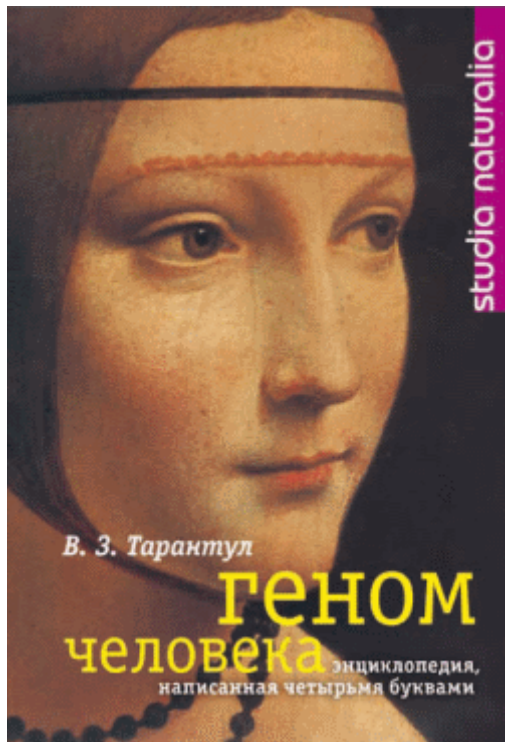




В. З. Тарантул

ГЕНОМ человека

энциклопедия,
написанная четырьмя буквами

studio naturalia

Вячеслав Залманович Тарантул

Геном человека: Энциклопедия, написанная четырьмя буквами

*Текст предоставлен правообладателем
http://www.litres.ru/pages/biblio_book/?art=180721*

*Геном человека: Энциклопедия, написанная четырьмя буквами. : Языки славянской культуры;
Москва, 2003
ISBN 5-94457-108-X*

Аннотация

26 июня 2000 года произошло очень знаменательное событие, о котором много писалось в прессе. На пресс-конференции с участием президента США и премьер-министра Великобритании представители двух исследовательских коллективов – международного консорциума Human Genome Project и компании Celera Genomics – торжественно объявили о том, что в результате многолетних усилий большого числа ученых и огромного финансирования закончена расшифровка генома человека (точнее, определена полная структура ДНК). Был успешно реализован не имеющий аналогов по масштабам Проект века. Что же человечество узнало и приобрело в результате этого? Как эти новые знания помогают человеку в решении его многих повседневных проблем? Какие тайны хранит человеческий геном?

Обо всем этом и идет речь в настоящей монографии, написанной в популярной форме и предназначенной для самого широкого круга читателей: для школьников старших классов, школьных учителей, студентов и преподавателей вузов и вообще для всех образованных людей, желающих больше знать о самих себе.

Содержание

НЕСКОЛЬКО ВСТУПИТЕЛЬНЫХ СЛОВ	4
ОБЩЕЕ ВВЕДЕНИЕ	6
ЧАСТЬ I	10
ЧТО ТАКОЕ ГЕНОМ?	10
НЕМНОГО ИСТОРИИ, НЕМНОГО ТЕРМИНОЛОГИИ	12
ДНК – МОЛЕКУЛЯРНАЯ ОСНОВА ГЕНОМА	15
Тонкое устройство ДНК	15
Генетическая грамматика	21
Размножение ДНК	23
Перезапись генетического текста и перевод в белковый текст	24
ДЛИННАЯ ИСТОРИЯ СО СЧАСТЛИВЫМ КОНЦОМ	28
Хромосомы дают первые сведения о структуре генома	28
Как двухметровая молекула уместается в микроскопическом ядре?	35
Анализ суммарной ДНК – новые сведения о структуре генома человека	36
Новый этап – анализ индивидуальных элементов ДНК (генная инженерия)	39
Конечная цель – определение полной последовательности нуклеотидов в ДНК человека	43
Секвенирование простых геномов	45
ТАЙНА ТАЙН ЗА 6 МИЛЛИАРДОВ ДОЛЛАРОВ	50
ОСНОВНЫЕ ВЕХИ ГЕНЕТИКИ И ГЕНОМИКИ	54
ГЕНЫ – ОСНОВНОЙ ТЕКСТ ГЕНОМА	60
«Жи, ши – пиши через и»	61
Знаки препинания (Генетическая пунктуация)	62
Гены построены из кусков	64
Перекрывание генных текстов	68
Не все гены кодируют белки	69
Ген в гене (генная матрешка)	70
Генные семейства	71
Новые предложения из фрагментов старых текстов	73
Сколько же всего генов у человека?	74
ОСНОВНАЯ ЧАСТЬ ГЕНОМА – ТЕРРА ИНКОГНИТА	77
ДНКовый текст человека перенасыщен повторами	78
Перевертыши (обращенные повторы)	78
Простые tandemные повторы (сателлиты)	80
Конец ознакомительного фрагмента.	82

Вячеслав Залманович Тарантул **Геном человека: Энциклопедия,** **написанная четырьмя буквами**

НЕСКОЛЬКО ВСТУПИТЕЛЬНЫХ СЛОВ

Посвящаю жене, дочери и внучке

Любому человеку – от младшего школьника до пожилого пенсионера – интересно знать, как устроен он сам, почему он болеет, как всегда быть здоровым и что для этого нужно делать. Человеку всегда были близки науки о жизни, биология – одна из самых древних наук. Однако уже в древности стало понятно, что любое живое существо устроено несомненно сложнее, чем любой неодушевленный предмет, любое вещество, самое сложное, проще любого существа, самого примитивного. Сложность познания жизни во всех её многообразных проявлениях связана с несколькими существенными обстоятельствами. Во-первых, жизнь невероятно многообразна. Что общего можно усмотреть между бактерией, не видимой простым глазом, и, на другом конце цепочки, – китом, или слоном, или, наконец, человеком? Во-вторых, живые существа непрерывно меняются, одни виды появляются, другие исчезают. Когда жили динозавры, не было людей, а когда размножились люди, исчезли мамонты и сотни других видов животных и растений. В-третьих, жизнь необычайно чувствительна к окружающей среде, нежна, очень легко ранима. Поместите любое животное на 5–10 минут в бескислородную среду, и живой организм станет мертвым. Наконец, живые организмы необычайно сложно устроены – в их состав входит практически вся таблица Менделеева, тогда как, для сравнения, любое вещество, даже сложное органическое, не использует для построения своих молекул более 2-х десятков элементов менделеевской таблицы.

Одной из самых больших загадок биологии всегда была проблема наследования: почему мышь рождает мышь, а не крокодила? Почему дети похожи на родителей? Почему некоторые болезни передаются у людей и животных по наследству?

Старая, классическая, дарвиновская биология дала качественные, приблизительные, описательные ответы на часть этих вопросов. Биология XX века совершила гигантский скачок в понимании природы жизни, её молекулярных, химических основ. Венцом, вершиной этих титанических усилий всего мирового научного сообщества стала расшифровка строения наследственного аппарата человека, который представляет собой совокупность генов и внегенных участков ДНК – главной молекулы жизни. Коротко всю ДНК, в которой записана наследственная информация о данном организме, называют «геномом».

Раскрытие вчерне строения генома (февраль 2001 года) – событие столь же ключевое для истории человечества, как открытие электричества, атомной энергии, изобретение радио и телефонной связи и компьютеров. Человек, наконец, получил химическую, молекулярную основу для того, чтобы познать самого себя.

XXI век, несомненно, будет веком биологии. Это признают не только биологи, но и ученые других, далеких от биологии специальностей. В этом веке мы узнаем, что такое память, творческие способности, мы научимся защищать себя от многих болезней и лечить заболевания, которые мы сейчас называем неизлечимыми (рак, например).

Эта книга рассказывает о геноме. Каждый, кому интересно узнать о себе что-то новое, прочтет её с интересом и пользой для себя. Она написана достаточно просто, чтобы одолеть её до конца, хотя, конечно, неподготовленному читателю, который давно забыл школьный

курс биологии, придется приложить некоторые усилия. Но они будут вознаграждены тем, что читатель узнает для себя массу нового, чего нет пока еще ни в одном учебнике, узнает многое из того, что некоторые биологи далеко еще сами не освоили.

Эта книга – слепок новой биологии, она объясняет (и достаточно понятно), почему наша жизнь все больше и больше будет зависеть от того, как мы сумеем распорядиться тем информационным богатством, которое нам подарил прошлый век. Сейчас главная проблема состоит в том, как превратить гигантский объем новых знаний о геноме в новое качество нашей жизни – здоровье, долголетие, победу над болезнями, использование всех научных способностей, заложенных в нас природой и родителями.

Не поленитесь дочитать книгу до конца. Дочитав последнюю страницу, вы почувствуете себя гордым за человечество, которое, наконец, начинает познавать самое себя. Мы на полпути к себе. Вы сможете сами участвовать в освоении второй половины пути – для многих читателей это может стать делом жизни.

Прочтя эту книгу, вы очень многое узнаете о себе и окружающих вас существах. Разве это не интересно?

Академик Л. Киселев

ОБЩЕЕ ВВЕДЕНИЕ

*Мир не просто удивительнее, чем
мы себе представляем, —
он удивительнее,
чем мы можем себе представить.*

Джон Бердон Холдейн

Человек – вершина эволюции живой природы на Земле. Такое утверждение общепризнанно, и оно обычно ни у кого не вызывает сомнения. Знаменитый молекулярный биолог, лауреат Нобелевской премии Дж. Уотсон писал: «Легко видеть в человеке существо совершенно особое, отличное от всех прочих живых существ. Лишь он изобрел сложные языки, дающие ему средство для обмена идеями и эмоциями. Он создал великие цивилизации, изменившие окружающий мир так, как не могло бы его изменить никакое другое животное. Поэтому человек всегда был склонен считать, что он занимает в природе не совсем обычное место». Но эта мысль принадлежит самому человеку, и поэтому в ней есть элемент субъективности. Безусловно, в том, что касается мыслительных функций, данное утверждение во многом правильно. Человек действительно царствует на Земле. Он уже давно в значительной мере встал над природой, перестал быть прямым объектом ее естественного отбора. Достижения человеческой цивилизации в целом огромны, они изменили облик всей нашей планеты, дошла даже очередь до других планет.

Однако царствование человека на Земле и в космосе не безгранично. До полного совершенства ему – царю природы – еще очень далеко. Существует множество неподвластных человеку проблем даже на планете Земля. Катаклизмы и стихии порой ставят человечество в тупик. Да и мы сами по ряду параметров сильно уступаем другим представителям животного мира, обитающим вокруг нас. Во многих отношениях мы чересчур слабы, ранимы. Это проявляется в нашей сильной уязвимости к воздействиям различных физических и химических факторов окружающей среды, в слабой устойчивости к действию различных патогенов, низкой плодовитости, относительно короткой продолжительности полноценной жизни. В отличие от других животных нашему зрению доступен лишь ничтожно малый участок электромагнитного спектра, а обоняние не способно воспринять и сотой доли окружающих нас запахов. Некоторые животные (рыбы, птицы) способны различать сигналы, которые человек вообще не воспринимает без специальных приборов. Гомер даже говорил: «Меж всевозможных существ, которые дышат и ходят, здесь, на нашей земле, человек наиболее жалок». Пожалуй, самую уничижительную оценку человеку дал один американец, который подсчитал, что в организме человека, весящего 70 кг, содержится 45,5 кг кислорода, 12,6 кг углерода, 7 кг водорода, 2,1 кг азота, 1,4 кг кальция, 700 г фосфора, а всех остальных элементов, вместе взятых (в основном калия, серы, натрия, хлора, магния, железа и цинка), – около 700 г. И всё это «богатство» стоит всего 1 доллар 98 центов. Но это уже крайность.

Куда серьезнее то, что человечество страдает от многочисленных наследственных патологий, болезней сердечно-сосудистой системы, злокачественных заболеваний, иммунодефицита и травматизма. Мы по-прежнему подвержены разнообразным вирусным инфекциям, идет широкое распространение СПИДа, гепатита и других грозных инфекционных заболеваний. Кроме того, человечество обременено и многочисленными пороками, которые сопровождают всю его историю. Наркомания, алкоголизм, эпилепсия и шизофрения – все это сопутствует пребыванию человека на Земле. В ходе всей истории человечество вело многочисленные войны, в которых были уничтожены сотни миллионов его представителей, и даже в XXI веке существует терроризм, растет преступность. Исчезают одни «сверх-человеки»,

появляются другие, даже в наше время, казалось бы, далекое от времени дикарей, отдельные страны объявляют себя «сверхдержавами» и начинают диктовать миру свои законы жизни и поведения. Нарушаются все возможные моральные и библейские принципы. Так что «царь природы» далеко не во всех отношениях выглядит совершенным, ему требуются значительная помощь и защита.

Но, несмотря на все «минусы», человеческая популяция продолжает неуклонно расти и процветать, в ней сохраняется и превалирует жизнеутверждающее и творческое начало. Достижения человека на планете Земля огромны. В чем же причина всего этого? В значительной мере это результат реализации той информации, которая заложена природой в нашем генетическом аппарате.

Известно, что мы все в той или иной мере эгоисты. Человека больше всего на свете интересует он сам. Все, что имеет к нему отношение, – предмет наивысшего внимания. Со временем пришло понимание того, что все упирается в биологию человека, а вся биология человека упирается в геном. Козьма Прутков говорил: зри в корень. В организме человека главный «корень» – это и есть геном. Стало ясно, что основной путь к исправлению многочисленных «слабостей» человека лежит через детальное изучение строения и работы его наследственной информации, содержащейся в геноме. Но долгое время это не удавалось сделать, поскольку ученые не располагали соответствующими подходами и методами. И как только к концу XX века появились реальные научные, методические и технические возможности определять структуры больших по размерам молекул, кодирующих наследственную информацию (геномов), сразу после расшифровки ДНК простых организмов (вирусы, бактерии, дрожжи, червь) одним из первых сложных организмов, чья ДНК была расшифрована, стал человек. Знаменательно, что даже структура генома такого классического генетического объекта, как дрозофила, который исследуется существенно дольше и который значительно проще устроен, была определена не раньше, а почти одновременно с геномом человека.

