопс

Получится:

Альтернативный + ища + цуюофе + ол + жиу + ке + ую + бюу + ха + про + бюу + ьцф + ооопс

Альт + ища + цуюофе + ол + жиуекеруюнабюутхаипровбюуньцфыйооопс

нативный + Аищалцуюофеьолтжиуекерую + бюу + ха + про + бюу + ьцф + ооопс

наивный + Аищалцуюофеьолтжиуекерую + бюутха + про + бюу + ьцф + ооопс

левый + Аища + цуюофеьолтжиу + керуюнабюутхаипро + бюуньцф + ооопс

лев + Аища + цуюофеьолтжиу + керуюнабюутхаипро + бюуньцфыйооопс

В итоге, из одного, казалось бы, бессмысленного слова, получено шесть вполне осмысленных. А если это слово — ген?

Один ген — множество белков

Действительно, путь стыковки экзонов, принадлежащих одному гену, может быть множественным. Некоторые экзоны могут удаляться вместе с интронами. Такой альтернативный сплайсинг приводит к тому, что один и тот же ген может кодировать семейство структурно схожих, но функционально разных белков. На данный момент известное максимальное количество разных белков, которое может кодировать один ген, составляет около 40 000! (Сумма прописью — сорок тысяч). Например, ген дрозофилы, который кодирует один из белков рецептора, аксона за счёт альтернативного сплайсинга может приводить к образованию 38016 различных информационных РНК. Этот ген содержит 95 альтернативных экзонов. Но все ли гены экспрессируются за счёт альтернативного сплайсинга? Согласно текущим знаниям, по крайней мере, 74% генов человека работает с помощью альтернативного сплайсинга!

Теперь самое время задаться вопросом: что такое ген?

Ген (эукариотный) это длинная и преимущественно случайная, не кодирующая последовательность нуклеотидов, в которой расположены участки (экзоны), способные после вырезания из транскрипта этого гена и их объединения в строго определённой очерёдности, кодировать определённую функцию.

Особо отметим, что при альтернативном сплайсинге порядок расположения экзонов не нарушается. В окончательном варианте сплайсированной РНК некоторые экзоны могут присутствовать или отсутствовать, но местами они не меняются. Например, в окончательно сплайсированной РНК экзоны 1–2–3–4–5–6 могут быть в последовательности 2–4–6, но не в последовательностях 4–2–6 или 6–4–2. Таким образом, из одного и того же транскрипта гена, используя разные варианты распознавания, вырезания и соединения разных экзонов можно получить множество разных изоформ белков, у которых будут общими некоторые аминокислотные последовательности, но которые будут отличаться по своим функциональным свойствам. И то, что сначала наивно полагали бессмысленным — интроны, перемежающие гены, на самом деле оказалось весьма эффективным и экономичным способом кодирования множества смыслов за счёт ограниченного числа знаков. Правда, это привело к значительному усложнению правил обнаружения этих смыслов. Путь альтернативного сплайсинга в большой степени определяется регуляторными сигналами клетки, характеризующими её состояние. В ответ на изменение физиологической ситуации из одного и того же гена реализуются разные функции.

Весьма принципиально, что при эволюционном усложнении организмов среднее количество интронов, приходящихся на один ген, возрастает. На основе статистического анализа сделаны выводы, что размер генома коррелирует с общей длиной интронов, содержащихся в гене данного вида; интроны беспозвоночных короче, чем интроны генов человека, а интроны дрожжей короче, чем интроны беспозвоночных. По мере усложнения организмов увеличивается и длина интронов. В общем, в гене суммарная длина интронов может превосходить суммарную длину экзонов в десятки и сотни раз.

Если секвенирование (определение нуклетотидной последовательности, от англ. sequence — последовательность) эукариотных генов привело к ошеломляющему открытию их мозаичной структуры, то массовое секвенирование целых геномов разных организмов привело к результатам просто изумляющим. У мыши, человека у рыбы фугу (рыба шар) количество генов практически одинаково — 30000 — 40000. Что же тогда определяет эволюционную сложность?

Более того, если сравнивать между собой кодирующие последовательности (экзоны) в геномах мыши и человека, то окажется, что они идентичны на 99%! Почему же мы так не похожи на мышей?

Может быть и потому, что несмотря на то, что наши гены похожи на мышиные, у нас альтернативный сплайсинг идёт или по другому пути, или более множественный. Или и то и другое одновременно. Ведь не зря же по мере прогрессивной эволюции среднее количество интронов (а значит, и экзонов), приходящихся на один ген, возрастает? Ведь это расширяет спектр белков, потенциально кодируемых одним геном. Не так ли? И в результате из-за разного альтернативного сплайсинга из почти одних тех же генов получается или мышь, или шимпанзе, или тот, кто в данный момент читает эти строки.