Успехи в изучении генома человека стали возможными в первую очередь благодаря усилиям международного проекта «Геном человека», который возник и реализовался в результате напряженной работы большого сообщества ученых. Инициатором этого движения с самого начала (конец 80–х годов) стал лауреат Нобелевской премии знаменитый ученый Дж. Уотсон. В России таким энтузиастом был академик А. А. Баев (1904—1994). В 1988 году он обратился к М. С. Горбачеву с письмом, в котором предложил организовать государственный научный проект по изучению генома человека. Как в нашей стране, так и за рубежом эта идея первоначально была встречена рядом ученых и общественных деятелей весьма критически. Им казалось, что задача просто невыполнима, а труд будет сизифов. Но время шло, и довольно скоро научное сообщество во всем мире стало обсуждать ее всерьез. Ученые смогли убедить свои правительства и бизнесменов, что определить полную структуру ДНК человека – важнейшая и насущнейшая задача для всего человечества. В результате этих усилий в 1989—1990 гг. в США и в СССР, а затем и в ряде других стран начали функционировать соответствующие научные программы. Чуть позднее возникла Международная организация по изучению генома человека (Human Genome Organization, сокращенно HUGO).

В самом начале задача казалась сверхсложной, а ее осуществление сверхдорогим. Но научная мысль, интеллект и энтузиазм множества исследователей из разных стран, помноженные на огромное финансирование, позволили в конечном итоге решить эту грандиозную задачу. На реализацию проекта были выделены огромные средства. Крупный вклад в решение этой проблемы (как финансовый, так и научный) был сделан в дальнейшем частной американской компанией «Celera Genomics». В сумме, по имеющимся оценкам, на определение полной структуры генома человека было затрачено свыше 6 млрд. долларов. Проект «Геном человека» стал одним из самых дорогостоящих из всех, когда-либо создаваемых на

земле. Для сравнения: весь годовой бюджет Российской Федерации на 2000 год был равен 60 млрд. долларов. Грандиозная работа проводилась одновременно в двух десятках стран мира, включая США, Англию, Францию, Германию, Японию, Китай и Россию, и была осуществлена за срок чуть более 10 лет. Основных инициаторов и исполнителей проекта «Геном человека» (Джеймса Уотсона, Френсиса Коллинза, Крега Вен-тера) уже называли «рыцарями двойной спирали».

Как сказал профессор Майкл Декстер (руководитель британской части проекта), определение полной структуры генома человека представляет собой одно из величайших достижений человечества, которое превосходит по своему значению высадку человека на Луне.

Правда, едва ли раскрытие тайн структуры генома человека станет панацеей для человечества. Ажиотаж вокруг проекта «Геном человека» создает несколько искаженную картину, будто бы знание молекулярной организации генома уже сейчас решает все проблемы человека. На этом фоне возникают даже рассуждения об узко прагматическом «геноцентризме», который формирует у общества лишь кратковременную иллюзию научного успеха. Но серьезные ученые понимают, что это всего лишь первый, начальный этап. Он уже получил и свое название – «структурный». Когда-то один из отцов молекулярной биологии, нобелевский лауреат Ф. Крик говорил: «Не знаешь функции – изучай структуру». Теперь неизбежно должен последовать второй, и, скорее всего, даже более важный этап – «функциональный» – изучение функции генов. Эрик Ландер, возглавлявший исследование генома человека в Массачусетском технологическом институте, сказал в 2000 году: «В июне люди, вероятно, полагали, что на этом исследование закончится. Но я не знаю, поняли ли люди, что мы обнаружили в геноме самую величайшую в мире книгу истории. Мы постоянно просыпаемся ночью для того, чтобы продолжить читать истории, которые заключены в геноме. Это так здорово!»

Однако и преуменьшать значение состоявшегося «структурного» этапа в изучении человеческого генома нельзя. В настоящее время биология в целом (как наука о жизни) встает на принципиально новые, геномные, рельсы. Реализация проекта «Геном человека», подобно любому другому крупному и хорошо организованному проекту, одновременно способствовала прогрессу молекулярной биологии и генетики в целом. Были в значительной мере усовершенствованы методическая и приборная база науки, мощный импульс получили такие, по сути дела, вновь созданные направления, как геномика (изучение целых геномов), протеомика (изучение совокупности белков) и ряд других. С самого начала работ по геномному проекту мир договорился об открытости, доступности всей получаемой информации для его участников, независимо от их вклада и государственной принадлежности. Сейчас существуют десятки мощных баз данных, доступных любому пользователю, в которых аккумулирована гигантская информация о структуре не только генома человека, но и геномов многих других организмов.

Только благодаря вновь полученной конкретной информации о генетическом аппарате человека становится возможным исправлять генетические дефекты, лечить различные заболевания, корректировать психику, бороться с вирусными инфекциями, разрабатывать совершенные лекарственные средства и выявлять мишени, на которые они направлены, создавать новые поколения диагностических и профилактических средств и в результате существенно продлить полноценную жизнь человека. По этой причине известный генетик Джереми Рифкин сказал: «Кому принадлежат гены, тому принадлежит 21-е столетие». А полный текст, записанный в геноме человека, иногда называют даже несколько возвышенно Энциклопедией жизни.

Итак, на рубеже третьего тысячелетия человечество наконец-то раскрыло одну из важнейших для него тайн – структуру генома *Homo sapiens* (человека разумного). Произошло то, что еще в 1957 году предсказывал российский академик Владимир Александрович

Энгельс-гардт: «Не нужно быть беспочвенным оптимистом, чтобы верить, что через пятьдесят лет биологический код – химическая зашифровка наследственных свойств – будет расшифрован и прочитан. С этого момента человек станет полным властелином живой материи».

Знаменательно, что структура генома человека вчера была определена к середине 2000 г., когда исполнилось 100 лет с официального начала генетики как науки (переоткрытие менделеевских фундаментальных законов наследственности произошло в 1900 г.). Полное завершение этой работы намечено к 2003 году, когда будет опубликован окончательный текст, записанный в геноме. Кстати, 2003 год юбилейный – 50 лет прошло с тех пор, как Дж. Уотсон и Ф. Крик открыли структуру ДНК, которая представляет собой двойную спираль.

Нельзя преувеличивать успехи российских ученых в завершившемся международном геномном проекте, но не стоит и преуменьшать их. В первую очередь следует вспомнить выдающегося ученого академика А. А. Баева, который стоял у истоков этого проекта в СССР. Заметный вклад в изучение генома человека был сделан российскими учеными в различных областях исследования, особенно в таких, как функциональная геномика, этногеномика и биоинформатика. В настоящее время около 70 ученых России являются членами всемирной организации HUGO, избрание в которую осуществляется тайным международным голосованием на основании оценки публикаций кандидатов. Все эти годы российская программа «Геном человека» в меру сил финансировалась Минпромнауки РФ и поддерживалась Президиумом РАН. Конечно, жаль, что этих средств было явно маловато на фоне общего финансирования геномной программы в мире, но, как это часто бывало в прошлом, мозги российских ученых в известной мере компенсировали недостаток материальных средств.

* * *

Успешное и широко разрекламированное определение полной структуры генома человека послужило основным поводом для написания этой книги. Набожный руководитель международной программы «Геном человека» Ф. Коллинз даже назвал текст, записанный в геноме человека, «языком Бога». Однако значение произошедшего события до сих пор не в полной мере понято широкой общественностью. Академик А. А. Баев еще в середине 90-х годов прошлого столетия отмечал с сожалением, что развитие современной биологии сделало ее малодоступной для неспециалистов. В этой связи, несмотря на гигантские научные успехи, мистика и суеверия не только продолжают существовать в современном обществе, но даже прогрессируют. Хотелось бы надеяться, что эта книга хотя бы в небольшой мере поможет закрыть имеющийся пробел в понимании сути и значения последних достижений ученых. Она предназначена не для узких специалистов, которые и так понимают всю суть проблемы, а для самой широкой аудитории. Но тема, которой посвящена книга, весьма специальная. Поэтому в самом начале нельзя было не остановиться хотя бы кратко на истории вопроса, необходимо также дать объяснение хотя бы части той специальной терминологии, которая используется на своей «кухне» генетиками. Жан-Жак Руссо писал: «Мудрецы, которые хотят говорить с простым народом своим, а не его языком, никогда не смогут стать ему понятными. Однако есть множество разного рода понятий, которые невозможно перевести на язык народа». Попробуем и мы для начала сделать такой перевод, отчетливо понимая при этом, что, согласно известному закону Мей-ера, «усложнять – просто, упрощать – сложно».

ЧАСТЬ I

СТРУКТУРА ГЕНОМА ЧЕЛОВЕКА

ЧТО ТАКОЕ ГЕНОМ?

Вопросы вечны, ответы обусловлены временем.

Е. Чаргафф

В диалоге с жизнью важен не ее вопрос, а наш ответ.

М. И. Цветаева

С самого начала определимся, что мы здесь будем подразумевать под словом **геном**. Сам этот термин впервые был предложен в 1920 году немецким генетиком Г. Винклером. Тогда уже существовал другой научный термин – **генотип**, введенный в арсенал генетиков В. Иогансеном еще в 1909 году, под которым подразумевалась совокупность всех наследственных задатков данной конкретной клетки или данного конкретного организма. Впоследствии Иогансен сам с удивлением говорил, что его «словечко» неожиданно материализовалось в возникшей позднее хромосомной теории Т. Моргана. Но вот появился новый термин – геном. В отличие от генотипа этот термин должен был стать *характеристикой целого вида организмов, а не конкретной особи*. И это стало новым этапом в развитии генетики.

В биологическом словаре понятие **геном** определяется как совокупность генов, характерных для гаплоидного (одинарного) набора хромосом данного вида организмов. Такая формулировка звучит не совсем понятно для неспециалиста, а главное, она неточна в современном понимании этого слова. Основу генома составляет молекула дезоксирибонуклеиновой кислоты, хорошо известная в сокращенном виде как ДНК. Ведь все геномы (ДНК) содержат по крайней мере два вида информации: кодированная информация о структуре молекул–посредников (так называемых РНК) и белка (эта информация содержится в генах), а также инструкции, которые определяют время и место проявления этой информации при развитии и дальнейшей жизнедеятельности организма (эта информация в основном расположена в межгенных участках, хотя частично и в самих генах). Сами гены занимают очень небольшую часть генома, но при этом составляют его основу. Информация, записанная в генах, – это своего рода «инструкция» для изготовления белков, главных строительных кирпичиков нашего тела. «На плечах» генов лежит огромная ответственность за то, как будет выглядеть и работать каждая клетка и организм в целом. Они управляют нашей жизнью от момента зачатия до самого последнего вдоха, без них не функционирует ни один орган, не течет кровь, не бьется сердце, не работают печень и мозг.

Однако для полной характеристики генома недостаточно заложенной в нем информации о структуре белков. Нужны еще данные об элементах генетического аппарата, которые принимают участие в работе (**экспрессии**) генов, регулируют их проявление на разных этапах развития и в разных жизненных ситуациях.

Но даже и этого мало для полного определения генома. Ведь в геноме присутствуют также элементы, способствующие его самовоспроизведению (**репликации**), компактной упаковке ДНК в ядре и еще какие-то непонятные пока еще участки, иногда называемые «эгоистичными» (то есть как бы служащими только для самих себя). По всем этим причинам сегодня, когда речь идет о геноме, обычно имеют в виду всю совокупность последовательностей ДНК, представленных в хромосомах ядер клеток определенного вида организмов,

включая, конечно, и гены. В этой книге мы будем подразумевать именно такое определение. Вместе с тем следует помнить, что в некоторых других структурах (органеллах) клетки также присутствует генетическая информация, необходимая для функционирования организмов. В частности, у всех животных организмов, в том числе и у человека, имеется еще и митохондриальный геном, то есть молекулы ДНК, присутствующие в таких внутриклеточных структурах, как митохондрии, и содержащие ряд так называемых митохондриальных генов. Митохондриальный геном человека очень небольшой по сравнению с ядерным геномом, расположенным в хромосомах, но, тем не менее, его вклад в клеточный метаболизм весьма существенен.

Понятно, что знание одной лишь структуры ДНК вовсе не достаточно для полного описания наследственной системы клетки. Этому выводу в литературе дана следующая аналогия: сведения о числе и форме кирпичей не могут раскрыть замысла готического собора и хода его постройки. В более широком смысле наследственную систему клетки составляют не только структура ДНК, но и другие ее компоненты, совокупность которых и факторы окружающей среды определяют, как геном будет работать, как пойдет ход индивидуального развития и как возникший организм будет жить потом.

НЕМНОГО ИСТОРИИ, НЕМНОГО ТЕРМИНОЛОГИИ

*По-видимому, на свете нет ничего,
что не могло бы случиться.*

М. Твен

Некоторым читателям может показаться скучным и малоинтересным содержание этого раздела книги. Но хотелось бы таким читателям напомнить, что даже в захватывающем детективе бывают вроде бы малоинтересные страницы, которые, однако, при дальнейшем изложении часто оказываются важны для понимания происходящего.

Сначала несколько слов о генетике. Без сомнения, наивысшим достижением в биологии ушедшего двадцатого века было появление и развитие этой науки, а также раскрытие с ее помощью молекулярной природы наследственности. И теперь уже мало верится, что было время, когда в нашей стране генетику называли «продажной девкой империализма», а ученых-генетиков – презренным тогда словосочетанием «вейсманисты–морганисты» (производное от имен немца Августа Вейсмана (1834—1914) и нобелевского лауреата американца Томаса Ханта Моргана (1866—1945), которые сформулировали представления о дискретности генов, их локализации в хромосомах, обосновали хромосомную теорию наследственности). Сейчас во всем мире генетика не только получила «путевку в жизнь», но пребывает на законных основаниях в славе и почете.

В отличие от многих других биологических наук генетика с момента своего возникновения стремилась быть точной наукой. И вся история генетики – это история создания и использования в эксперименте все более и более точных методов и подходов, что сближает ее с такими точными науками, как физика, химия и математика.

А начало всему было положено чешским монахом Грегором Менделем, который в 1865 году опубликовал свой фундаментальный труд

с математическими расчетами, указывающими на существование неких абстрактных дискретных частиц, передающих наследственные свойства («частицы наследственности»), названных позднее **генами**. Эта феноменальная работа Менделя, осуществленная на горохе, не произвела особого эффекта на его современников и была забыта вплоть до 1900 года (научная мысль в то время еще не созрела для ее восприятия). Сам Мендель после неудачных попыток получить аналогичные результаты при скрещивании других растений прекратил опыты и до конца жизни занимался садоводством, пчеловодством и метеорологическими наблюдениями.

Лишь спустя 35 лет произошло то, что и должно было случиться: законы Менделя были переоткрыты независимо и одновременно тремя разными исследователями (Г. де Фризом, Э. Чермаком и К. Корренсом), после чего и начала интенсивно развиваться наука, получившая позднее название **генетика**. С тех пор эта наука и рожденные на ее основе молекулярная генетика и геномика занимают лидирующее положение среди прочих наук о природе, став в итоге одними из основных, определяющих сегодняшний и завтрашний день развития человечества.

Вскоре после открытия основных законов генетики было установлено, что маленькие продолговатые тельца, наблюдаемые под микроскопом в ядрах клеток, которые были названы **хромосомами**, ведут себя именно так, как это ожидалось от «единиц наследственности» Менделя. Но уже тогда было ясно, что число генов должно быть больше, чем число хромосом. В 1910 году Томас Хант Морган начал изучать относительно простой и удобный для анализа генетический аппарат плодовой мушки дрозофилы, что привело в конечном итоге к созданию хромосомной теории наследственности. Согласно этой тео-

рии, существуют многочисленные гены, которые линейно расположены в хромосомах, и их последовательность в будущем может быть расшифрована.

Известно, что гены управляют развитием любого живого организма с момента его рождения и до смерти. Гены достаются нам от родителей, и от них в значительной мере зависят наши физические параметры, внешность, склонность к различным заболеваниям или, наоборот, своего рода иммунитет к ним. При этом следует обратить внимание на то, что такие черты, как характер, убеждения, привычки, поведение и даже способности также определяются в значительной мере генетически, хотя здесь существенную роль могут играть и социальные факторы, такие, как условия жизни, воспитание, образование, окружение.

Лишь в 40-е годы прошлого века была установлена материальная основа генов. Выяснилось, что ею служит одна из так называемых нуклеиновых кислот, а именно **дезоксирибонуклеиновая кислота** (сокращенно, ДНК). Само существование нуклеиновых кислот было обнаружено швейцарским биохимиком Ф. Мишером еще в 1868 году, то есть всего через три года после открытия Менделеем своих законов (случайность или закономерность?). Тогда из спермы лосося Ми-шер выделил фосфорсодержащее вещество, происходящее из клеточных ядер, которое он назвал **нуклеином** (от слова нуклеус – ядро), а мы теперь его называем дезоксирибонуклеиновой кислотой. Примечательно, что два таких важных открытия, как обнаружение единицы наследственности и ее физического носителя, были сделаны почти одновременно. Однако, как и в случае с законами Г. Менделя, практически никто из исследователей в то время не смог оценить важность открытия Мишера.

В дальнейшем существенный вклад в изучение нуклеиновых кислот внесли немецкий химик Альбрехт Кёссель и американский биохимик русского происхождения А. Ф. Левин. Первый установил, что в состав нуклеина входят четыре азотсодержащих вещества (их называли азотистыми основаниями): **аденин** (А), **гуанин** (Г), **цитозин** (Ц) и **тимин** (Т) (за это в 1910 г. Кёссель получил Нобелевскую премию по физиологии и медицине). Затем А. Ф. Левин показал, что в состав нуклеина, кроме тетрады А, Г, Ц и Т, входит добавок к фосфорной кислоте еще и сахар **дезоксирибоза**, то есть «рибоза без кислорода». Так и возникло название дезоксирибонуклеиновая кислота (ДНК). Позднее в составе другой нуклеиновой кислоты – **рибонуклеиновой кислоты** (РНК) – была обнаружена рибоза вместо дезок-сирибозы.

А. Ф. Левин предложил одну из первых гипотез о структуре нуклеиновых кислот. Согласно этой гипотезе, нуклеиновые кислоты построены как линейная комбинация связанных друг с друга химической связью нуклеотидов. По мнению Левина, четыре разных нук-леотида, входящие в состав нуклеиновых кислот, связаны последовательно в стандартный тетра-нуклеотид, который многократно повторяется в структуре нуклеиновой кислоты. И многие исследователи приняли эту гипотезу на веру. Однако столь однообразная и монотонная последовательность не годилась на роль материальной структуры генов. По этой причине долгое время бытовало мнение, что ДНК выполняет какую-то чисто структурную функцию в хромосомах.

Серия открытий, которые привели к современному пониманию генетической важности ДНК и ее основополагающей роли в организации хромосом, началась в середине 20-х годов прошлого века, когда Л. Зильберт на протейях и Ф. Гриффит на пневмококках опирали опыты по серологической трансформации. В 1944 году американский биолог Освальд Теодор Эвери с соавторами в опытах с пневмококками показали, что с помощью чистого препарата ДНК могут быть специфически изменены их наследственные свойства. Однако даже эти безупречные результаты не убедили ученых полностью в том, что ДНК – это вещество наследственности. Они заставили ученых только усомниться в том, что ДНК играет чисто структурную роль в хромосомах. Продолжала господствовать теория белковой природы гена.

И такая ситуация сохранялась вплоть до апреля 1953 года.

ДНК – МОЛЕКУЛЯРНАЯ ОСНОВА ГЕНОМА

*Сколько истин, признаваемых нами
в настоящее время бесспорными,
в момент провозглашения их казались
лишь парадоксами или даже ересьми!*

Екатерина II

*Что мыслимо – то возможно,
что возможно – то мыслимо.*

Г. Лейбниц

Тонкое устройство ДНК

Чтобы дальнейшее повествование было более ясным для читателя, рассмотрим сначала подробнее, как же устроена эта странная и загадочная молекула ДНК.

Итак, ДНК состоит из 4-х азотистых оснований, а также сахара (дезоксирибозы) и фосфорной кислоты. Два азотистых основания (сокращенно называемых Ц и Т) относятся к классу так называемых пиримидиновых основания, а два других (А и Г) – к пуриновым основаниям. Такое разделение связано с особенностями их структур, которые показаны на рис. 1.

Отдельные основания связаны в цепочке ДНК сахаро-фосфатными связями. Эти связи изображены на следующем рисунке (рис. 2).

Все это известно уже довольно давно. Но детальное устройство молекулы ДНК стало понятно лишь спустя почти 90 лет после знаменитых работ Менделя и открытия Мишера. 25 апреля 1953 г. в английском журнале «Nature» было опубликовано небольшое письмо молодых и тогда еще мало известных ученых Джеймса Уотсона и Френсиса Крика редактору журнала. Оно начиналось словами: «Мы хотели бы предложить свои соображения по поводу структуры соли

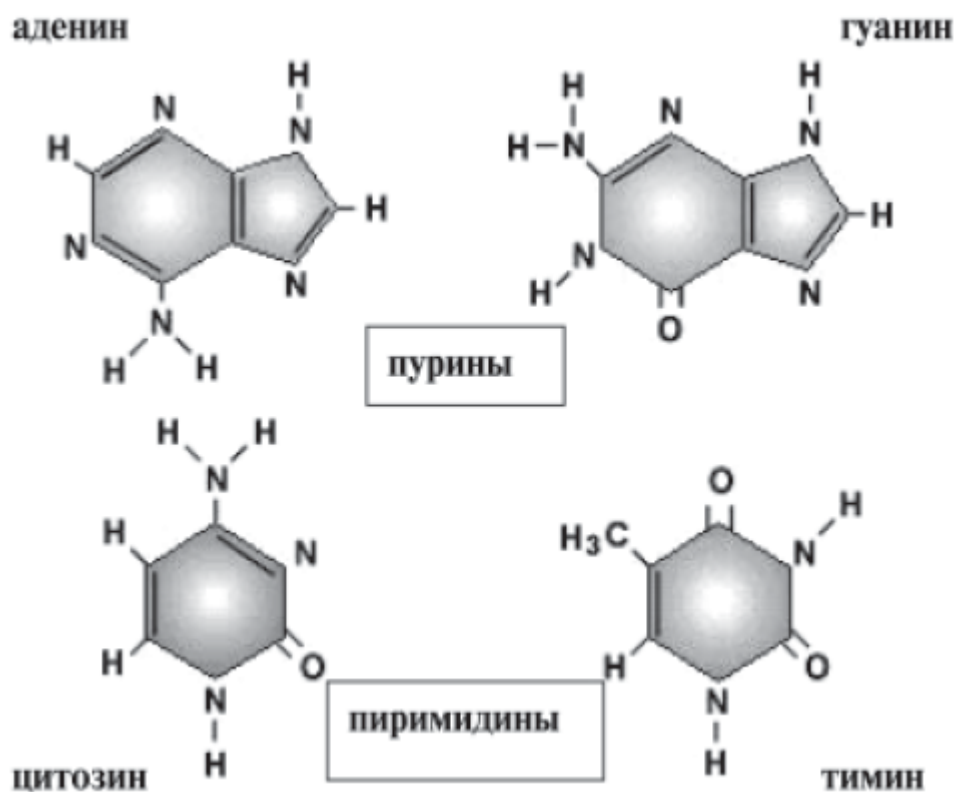


Рис. 1. Структура азотистых оснований (элементарных «букв»), из которых построена молекула ДНК

ДНК. Эта структура имеет новые свойства, которые представляют большой биологический интерес». Статья содержала всего около 900 слов, но – и это не преувеличение – каждое из них оказалось на вес золота.

А началось все так. В 1951 году на симпозиуме в Неаполе американец Джеймс Уотсон встретился с англичанином Морисом Уилкинсом. Конечно же, они тогда не могли себе даже представить, что в результате этой встречи они станут нобелевскими лауреатами. В то время Уилкинс со своей коллегой Розалиндой Франклин проводили в Кембриджском университете рентгеноструктурный анализ ДНК и определили, что молекула ДНК представляет собой, скорее всего, спираль. После разговора с Уилкинсом Уотсон «загорелся» и решил заняться исследованием структуры нуклеиновых кислот. Он перебрался в Кембридж, где познакомился с Френсисом Криком. Ученые решили совместными усилиями попытаться понять, как устроена ДНК. Работа началась не на пустом месте. Исследователи уже знали

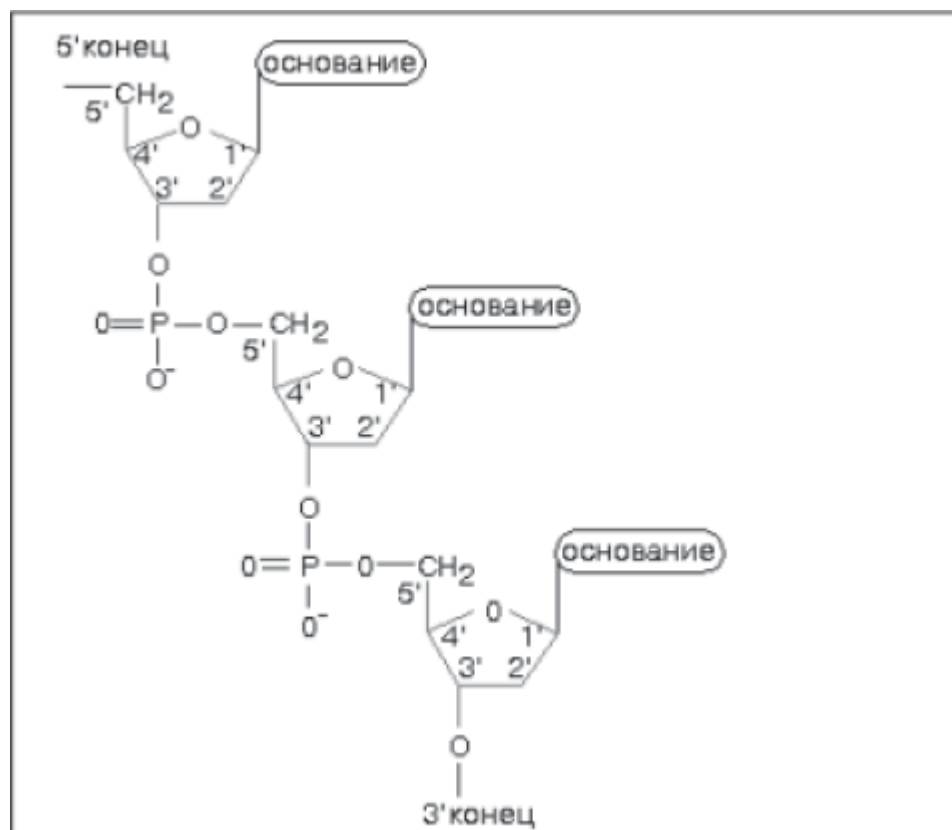


Рис. 2. Химическая структура цепи ДНК

о существовании двух типов нуклеиновых кислот (ДНК и РНК), знали и то, из чего они состоят. В их распоряжении были фотографии рентгеноструктурного анализа, полученные Р. Франклин. Кроме того, Эрвин Чаргафф сформулировал к тому времени очень важное правило, согласно которому в ДНК число А всегда равно числу Т, а число Г равно числу Ц. А далее сработала «игра ума». Результатом этой «игры» и стала статья в журнале «Nature», в которой Дж. Уотсон и Ф. Крик описали созданную ими теоретически модель строения молекулы ДНК. (Уотсону к этому времени еще не исполнилось и 25 лет, а Крику было 37). Согласно их «научной фантазии», основанной тем не менее на определенных твердо установленных фактах, молекула ДНК должна состоять из двух гигантских полимерных цепочек. Звенья каждого полимера состоят из **нуклеотидов**: углевода дезоксирибозы, остатка фосфорной кислоты и одного из 4 азотистых оснований (А, Г, Т или Ц). Последовательность звеньев в цепочке может быть любой, но эта последовательность строго связана с последовательностью звеньев в другой (парной) полимерной цепочке: напротив А должно быть Т, напротив Т должно быть А, напротив Ц должно быть Г, а напротив Г должно быть Ц (**правило комплементарности**) (рис. 3).