Альтернативный сплайсинг предоставляет эволюции практически безграничные возможности. Материал эволюции — генетическое разнообразие, а двигатель — естественный отбор. А ведь альтернативный сплайсинг приводит к такому разнообразию белков, что… Посудите сами. Комбинация только из трёх генов, каждый из которых может кодировать только 1 000 вариантов белков, даёт 1 000 000 000 возможностей для естественного отбора (1 млрд изоформ трёх белков). А если таких генов 1000? А если 10000?

Каковы молекулярные механизмы эволюции? Как происходит видообразование? Поисками ответов на эти вопросы занимаются такие новые научные направления, как молекулярная генетика развития и эволюционная биология развития (по-английски сокращённое название этой волнующей науки звучит очень романтично — evodevo, от evolutionary developmental biology). Большие успехи в выяснении молекулярных механизмов эволюции делает сравнительная геномика. которая с помощью мощнейших методов компьютерного анализа анализирует и сравнивает гены и геномы разных организмов.

Итак, открытие принципов организации и экспрессии эукариотных генов поставило перед наукой новые проблемы и новые задачи. Это нормально. А что дали эти открытия для практики?

Не приведёт ли альтернативный сплайсинг к банкротству геномных фирм?

Проект „Геном человека“ — секвенирование генома Homo sapiens, содержащего около 3 млрд нуклеотидов, по свой научной значимости и амбициозности сравнивают с программой пилотируемых полётов на Луну „Аполлон“. Стоимость этих программ соизмерима. Миллиарды долларов. Но инициаторы проекта „Геном человека“, кроме научных целей имели ещё и грандиозные планы практического использования генетической информации, которая должна быть получена в результате его выполнения. А там, где практическое использование — там и коммерческий интерес. Предполагалось, что информация о строении генома человека будет полезна для ранней молекулярной диагностики наследственных заболевания и для их лечения путём, так называемой, заместительной генетической терапии (дефектный ген замещается на нормальный). Планировалась, что информация о геноме человека приведёт к революции в разработке нового поколения лекарственных препаратов, создаваемых на основе знаний о нарушениях функционирования определённых белков. Если знать, как кодируется и экспрессируется конкретный белок, приводящий к первопричине заболевания, то будет возможным, как полагалось, создать такие искусственные молекулы, которые будут направленно, а не в слепую, исправлять патологические процессы уже на молекулярном уровне. А это, как прогнозировалось, может принести многомиллиардные прибыли. Но сначала были необходимы многомиллионные инвестиции. И они были сделаны. Были созданы коммерческие фирмы, которые проводили секвенирование генома человека. Информация о нуклеотидных последовательностях генов, которая, как полагалась, может привести к разработке новых методов диагностики и терапии, патентовалась.

Но разве нуклеотидная последовательность эукариотного гена даёт однозначную информацию о том, какой именно белок он кодирует? Нет. Всё зависит от пути альтернативного сплайсинга. А этих путей для одного гена может быть почти 40000. Выделить и охарактеризовать все 40000 белков и установить, из-за какой именно изоформы белка происходит патология задача практически не решаемая. По крайней мере, пока. И знаменитый дешифровальщик генома человека Крейг Вентер, бывший директор геномной фирмы Celera Genomics, круто меняет направление своего бизнеса. Теперь он занимается массовой дешифровкой геномов бактерий. Его научный корабль бороздит волны Саргассова моря (в районе Бермудского треугольника), научные сотрудники отлавливают морских микробов и секвенируют их геномы прямо на корабле. Суммарная длина всех уже просеквенированных нуклеотидных последовательностей — более 1 млрд нуклеотидов, она относится к 1800 видам бактерий, из которых 148 видов ранее известными не были. Цель проекта — поиск новых генов, имеющих прикладное значение. Разумеется, работать с прокариотными генами несоизмеримо проще: один ген — один белок. Если Крэйг Вентер уже подал в отставку с должности директора, то директор другой крупнейшей геномной фирмы Human Genome Science — Уильям Хэзелтайн пока только объявил, что собирается „перейти на другую работу“.

Разумеется, совершенно непредвиденная сложность эукариотных генов и механизмов их экспрессии не должны приводить нас в отчаяние и разочарование. Наоборот, чем сложнее задачи, тем более интересными и неожиданными будут их решения. А они обязательно будут.

Что ж, изобретя мозаичный ген и альтернативный сплайсинг, эволюция преподнесла своему результату — Человеку разумному ошеломляющий сюрприз.

Но благодаря этому изобретению, благодаря альтернативному сплайсингу из мозаики экзонов и интронов выросло чарующее Древо Прогрессивной Эволюции Жизни.

**Список литературы**

1. Жимулев И.Ф., Общая и молекулярная генетика; Учеб. пособие.-2-е изд., Новосибирск: Сиб.унив. изд-во, 2003, 479 с.; ил.

2. Modrek B, Lee C., A genomic view of alternative splicing. Nature Reviews, 2002, v.30, №1, 13-19.

3. Silverman P H, Rethinking Genetic Determinism: With only 30 000 genes what is it that makes humans human? The Scientist, 2004, May 24, v.18, № 10.