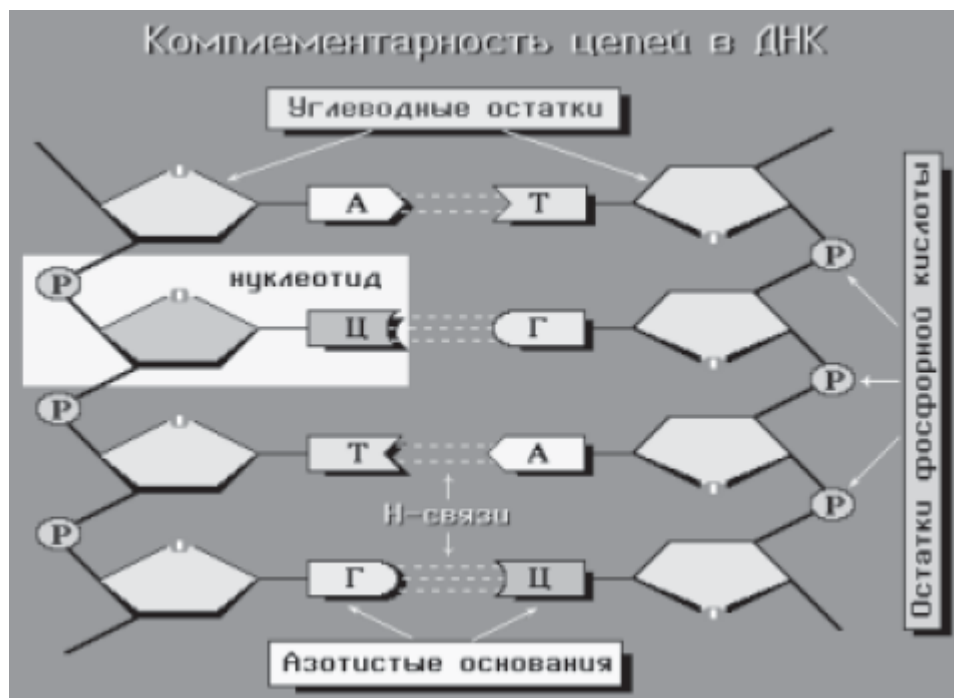
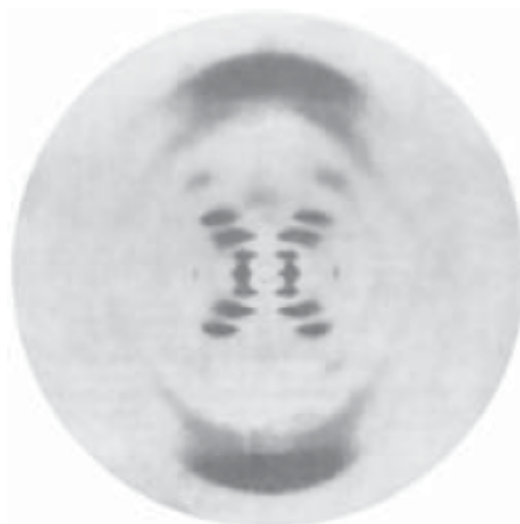


Рис. 3. Схема взаимодействия двух комплементарных цепей в молекуле ДНК

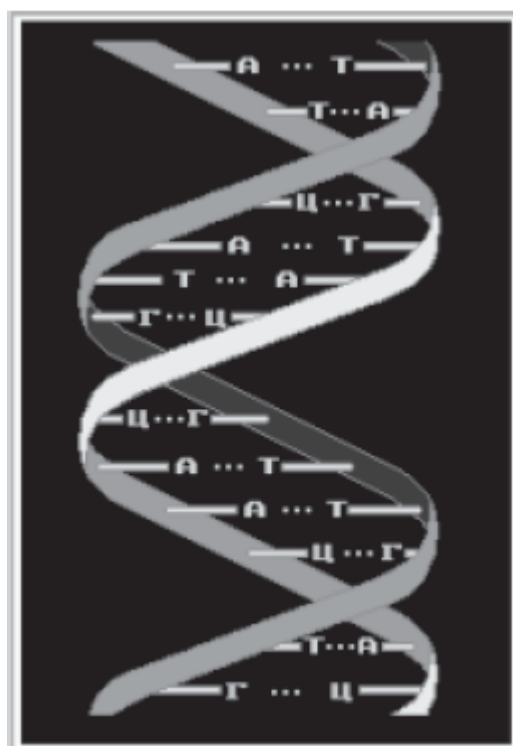
Две полимерные цепи закручены в правильную двойную спираль. Они удерживаются вместе посредством водородных связей между парами оснований (А—Т и Г—Ц) подобно ступенькам лестницы. По этой причине говорят, что две цепи ДНК комплементарны. Для природы это не удивительно. Известно множество примеров комплементарности. Комплементарны, например, древнекитайские символы «инь» и «янь», гнезда розетки и штырьки вилки.

Двойная спираль ДНК схематически изображена на рис. 4. Внешне она напоминает веревочную лестницу, завитую в правую спираль. Ступенями в этой лестнице являются пары нуклеотидов, а связывающие их «боковинки» состоят из сахаро-фосфатного остова.

Так была открыта знаменитая «двойная спираль». Если последовательность звеньев (нуклеотидов) в ДНК рассматривать как ее первичную структуру, то двойная спираль – это уже вторичная структура ДНК. Предложенная Уотсоном и Криком модель «двойной спирали» изящно решала не только проблему кодирования информации,



а



б

Рис. 4. Знаменитая двойная спираль ДНК а – Рентгенограмма ДНК, полученная Р. Франклин, которая помогла Уотсону и Крику найти ключ к двухспиральной структуре ДНК; б – Схематическое изображение двухспиральной молекулы ДНК

но и удвоения (репликации) гена. В 1962 году Дж. Уотсон, Ф. Крик и Морис Уилкинс получили по достоинству за это достижение Нобелевскую премию. А ДНК была названа самой главной молекулой живой природы. Во всем этом, конечно же, сыграли свою роль точные сведения о строении ДНК, но не в меньшей мере и «провидческие» построения сложной пространственной структуры, что потребовало от исследователей не только логики, но и творческого воображения – качества, присущего художникам, писателям и поэтам. «Здесь, в Кембридже, произошло, быть может, самое выдающееся после книги Дарвина событие в биологии – Уотсон и Крик раскрыли структуру гена!» – писал в то время в Копенгаген Нильсу Бору его бывший ученик М. Дельбрюк. Известный испанский художник Сальвадор

Дали после открытия двойной спирали сказал, что это для него явилось доказательством существования Бога, и изобразил ДНК на одной из своих картин.

Итак, интенсивный мозговой штурм, предпринятый учеными, завершился полным успехом! В историческом масштабе открытие структуры ДНК сопоставимо с открытием структуры атома. Если выяснение строения атома привело к появлению квантовой физики, то открытие структуры ДНК дало начало молекулярной биологии.

Каковыми же оказались главные физические параметры ДНК человека – этой главной его молекулы? Диаметр двойной спирали равен 2 нанометрам ($1 \text{ нм} = 10^{-9} \text{ м}$); расстояние между соседними парами оснований («ступеньками») составляет 0,34 нм; один поворот спирали состоит из 10 пар оснований. Последовательность пар нуклеотидов в ДНК нерегулярна, но сами пары уложены в молекуле как в кристалле. Это дало основание характеризовать молекулу ДНК как линейный аperiодический кристалл. Число отдельных молекул ДНК в клетке равно числу хромосом. Длина такой молекулы в наибольшей по размеру хромосоме 1 человека составляет около 8 см. Подобных гигантских полимеров пока не выявлено ни в природе, ни среди искусственно синтезированных химических соединений. У человека длина всех молекул ДНК, содержащихся во всех хромосомах одной клетке, составляет примерно 2 метра. Следовательно, длина молекул ДНК в миллиард раз больше их толщины. Так как организм взрослого человека состоит примерно из $5 \times 10^{13} - 10^{14}$ клеток, то общая длина всех молекул ДНК в организме равна 10^{11} км (это почти в тысячу раз больше расстояния от Земли до Солнца). Вот такая она, суммарная ДНК всего лишь одного человека!

Когда говорят о размере генома, то подразумевают общее содержание ДНК в единичном наборе хромосом ядра. Такой набор хромосом называют гаплоидным. Дело в том, что большинство клеток нашего организма содержит двойной (диплоидный) набор совершенно одинаковых хромосом (только у мужчин 2 половые хромосомы отличаются). Измерения размера генома приводятся в дальтонах, парах нуклеотидов (п. н.) или пикограммах (пг). Соотношение между этими единицами измерения следующие: $1 \text{ пг} = 10^{-9} \text{ мг} = 0,6 \times 10^{12} \text{ дальтон} = 0,9 \times 10^9 \text{ п. н.}$ (далее мы будем использовать в основном п. н.). В гаплоидном геноме человека содержится около 3,2 млрд. п. н., что равно 3,5 пг ДНК. Таким образом, в ядре одной клетки человека содержится около 7 пг ДНК. Если учесть, что средний вес клетки человека равен примерно 1000 пг, то легко рассчитать, что ДНК составляет менее 1% от веса клетки. И тем не менее, чтобы воспроизвести самым мелким шрифтом (как в телефонных справочниках) ту огромную информацию, которая содержится в молекулах ДНК одной нашей клетки, понадобилось бы тысяча книг по 1000 страниц в каждой! Вот таков полный размер генома человека – Энциклопедии, написанной четырьмя буквами.

Но не следует думать, что геном человека наибольший из всех существующих в природе. Например, у саламандры и лилии длина молекул ДНК, содержащихся в одной клетке, в тридцать раз больше, чем у человека.

Поскольку молекулы ДНК имеют гигантский размер, их можно выделить и увидеть даже в домашних условиях. Вот как описывается эта простая процедура в рекомендации для кружка «Юный генетик». Во-первых, надо взять любые ткани животных или растительных организмов (например, яблоко или кусок курицы). Затем надо нарезать ткань на кусочки и положить 100 г в обычный миксер. После добавления 1/8 чайной ложки соли и 200 мл холодной воды вся смесь взбивается на миксере в течение 15 секунд. Далее взбитая смесь процеживается через ситечко. В полученную мякоть надо добавить 1/6 от ее количества (это будет примерно 2 столовые ложки) моющего средства (для посуды, например) и хорошо размешать. Через 5–10 минут жидкость разливается по пробиркам или любым другим стеклянным емкостям, чтобы в каждой из них было заполнено не больше трети объема. Затем к ней добавляется по чуть-чуть либо сок, выжатый из ананаса, либо раствор, используе-

мый для хранения контактных линз. Все содержимое встряхивается. Делать это надо весьма осторожно, так как если трясти слишком сильно, то гигантские молекулы ДНК поломаются и после этого ничего нельзя будет увидеть глазами. Далее в пробирку медленно вливается равный объем этилового спирта, чтобы он образовал слой поверх смеси. Если после этого покрутить в пробирке стеклянной палочкой, на нее «наматывается» вязкая и почти бесцветная масса, которая и представляет собой препарат ДНК.

Генетическая грамматика

После установления химического строения и пространственной структуры ДНК осталось еще множество вопросов, основной из которых заключался в том, как же ДНК кодирует белки, то есть, что представляет из себя генетический код этой молекулы, какую «грамматику» она использует? На это в первую очередь и были направлены дальнейшие усилия исследователей.

Итак, установлено, что «буквами» в ДНКовом тексте служат нук–леотиды – элементарные звенья полимерной молекулы ДНК. В ДНК всего 4 нуклеотида (А, Т, Г, Ц). Следовательно, если сравнить каждый из этих нуклеотидов с отдельной буквой, то **алфавит ДНКового текста содержит всего 4 «буквы»**. Как же из этих «букв» формируются «слова» и «предложения»?

Белковые молекулы всех существующих на земле организмов построены всего из 20 аминокислот. Сразу после создания модели ДНК стало ясно, что существует некий код, переводящий четырехбуквенный ДНКовый текст в двадцатибуквенный аминокислотный текст. Элементарные расчеты говорили о том, что число возможных сочетаний, в которых четыре нуклеотида могут быть по–разному расположены в «тексте», достигает астрономических значений. Так, молекула ДНК, состоящая, к примеру, всего из 100 пар нуклеотидов, может теоретически кодировать 4^{100} различных белковых «текстов». Какова же ситуация на самом деле?

Одним из первых в этом пытался разобраться русский физик Г. Гамов, эмигрировавший в то время в Америку. Наслушавшись многочисленных разговоров о ДНК и узнав, что она содержит – как и карты – всего четыре «масти», Гамов решил «разложить пасьянс» с целью понять устройство генетического кода. Ему сразу стало ясно, что код не может быть «двоичным», то есть одну аминокислоту в белке должна кодировать не двойка нуклеотидов – «букв», а как минимум тройка. Дело в том, что сочетание из 4 по 2 дает всего 16 комбинаций, а этого недостаточно для кодирования всех 20 аминокислот. Следовательно, рассуждал Гамов, код должен быть по крайней мере трехбуквенным, то есть каждую аминокислоту должна кодировать тройка «букв» в любых сочетаниях. На этом он и остановился, поскольку далее возникало множество вопросов. В частности, такой: число сочетаний из 4 по 3 равно 64, а аминокислот всего 20. Зачем же такая избыточность в трехбуквенном коде?

В то время уже существовал хорошо известный путь, который, в частности, был проделан в свое время французом Жаном Шамполь–оном при дешифровке иероглифов древнего Египта. В качестве основного подспорья для решения стоящей перед ним задачи он использовал базальтовую плиту, которую обнаружили во время военной компании Наполеона в Египет и которая получила название Розеттский камень. На плите одновременно присутствовали две надписи: одна была иероглифическая, а другая – сделанная греческими буквами на греческом языке. К счастью, и язык, и письмо древних греков были в то время уже хорошо известны ученым. В результате сравнение двух текстов Розеттского камня привело к расшифровке египетской иероглифики. Этим путем и двинулись ученые при расшифровке генетического кода. Надо было сравнить два текста: текст, записанный в ДНК, с текстом, записанным в белке. Однако первоначально ученые не умели «читать» ДНК, а одного

известного в то время белкового текста было недостаточно. Пришлось искусственно синтезировать разнообразные короткие фрагменты РНК и синтезировать на них в искусственных системах фрагменты белка. Весной 1961 года в Москве на Международном биохимическом конгрессе М. Ниренберг сообщил, что ему удалось «прочитать» первое «слово» в ДНК-овом тексте. Это была тройка букв—ААА (в РНК, соответственно, УУУ), то есть три аденина, стоящие друг за другом, – которая кодирует аминокислоту фенилаланин в белке. Так было положено начало расшифровке генетического кода.

Такой путь в конечном итоге вскоре привел к полной расшифровке генетического кода. Подтвердилось предположение Гамова, что код триплетный: одной аминокислоте в белках соответствует последовательность из 3 нуклеотидов в ДНК и РНК. Такие кодирующие тройки нуклеотидов – «слова» – получили название **кодонов**.

Напомним, что еще Гамов столкнулся с парадоксом: из четырех нуклеотидов может быть построено 64 разных кодонов, а для построения белков используется только 20 различных аминокислот. Решение этого парадокса оказалось в следующем. Большинство аминокислот может кодироваться несколькими кодонами. После выяснения этого обстоятельства генетический код назвали **вырожденным**.

В таблице 1 приведены кодоны, но не в самой ДНК, а в РНК–посреднике (матричной РНК, или мРНК), образующейся на ДНК, и соответствующие им аминокислоты в белках.

Кроме того, как видно из таблицы, реально для кодирования используются не все возможные кодоны. Три из этих «лишних» кодонов выполняют функцию стоп–сигналов, обеспечивая прекращение синтеза белковой цепи.

Если внимательно посмотреть на таблицу 1, то видно, что вырожденность генетического кода носит не совсем случайный характер. Хотя код триплетный, основную нагрузку несут первые два нуклеотида в каждом кодоне. Чаще всего в разных кодонах, кодирующих одну и ту же аминокислоту, отличается лишь третий нуклеотид.

Таблица 1. Генетический словарь. Указаны аминокислоты, встречающиеся в белках, и соответствующие им кодоны в комплементарной ДНК матричной РНК

Аланин	Аргинин	Аспарагин	Аспарагиновая кислота	Валин
ГЦУ, ГЦА, ГЦЦ, ГЦГ	ЦГУ, ЦГА, ЦГЦ, ЦГГ, АГА, АГГ	ГАУ, ГАЦ	ААУ, ААЦ	ГУУ, ГУЦ, ГУА, ГУГ
Гистидин	Глицин	Глутамин	Глутаминовая кислота	Изолейцин
ЦАУ, ЦАЦ	ГГУ, ГГА, ГГЦ, ГГГ	ГАА, ГАГ	ЦАА, ЦАГ	АУУ, АУА, АУЦ
Лейцин	Лизин	Метионин	Пролин	Серин
УУА, УУГ, ЦУЦ, ЦУГ, ЦУА, ЦУЦ	ААА, ААГ	АУГ	ЦЦГ, ЦЦЦ, ЦЦА, ЦЦУ	АГУ, АГЦ, УЦА, УЦГ, УЦУ, УЦЦ
Тирозин	Треонин	Фенилаланин	Триптофан	Цистеин
УАУ, УАЦ	АЦУ, АЦА, АЦГ, АЦЦ	УУУ, УУЦ	УГГ	УГУ, УГЦ
Нет аминокислоты				
УАА, УАГ, УГА				

Генетический код первоначально был расшифрован у таких простых организмов, как фаги и бактерии. В дальнейшем оказалось, что он универсален (за очень редким исключением) для геномов всех существующих ныне живых организмах (от бактерий до человека). Небольшие отличия, о которых мы поговорим далее, были выявлены при сравнении ядерного и митохондриального геномов.

Итак, как в привычном нам тексте книги, вся информация записана в ДНК последовательностью расположения четырех составляющих ее «букв» – нуклеотидов. Таким образом, ДНК-текст написан с помощью А, Т, Ц, Г-алфавита. При этом только текст одной из двух цепей ДНК обычно кодирующий, а другая цепь, как правило, не кодирующая. Хотя известно, что в каждом правиле есть исключения. Если читатель попытается написать этими четырьмя буквами какие-нибудь русские слова, то у него ничего не получится. «Словом» в ДНК-тексте, условно говоря, служит определенное сочетание трех нуклеотидов, которому соответствует конкретная аминокислота в белке, являющемся также полимером. Таким образом, в клетке четырьмя буквами записано два десятка «слов» (аминокислот – составных частей белков). И, наконец, как «предложение» в ДНК-тексте можно рассматривать полный набор триплетов, кодирующих определенный белок, то есть ген. Таким образом, генетический алфавит состоит всего из 4 букв, а генетический словарь из 20 слов. В этой связи вспомним, что даже словарь Элочки-людоедки из романа И. Ильфа и Е. Петрова «Двенадцать стульев» состоял из 30 слов, а «Словарь языка произведений А. С. Пушкина» насчитывает примерно 20 тыс. слов.

Существует строгая закономерность: чем длиннее код (чем больше в нем знаков), тем короче тексты. Огромный по размерам код представляют собой, например, китайские иероглифы. В результате этого иероглифические тексты существенно более кратки по сравнению с другими системами письма, в том числе и нашей. Однако для создания генетического кода природа выбрала всего 4 «буквы». Такой код предполагает наличие длинных текстов, что и реализовалось природой в виде создания гигантских молекул ДНК. При написании полного «текста» генома человека потребовалось около 3,2 млрд. «букв». Для сравнения: в священной книге Бытия, написанной на древнееврейском языке, содержится всего 78100 букв.

Размножение ДНК (репликация)

Важно то, что структура ДНК, открытая Уотсоном и Криком, многое прояснила относительно разных механизмов функционирования этой молекулы в клетке. ДНК не только кодирует генетическую информацию, но и самовоспроизводится (удваивается) при каждом клеточном делении. И вскоре уже было экспериментально установлено, что одновременно с делением клетки ДНК снимает с самой себя точные копии в процессе **удвоения**, или **репликации**. Во время клеточного деления слабые связи между двумя цепями двойной спирали ДНК разрушаются, в результате чего нити разделяются. Затем на каждой из них строится вторая «дочерняя» (комплементарная) цепь ДНК. В результате этого молекула ДНК удваивается, как и клетка, и в обеих клетках оказывается по одной полной копии ДНК. Копии должны быть полностью идентичными, чтобы сохранить всю генетическую информацию.

Процесс репликации играет ключевую роль в сохранении одной и той же генетической информации в разных клетках, образующихся при их делении. В общем виде художественно он изображен на рис. 5. Однако реальные механизмы репликации довольно сложны, и до настоящего времени еще не все тонкие детали этого процесса известны, особенно применительно к геномам высших животных организмов, включая человека.



Рис. 5. Схема репликации ДНК

В общем виде этот процесс выглядит следующим образом. В каждой хромосоме ДНК удваивается не с начала до конца, а отдельными кусками (**репликонами**). Средний размер репликона составляет около 30 мкм. Тем самым в составе генома человека должно встречаться более 50 000 репликонов, участков ДНК, которые синтезируются в ядре как независимые единицы. И это имеет свой глубокий смысл. Если бы каждая из молекул ДНК удваивалась как один репликон от начала до конца молекулы, то при скорости синтеза 0,5 мкм в минуту (а она именно такова у человека) удвоение первой хромосомы, имеющей длину ДНК около 7 см, занимало бы 140 000 минут, или около трех месяцев. На самом деле благодаря полирепликонному строению молекул ДНК весь процесс занимает всего 7—12 часов. Отдельные относительно короткие репликоны соединяются друг с другом, обеспечивая этим процесс воспроизведения целой молекулы ДНК.

Перезапись генетического текста и перевод в белковый текст (транскрипция и трансляция)

В клетке ДНК служит в качестве матрицы, на которой первоначально происходит синтез разных РНК. Процесс перезаписи генетической информации из ДНК в РНК-овый текст получил название **транскрипция**. Этот процесс, как и репликация ДНК, осуществляется в ядрах клеток. Первоначально на генах, кодирующих белки, образуются РНК—предшествен-

ники, которые после ряда модификаций превращаются в так называемые матричные РНК (мРНК). Они – то непосредственно и служат матрицей для синтеза белков, то есть их кодируют. Установлено, что мРНК служат не только носителями ДНКовой информации, но и переносчиками этой информации из ядра в цитоплазму клетки. Только там мРНК может играть роль матрицы для синтеза белковых молекул (этот процесс назван **трансляцией**). В цитоплазме на специфических «машинах» – рибосомах – осуществляется при трансляции мРНК синтез молекулы белка, т. е. происходит перевод информации с четырехбуквенного языка мРНК на двадцатибуквенный язык белка. Схематически этот процесс изображен на рис. 6.

В середине 60-х годов был сформулирован основной постулат (центральная догма) новой науки – молекулярной биологии, который первоначально выглядел следующим образом:

ДНК → РНК → белок.

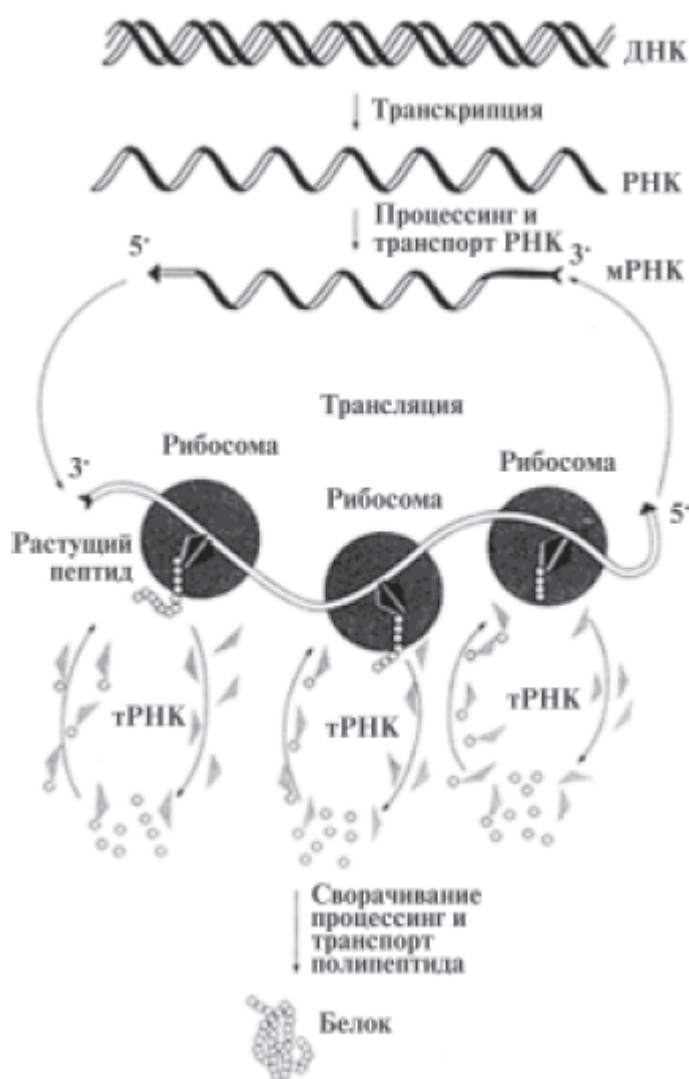


Рис. 6. Процесс трансляции мРНК на рибосомах с образованием белка. Рибосома играет роль «машины», читающей генетическую информацию, записанную в мРНК. Это молекулярная машина, построенная по единой схеме у всех организмов (включая человека) с небольшими вариациями. Она состоит из двух рибонуклеопротеидных (состоящих из РНК и белков) субчастиц: малой и большой. На рибосоме происходит взаимодействие мРНК с

транспортными РНК (тРНК), несущими по одной аминокислоте каждая. Белок синтезируется путем образования связей между последовательно доставляемыми тРНК аминокислотами. Этим процессом «руководит» сама рибосома

Позднее этот постулат был уточнен, и теперь он выглядит как



Сделанное дополнение вызвано обнаружением в 1970 году такого явления, как **обратная транскрипция**. Выяснилось, что у некоторых вирусов, генетический аппарат которых представлен РНК, а не ДНК, как у других живых организмов, имеется специальный фермент, который позволяет осуществлять такой обратный процесс, как синтез ДНК на РНК. Этот фермент получил название **обратная транскриптаза**, или **ревертаза**.

Казалось бы, теперь уже все стало ясно и однозначно и более ничего не может измениться. Но, как писал Марк Твен: «Из жизненного опыта следует извлекать только полезное и ничего больше, – иначе мы уподобимся кошке, присевшей на горячую печку. Она никогда больше не сядет на горячую печку – и хорошо сделает, но она никогда больше не сядет и на холодную». Ряд открытий последних лет продолжает вносить уточнения в казавшуюся незыблемой центральную догму молекулярной биологии.

Во–первых, выяснилось, что центральная догма молекулярной биологии постулирует лишь путь передачи генетической информации от нуклеиновых кислот к белкам и, следовательно, к конкретным свойствам и признакам живого организма. Однако исследование механизмов реализации этого пути на протяжении нескольких десятилетий, последовавших за формулировкой центральной догмы, вскрыло гораздо более разнообразные функции РНК, помимо известных ранее.

Во–вторых, согласно первоначальной догме, все носители инфекционных болезней должны иметь генетический материал – ДНК или РНК. Оказалось, что и здесь есть исключения. В 1997 году Нобелевская премия была вручена Стэнли Прузинеру за открытие белковых инфекционных частиц, вызывающих такое заболевание, как, например, болезни «коровьего бешенства», или почесухи. Эти частицы были названы **прионами**. Проникая в клетку–хозяина, прионы «навязывают» свою болезнетворную конформацию (измененную пространственную структуру) нормальным белкам–аналогам, содержащимся в клетках. При этом ни РНК, ни ДНК никак не участвуют в развитии заболевания. Иными словами, при прионовых заболеваниях информация передается не от одной нуклеиновой кислоты к другой нуклеиновой кислоте, а от белка к белку, что в исходном варианте центральной догмы не предусматривалось.

Однако все эти дополнения в «центральную догму» молекулярной биологии не повлияли принципиально на ее общую закономерность. В центре всего и вся стоял и стоит ген. Сегодня он перестал быть чем–то таинственным, стал реальным химическим веществом, появилась возможность судить о нем так же, как и о других химических соединениях живых организмов, изучать с помощью доступных генетикам и биохимикам методов. Стало понятным, что такое генетический код и как реализуется в клетке та информация, которая записана в ДНКовом тексте. В результате этого в 1953 году родилась молекулярная генетика – раздел науки, который занялся детальным изучением процессов работы ДНК в клетке на молекулярном уровне.

Расшифровка химической и пространственной структуры ДНК – носителя генетической информации – оценена во всем мире как одно из наиболее выдающихся открытий XX

века. Геном стоит в центре всех биологических проблем, всех свойств и способностей человека, всего разнообразия человека. Теперь это уже аксиома. Как говорил Козьма Прутков: «Многие люди подобны колбасам: чем их начиняют, то и носят в себе». Так вот, мы «начинены» ДНК, носим ее в себе, а она—то, главным образом, и определяет многое в нас.

ДЛИННАЯ ИСТОРИЯ СО СЧАСТЛИВЫМ КОНЦОМ

Сначала неизбежно идут: мысль, фантазия, сказка.

За ними шествует научный расчет.

И уже, в конце концов, исполнение венчает мысль.

К. Э. Циолковский

Начало – половина целого.

Пифагор

После открытия структуры ДНК и расшифровки содержащегося в ней генетического кода сразу же начался следующий важный этап – изучение организации целых геномов. Но до прямого определения всех нуклеотидных последовательностей гигантских молекул ДНК было еще очень далеко. Дело в том, что в 60–х и начале 70–х годов прошедшего столетия технические возможности и методические подходы к исследованиям столь больших биологических молекул были еще весьма несовершенны. Однако энтузиазм ученых после установления основных особенностей структуры ДНК был огромным, и на эту главную молекулу жизни навалились со всех возможных сторон многочисленные специалисты, используя все доступные на то время методы и средства.

Хромосомы дают первые сведения о структуре генома

Выше уже говорилось, что в ядре клетки молекулы ДНК расположены в особых структурах, получивших название **хромосомы**. Их исследование началось еще свыше 100 лет назад с помощью обычного светового микроскопа. Уже к концу XIX века выяснилось кое-что о поведении хромосом в процессе деления клеток и высказывалась мысль об их участии в передаче наследственности.

Хромосомы становятся видимыми в микроскопе при делении клетки на определенной стадии клеточного цикла, называемой **митозом**. Хромосомы в этом состоянии представляют собой компактные палочковидные структуры разной длины с довольно постоянной толщиной, у большей части хромосом имеется перетяжка, которая делит хромосому на два плеча. В области перетяжки расположена важная для удвоения хромосом структура, называемая **центромерой**. При делении клетки в ходе митоза происходит удвоение числа хромосом, в результате которого обе вновь образующиеся клетки в конечном итоге обеспечиваются одним и тем же стандартным набором хромосом.

Лишь в 1956 г. впервые Ю. Тио и А. Леван описали хромосомный набор человека, определили количественный состав хромосом и дали их общую морфологическую характеристику. По сути дела эти работы и положили начало изучению структуры генома человека. У человека в каждой клетке тела содержится 46 хромосом, физические длины которых находятся в пределах от 1,5 до 10 мкм (рис. 7).

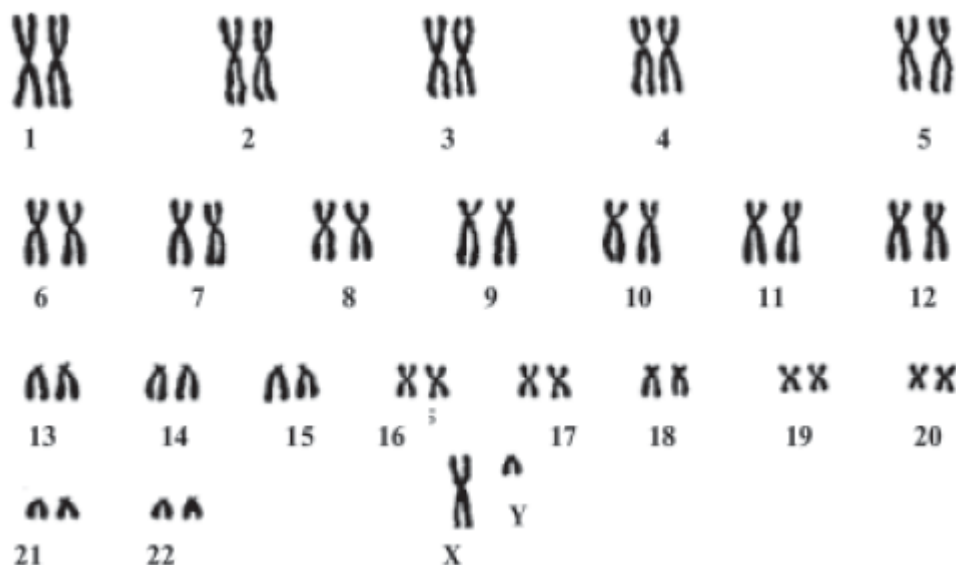


Рис. 7. Вид под микроскопом полного набора хромосом, содержащихся в ядре каждой отдельной клетки человека

Напомним читателю, что набор хромосом во всех клетках человека (за исключением половых) называют диплоидным (двойным), поскольку каждая из хромосом представлена двумя копиями (всего 23 пары). Каждая соматическая клетка человека (кроме красных кровяных клеток крови) содержит по 2 полных набора хромосом. В каждом единичном (гаплоидном) наборе присутствует 23 хромосомы – 22 обычные хромосомы (аутосомы) и по одной половой хромосоме – X или Y. Таким образом, геном каждого конкретного человека состоит из 23 пар гигантских молекул ДНК, распределенных в разных хромосомах, а если говорить о геноме человека вообще (мужчин и женщин), то общее число таких молекул равно 24. Это первое базовое сведение, которое было получено о геноме человека при анализе хромосом.

Изучение строения (размера и формы) хромосом человека показало, что большинство из них по внешнему виду напоминают кегли, состоящие из двух толстых частей (хроматид) и тонкой перетяжки (центромеры) между ними. Сходство с кеглями, а не с гантелями заключается в том, что центромера чаще всего расположена не в центре хромосомы, а смещена к одному из ее концов. Размеры хромосом сильно варьируют, самая короткая хромосома примерно в десять раз меньше, чем самая длинная. Это второе принципиально важное сведение о структуре генома человека – составляющие его 24 молекулы ДНК имеют разный размер.

Если сравнивать число и размер хромосом у человека и у других видов организмов, то можно увидеть огромные отличия. Например, у коровы, размер генома которой примерно равен геному человека, имеется 60 пар хромосом. У шпорцевой лягушки содержится всего 18 хромосом, но даже самые маленькие из них больше, чем самые крупные хромосомы человека. У птиц, наоборот, число хромосом достигает 40 и более и все они очень небольшие по размерам. Таким образом, разнообразие хромосом в природе весьма велико.

С помощью световой микроскопии были определены размеры всех хромосом человека. Затем все неполовые хромосомы были пронумерованы по уменьшению размера – от 1 до 22. Половым хромосомам не присвоили номер, а назвали X и Y. Как показали более точные последующие исследования, хромосома 21 реально оказалась чуть меньше 22, однако нумерацию хромосом не изменили (чтобы не вносить путаницу). Различие в хромосомных наборах между мужчинами и женщинами состоит в том, что у женщин имеются две половые X-хромосомы (т. е. хромосомы во всех 23-х парах одинаковы), а у мужчин пару с X-хромо-

сомой образует мужская половая хромосома – Y. Каждую хромосому можно рассматривать как отдельный том большого двадцатичетырехтомного собрания сочинений под названием Энциклопедия человека.

Половые клетки человека, в отличие от клеток тела взрослого организма (соматических клеток), содержат не 2 набора томов ДНК-ового текста, а всего лишь один. Перед зачатием каждая отдельная хромосома (отдельный том в Энциклопедии человека) сперматозоида отца и яйцеклетки матери состоят из смешанных в разном сочетании различных глав ДНК-ового текста их родителей. Любая из хромосом, полученная нами от отца, образовалась в его семенниках незадолго до того, как мы были зачаты. Ранее, за всю историю человечества, точно такая хромосома никогда не существовала. Она была сформирована в процессе случайного перемешивания, происходящего при делении, постепенно образуясь из объединяющихся друг с другом участков хромосом предков со стороны отца. Также обстоит дело и с хромосомами яйцеклеток, за исключением того, что они формируются в организме нашей матери задолго до нашего рождения (почти сразу после рождения самой матери).

В зиготе, образующейся в результате слияния сперматозоида и яйцеклетки, материнские и отцовские гены смешиваются и перетасовываются в разных сочетаниях. Это происходит в результате того, что хромосомы не остаются неизменными в поколениях – они вступают во взаимодействие со своей случайно встреченной парой, обмениваясь с ней материалом. Такой постоянно идущий процесс получил название **рекомбинации**. И следующему поколению часто достается уже гибридная хромосома – часть от бабушки и часть от дедушки. Далее в ряду поколений пути генов постоянно пересекаются и расходятся. В результате слияния уникальной яйцеклетки с уникальным сперматозоидом и возникает уникальный во всех отношениях геном. И в этом смысле все мы уникамы. Каждый человеческий индивид хранит уникальную генетическую информацию, состоящую из случайной комбинации разных вариантов генов.

Отдельный ген можно рассматривать как единицу, продолжающую существовать в ряду многочисленных поколений. И в этом смысле ген бессмертен! Существует даже такая оригинальная точка зрения, что не сами люди, а их гены правят миром, а каждый конкретный живой организм служит лишь временным прибежищем для них. Эта не бесспорная мысль принадлежит Ричарду Докинзу, автору книги «Эгоистичный ген». По его мнению, гены практически бессмертны в отличие от живых организмов, в которых они существуют. Некоторым генам десятки и даже сотни миллионов лет. Гены, пользуясь терминологией Докинза, делают все возможное, чтобы выжить. Приспосабливаются к жаре и холоду, выбирая себе местечко получше, мигрируют с помощью человека и вступают в новые комбинации. Человек оказался довольно непоседливым хозяином. За тысячи лет он сильно исколесил мир, распространяя свое присутствие, влияние и свою начинку – гены. (Подробнее с идеями и аргументацией Р. Докинза любознательный читатель может познакомиться в **Приложении 1**). Такая точка зрения далеко не бесспорна, и из дальнейшего изложения нам станет понятно, что гены – это в первую очередь не эгоисты, а трудоголики. Имеются гены – «сторожа» генома, гены – «дворники», гены – «повара» и гены – «домоправители». Обеспечивая свое существование, они обеспечивают и существование нас.

Сразу после зачатия будущий человек представляет собой всего одну клетку (зиготу), наделенную одной исходной ДНК-овой библиотекой, содержащей 46 томов. Среди 46 томов всегда 23 получены от отца, а другие 23 – от матери. Тексты 23 отцовских и 23 материнских томов хотя и очень сходны в целом, тем не менее отличаются в деталях. Например, в отцовском томе N18 на странице 253 существует предложение–предписание (в виде гена), в котором сказано, что глаза у ребенка должны быть карими, а в этом же материнском томе на той же странице тоже написано о цвете глаз, но согласно этому тексту цвет должен быть голубыми. Первое указание более строгое (доминирующее), чем второе, и в результате у ребенка

глаза будут иметь карий цвет. Ген, который диктует свои права, называют доминирующим, а тот, который уступает свои права, – рецессивным. Голубой цвет глаз имеют только те люди, у которых и в материнском, и в отцовском тексте содержатся рецессивные гены, в которых есть указание на голубоглазость. Затем зигота делится на две клетки, каждая из них вновь делится и так до появления миллиардов клеток. Схематически процесс деления клеток изображен на рис. 8.

При каждом делении клетки содержащиеся в библиотеках тома ДНКового текста точно копируются, причем практически без ошибок. Организм взрослого человека состоит в среднем из 10^{14} клеток. Например, в головном мозге и печени насчитывается примерно по 10 млрд. клеток, в иммунной системе – 300 млрд. клеток. В течение всей жизни человека в его организме происходит около 10^{16} клеточных делений. Клеточный состав многих органов за 70 лет жизни обновляется несколько раз. И каждая из этих клеток содержит одни и те же 46 томов ДНКового текста.

В конце 60–х годов XX века был осуществлен важный прорыв в исследовании хромосом. Обусловлен он был всего лишь тем, что для их окраски стали использовать специальное контрастное вещество – акрихин–иприт, а затем и другие сходные с ним соединения. Такая окраска позволила выявить внутри хромосом большое число разных субструктур, которые не обнаруживались под микроскопом без окрашивания. После окрашивания хромосом специфическим красителем Гимза–Романовского они выглядят как зебры: вдоль всей длины видны поперечные светлые и темные полосы, имеющие окраску разной интенсивности. Эти полосы получили название хромосом



Рис. 8. Основные стадии клеточного цикла, приводящего к делению клетки

ных G – сегментов или полос (рис. 9). Картина сегментации сильно отличается у разных хромосом, но расположение хромосомных сегментов постоянно у каждой из хромосом во всех типах клеток человека.

Природа полос, выявляемых при окраске, до конца еще не ясна. Сейчас установлено только, что участки хромосом, соответствующие темным полосам (названные R – полосами), реплицируются раньше, чем светлые участки (названные G – полосами). Таким образом, поло-сатость хромосом скорее всего все же имеет некий до конца еще не понятый смысл.

Окрашивание хромосом очень облегчило их идентификацию, а в дальнейшем способствовало определению расположения на них генов (картированию генов). Хотя детальные процессы, происходя

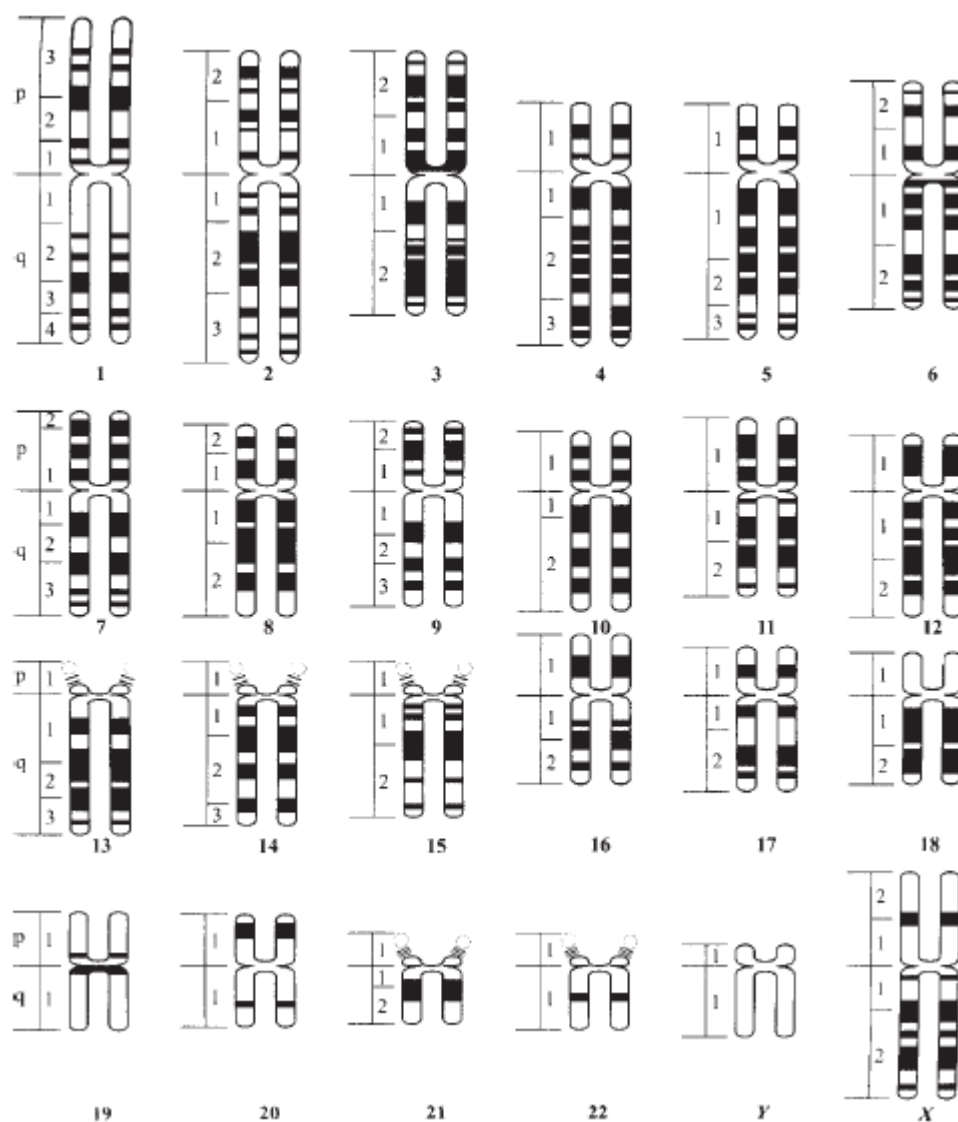


Рис. 9. Специфические хромосомные G – сегменты, выявляемые при окраске хромосом человека, и система их обозначения согласно решению международной конференции в Париже в 1971 году. Цифрами под хромосомами указаны их номера. X и Y —половые хромосомы, p – короткое плечо, q – длинное плечо хромосом

щие при окрашивании, еще не до конца ясны, очевидно, что картина окраски зависит от такого параметра, как увеличенное или уменьшенное содержание в отдельных полосах хромосом АТ–или ГЦ–пар. И это еще одно общее сведение о геноме – он не однороден, в нем есть районы, обогащенные определенными парами нуклеотидов.

Это, в частности, может быть связано с повторяемостью некоторых типов нуклеотидных последовательностей ДНК в определенных районах.

Дифференциальная окраска хромосом нашла широкое применение для выявления и идентификации небольших индивидуальных изменений генома конкретного человека (**полиморфизма**), в частности, приводящих к различным патологиям. Примером этому может служить обнаружение так называемой филадельфийской хромосомы, встречающейся у больных с хроническим миелоидным лейкозом. С помощью окраски хромосом установлено, что у пациентов с этим заболеванием определенный фрагмент исчезает на хромосоме 21 и появляется на конце длинного плеча хромосомы 9 (перенос фрагмента или **транслокация**, сокращенно t). Генетики обозначают такое событие как t(9; 21). Таким образом, хромосомный анализ свидетельствует о том, что разные молекулы ДНК могут обмениваться между собой отдельными участками, в результате чего в геноме образуются «гибриды», состоящие из молекул ДНК разных хромосом. Анализ уже изученных свойств хромосом позволил сформировать представление о полиморфизме генома человека.

Для выяснения локализации отдельных генов на хромосомах (то есть картирования генов) используют целый арсенал специальных зачастую весьма сложных по замыслу и исполнению методов. Один из основных – молекулярная гибридизация (образование гибрида) гена или его фрагмента с фиксированными на твердой подложке препаратами хромосом, выделенными из клеток в чистом виде (это называют гибридизацией *in situ*). Суть метода гибридизации *in situ* заключается во взаимодействии (гибридизации) между денатурированными (расплетенными) нитями ДНК в хромосомах и комплементарными нуклеотидными последовательностями добавленных к препарату хромосом, индивидуальных одонитевых ДНК или РНК (их называют **зондами**). При наличии комплементарности между одной из нитей хромосомной ДНК и зондом между ними образуются довольно стабильные молекулярные гибриды. Зонды маркируют предварительно с помощью разных меток (радиоактивных, флуоресцентных или др.). Места образования гибридов на хромосомах выявляют по положению этих меток на препаратах хромосом. Так, еще до появления методов генной инженерии и секвенирования ДНК выяснили, например, расположение в геноме человека генов, кодирующих большие и малые рибосомные РНК (рРНК). Гены первых оказались локализованными в пяти разных хромосомах человека (13, 14, 15, 21 и 22), тогда как основная масса генов малой рРНК (5S РНК) сконцентрирована в одном месте на длинном плече хромосомы 1.

Пример картины, получаемой при гибридизации меченых флуоресцентным красителем генов–зондов, приведен на рис. 10 на цветной вклейке.

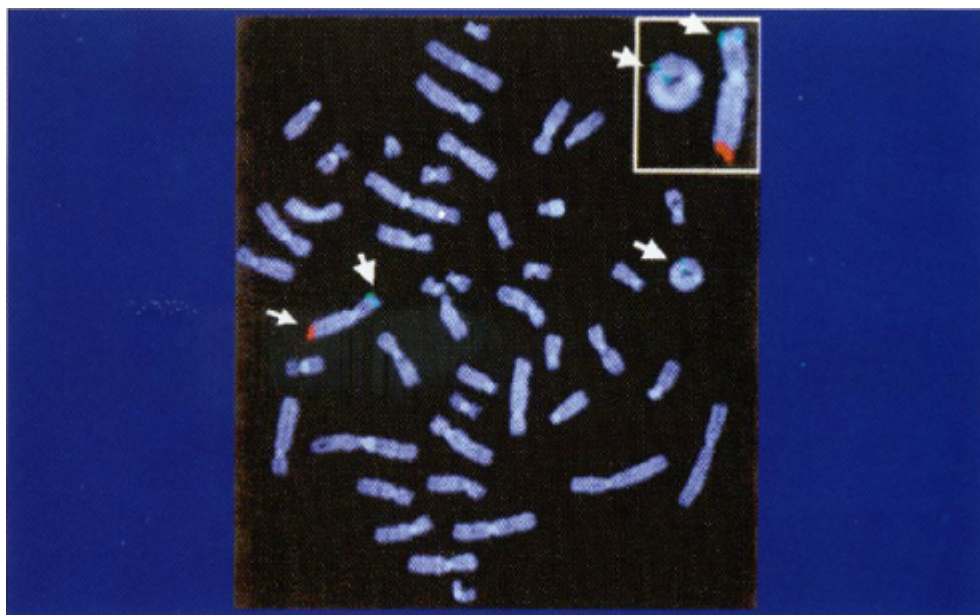


Рис. 10. Гибридизация хромосом человека с генами-зондами, мечеными красным и зеленым флюоресцентными красителями. Стрелками указано расположение соответствующих генов на концах двух разных хромосом (справа вверху дано увеличение картины гибридизующихся хромосом).

Гены, расположенные на одной хромосоме, определяют как сцепленные (связанные) гены. Если гены расположены на разных хромосомах, они наследуются независимо (независимая сегрегация). Когда же гены находятся на одной и той же хромосоме (т. е. сцеплены), они неспособны к независимой сегрегации. Изредка в половых клетках могут происходить различные изменения хромосом в результате рекомбинационных процессов между гомологичными хромосомами. Один из таких процессов получил название **кроссинговера**. Из-за кроссинговера сцепление между генами одной группы никогда не бывает полным. Чем ближе расположены друг к другу сцепленные гены, тем меньше вероятность изменения расположения таких генов у детей по сравнению с родителями. Измерение частоты рекомбинаций (кроссинговера) используется для установления линейного порядка генов на хромосоме внутри группы сцепления. Таким образом, при картировании хромосом первоначально устанавливают, находятся ли данные гены в одной и той же хромосоме, без уточнения, в какой именно. После того, как хотя бы один из генов данной группы сцепления локализуют в определенной хромосоме (например, с помощью гибридизации *in situ*), становится ясным, что все другие гены этой группы сцепления находятся в той же самой хромосоме.

Первым примером связи генов с определенными хромосомами может служить обнаружение сцепления определенных наследуемых признаков с половыми хромосомами. Чтобы доказать локализацию гена в мужской половой Y – хромосоме, достаточно показать, что данный признак всегда встречается только у мужчин и никогда не обнаруживается у женщин. Группа сцепления женской X-хромосомы однозначно характеризуется отсутствием наследуемых признаков, передающихся от отца к сыну, и наследованием признаков матери.

Особенно важным для изучения генома человека на первых этапах его исследования стал метод, называемый **гибридизацией соматических клеток**. При смешивании соматических (неполовых) клеток человека с клетками других видов животных (чаще всего для этой цели использовали клетки мышей или китайских хомячков) в присутствии определенных агентов может происходить слияние их ядер (гибридизация). При размножении таких гибридных клеток происходят потери некоторых хромосом. По счастливой для эксперимен-

таторов случайности в гибридных клетках человек–мышь происходит потеря большей части хромосом человека. Далее отбираются гибриды, в которых остается только какая–нибудь одна человеческая хромосома. Исследования таких гибридов позволили связать некоторые биохимические признаки, свойственные клеткам человека, с определенными хромосомами человека. Постепенно благодаря использованию селективных сред научились добиваться сохранения или потери отдельных хромосом человека, несущих определенные гены. Схема отбора, хотя и не очень проста на первый взгляд, довольно хорошо показала себя в эксперименте. Так, придумали специальную селективную среду, на которой могут выживать только те клетки, в которых синтезируется фермент тимидинкиназа. Если для гибридизации с клетками человека взять в качестве партнера мутантные клетки мыши, не синтезирующие тимидинкиназу, то будут выживать только те гибриды, которые содержат хромосомы человека с геном тимидинкиназы. Таким путем впервые удалось установить локализацию гена тимидинкиназы на хромосоме 17 человека.

Несмотря на то, что изучение генома человека на уровне хромосом дало ряд важных его характеристик, они были самыми общими и дали относительно мало для полного понимания устройства и функционирования генетического аппарата человеческих клеток.

Как двухметровая молекула умещается в микроскопическом ядре?

Как уже говорилось, молекулы ДНК, содержащиеся во всех хромосомах одной клетки человека, имеют общую длину около 2 метров. Это в среднем в миллион раз больше, чем диаметр клеточного ядра, составляющий порядка микрометра. Как же умещаются столь гигантские молекулы в таком маленьком ядре? Изучение структуры хромосом позволило ответить на этот вопрос. Оказывается, в ядре осуществляется «насильственная» упаковка молекул ДНК. Это достигается с помощью специальных механизмов, обеспечивающих изгибание двойной спирали ДНК. Существует несколько уровней «компактизации» ДНК в клетке (рис. 11).

Некоторые из особенностей упаковки ДНК изучены хорошо, а о некоторых существуют пока лишь приблизительные представления. Первый уровень компактизации заключается в накручивании нити ДНК, как нитки на катушку, на специальный комплекс ядерных белков (**гистонов**). Нить ДНК делает около двух оборотов вокруг одно



Рис. 11. Стадии упаковки ДНК в хромосомах (от двойной спирали до целых хромосом)

го комплекса, а затем снова около двух оборотов вокруг второго комплекса и т. д. В результате образуется структура, напоминающая бусы. Отдельные бусинки в этой структуре получили название **нуклеосомы**. На одной нуклеосоме размещается около 200 пар нуклеотидов ДНК. Между нуклеосомами остается фрагмент ДНК размером до 60 пар оснований, называемый линкером. Этот уровень укладки позволяет уменьшить линейные размеры ДНК в 6—7 раз.

На втором уровне компактизации «бусы» укладываются в спираль, состоящую из шести нуклеосом на виток. При этом линейные размеры ДНК уменьшаются в сумме до 1 мм, т. е. в 25—30 раз.

Третий уровень компактизации молекул ДНК изучен еще плохо. Скорее всего, это петельная укладка фибрилл – образование петельных доменов, которые под углом отходят от основной оси хромосомы (уплотнение в 680 раз). Их можно видеть в обычный световой микроскоп.

На последнем уровне компактизации ДНК происходит ее уплотнение примерно в 10000 раз.

Анализ суммарной ДНК – новые сведения о структуре генома человека

На первом этапе непосредственного исследования структуры генома человека, когда еще не существовала методология геномной инженерии, для изучения ДНК применяли традиционные физико-химические методы. В этих опытах использовали суммарные препараты ДНК, целиком выделенные из ядер клеток человека.

Пожалуй, первые сведения о молекулярной структуре генома человека были получены в результате центрифугирования в пробирке растворов ДНК в хлористом цезии при довольно высоких скоростях. В процессе вращения соли цезия создают тонкие слои раствора с различной плотностью (градиент плотности) вдоль пробирки и молекулы ДНК

перемещаются в этом градиенте, пока не достигнут такой области, где плотность солевого раствора будет точно такой же, как их собственная. А плотность ДНК сильно зависит от содержания АТ–и ГЦ–пар нуклеотидов, т. е., как говорят, от нуклеотидного состава. Оказалось, что основная масса ДНК человека после центрифугирования располагается преимущественно в одной зоне градиента (это соответствует среднему содержанию ГЦ–пар в геноме человека, равному 42%). Однако наряду с этим неожиданно обнаруживались и небольшие (минорные) полосы, в которых также содержались молекулы ДНК, но с иной плотностью и, следовательно, с иным содержанием нуклеотидных пар. Такие минорные, или дополнительные фракции ДНК получили название «сателлитных». Такое имя дали этим фракциям не случайно. В то время как раз был запущен первый советский спутник (лат. satellitis – спутник). Это и натолкнуло исследователей на такое название.

Вскоре после того, как была установлена двухспиральная структура ДНК, обнаружили, что при сильном нагревании ДНК две ее цепи расходятся (расплавляются) и ДНК из двуни-тевой превращается в однонитевую. Это приводит к нарушению ее естественной структуры, что получило отражение в названии данного процесса – **денатурация** ДНК. Однако при охлаждении денатурированной ДНК комплементарные цепи находят друг друга и соединяются строго так, как они располагались в исходной неденатурированной молекуле, по типу застешки – «молнии». Этот процесс получил название **рена-турации** или **реассоциации**. Сразу же после обнаружения этого явления оно было использовано экспериментаторами в целях изучения структуры генома.

Немного совсем простой математики для тех самых любопытных, кто недавно закончил ВУЗ, изучал химию и еще не забыл все, чему его учили. Процесс реассоциации ДНК во многом сходен с обычной химической реакцией второго порядка и, по этой причине, может быть описан довольно простой формулой:

$$C_t/C_0 = 1/(1 + k_2(C_0t)),$$

где C_0 и C_t – концентрации однонитевых ДНК соответственно в начальный (нулевой) момент времени и в момент времени t после начала реакции реассоциации, k_2 – константа скорости реакции второго порядка.

Для удобства изображения процесса реассоциации строят график, в котором по оси ординат откладывают долю реассоциированных молекул, а по оси абсцисс – величину C_0t . При таком изображении кривые кинетики реассоциации ДНК простейших организмов (вирусов и бактерий) имеют S – образный вид. Это соответствует кинетике химической реакции второго порядка. Если в эксперименте наблюдают отклонения от этой кривой, то это должно указывать на гетерогенность цепей ДНК по скорости взаимодействия друг с другом: одни реагирует быстрее, другие, соответственно, медленнее. Когда ДНК человека порезали на куски небольшого размера и измерили кинетику их реассоциации, то обнаружилось, что кривая, описывающая этот процесс, далека от стандартной (рис. 12).

Это явилось результатом того, что в ДНК человека имеются нук–леотидные последовательности, реассоциирующие с разной скоростью, а суммарная кривая реакции, наблюдаемая в опыте, отражает совокупность множества независимых реакций второго порядка. Когда были проведены эти эксперименты, уже существовал математический аппарат, позволяющий, хотя и очень грубо, вычленять из сложной суммарной кривой отдельные относительно однородные кинетические компоненты. Различия в скоростях реассоциации разных компонентов ДНК человека были связаны с разной представленностью в ДНК отдельных нуклеотидных последовательностей. Участки, которые присутствуют в геноме всего один раз, называли уникальными. Если в геноме определенная нуклеотидная последовательность не уникальна, а представлена неким числом одинаковых копий (т. е. повторяется, отсюда и название – повторяющаяся), то такая последовательность, естественно, по чисто химиче-

ским законам, будет в растворе находить комплементарную цепь и взаимодействовать с ней (реассоциировать) значительно быстрее первой.

В результате такого анализа нашли, что в геноме человека около 75% участков ДНК представлены 1 копией (уникальные) на гаплоидный геном (естественно, в ядре, являющемся диплоидным, имеется 2 копии каждой уникальной нуклеотидной последовательности). Остальную часть генома составляют повторяющиеся последовательности (повторы), среди которых 10% представлены очень быстро реассоциирующими повторами (10^4 и более копий на геном) и около 15% – умеренными повторами. Заметим сразу, что в дальнейшем эти оценки были существенно пересмотрены. Однако некоторые из этих результатов остались в силе до сих пор.

Первоначальные анализы показали, что среди быстро взаимодействующих друг с другом при реассоциации фрагментов ДНК присутствует некоторое количество таких, кинетика реассоциации которых отличается от реакции второго порядка. Причина этому оказалось в том, что эти участки представляют собой так называемые **обращенные повторы**, или **палиндромы** – взаимокплементарные последовательности, расположенные не на разных, а на одной нити ДНК.

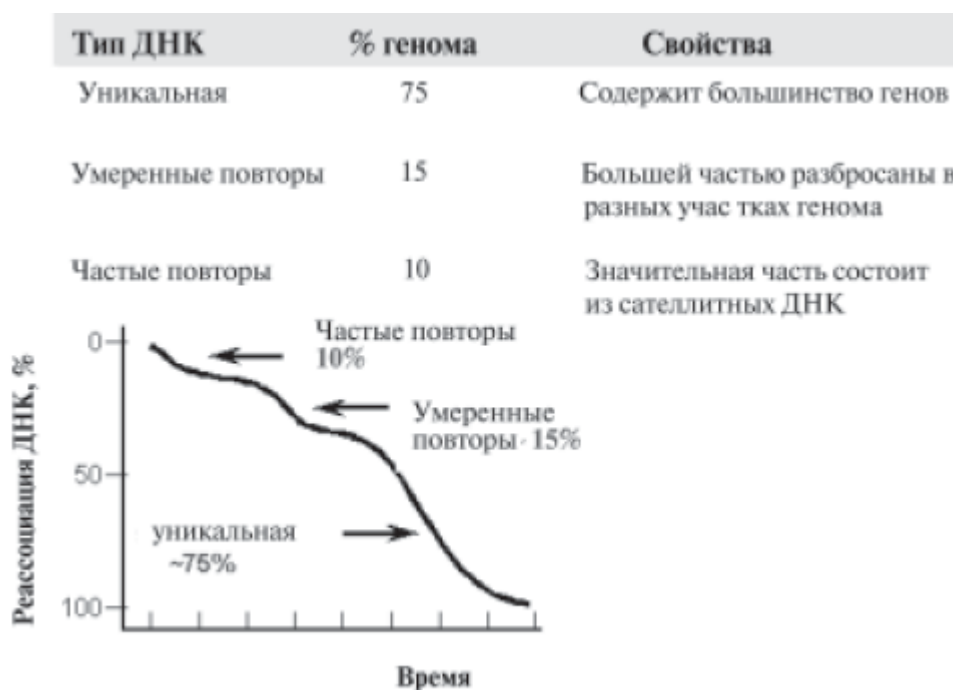


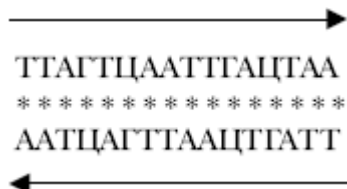
Рис. 12. Кинетика восстановления двунитевых молекул из искусственно разделенных комплементарных цепей ДНК человека (реассоциация). ДНК разбивают на небольшие фрагменты, денатурируют путем нагревания, а затем при охлаждении разошедшиеся цепи ДНК вновь соединяются. Чем чаще в смеси встречаются те или иные последовательности, тем быстрее они находят друг друга в растворе и реассоциируют. По этой кинетике определяли общее содержание повторяющихся и неповторяющихся (уникальных) нуклеотидных последовательностей в геноме человека

Таким образом, в ДНКовом тексте присутствуют «предложения» – палиндромы («перевертыши»), одинаково читаемые слева направо и справа налево. Перевертыши хорошо известны из литературы – это предложения, которые читаются одинаково слева направо и

справа налево без учета знаков препинания и интервалов между словами. В качестве примера приведем один из таких перевертышей:

УЖРЕДКОРУКОЮОКУРОКДЕРЖУ.

В ДНК перевертышами называют отрезки двойной антипараллельной спирали, которые имеют одинаковую нуклеотидную последовательность при чтении по обеим цепям в одинаковом направлении. Это выглядит как, например, в ниже приведенном случае:



Здесь стрелками показано направление цепей ДНК, а звездочками – водородные связи, образуемые между парами нуклеотидов. Общее число таких «перевертышей» в геноме человека оценено в интервале от 10^5 до 10^6 . При этом они относительно равномерно распределены по ДНК.

Имеются в геноме человека и нуклеотидные последовательности, которые на всем своем протяжении построены из одной единственной «буквы». Если в одной цепи ДНК эта буква А, то в другой цепи, соответственно, будет буква Т. Такие участки названы **гомополимерными**. В случае приведенного выше примера последовательность нуклеотидов записывается, как поли(А) – поли(Т). Таким образом, выяснилось, что в геноме человека (а параллельно это изучали и в геномах других организмов) имеются все варианты нуклеотидных последовательностей, состоящих из 4 «букв», которые только можно себе мысленно представить.

Используя различные модификации метода реассоциации ДНК, установили также, что в большей части генома человека повторяющиеся и уникальные нуклеотидные последовательности перемежаются друг с другом, а средние длины перемежающихся повторяющихся и уникальных фрагментов ДНК составляют соответственно 300 и 2000 п. н.

Приведенные выше данные были, конечно же, усредненными и поверхностно отражали общую картину устройства генома человека. Тем не менее они послужили хорошей основой для дальнейших более углубленных исследований. Важно, что одновременно такие же работы проводились и с использованием других эукариотических организмов. Многие моменты оказались сходными у разных высших организмов и растений. Так постепенно начали вырисовываться в общих чертах основные принципы организации генома человека.

Новый этап – анализ индивидуальных элементов ДНК (генная инженерия)

Мысль ученых продолжала интенсивно работать. И в результате в 1970–х годах были разработаны принципиально новые методы работы с ДНК. Ее научились резать на фрагменты, сшивать их друг с другом, а затем переносить в клетки и целые организмы. На практике это сводится к созданию в пробирке (т. е. *in vitro*) новых молекул ДНК, состоящих из фрагментов разного происхождения (так называемых **рекомбинантных ДНК**), которые в естественных условиях не могут возникнуть благодаря установленным и тщательно охраняемым Природой межвидовым барьерам. Датой рождения нового революционизирующего направления, получившего название **генной инженерии**, принято считать 1972 год, когда в США была создана первая рекомбинантная молекула ДНК.

Становление генной инженерии происходило очень непросто. Уже в самом начале (в 1973 году) участники Гордоновской конференции направили письмо президенту Национальной Академии наук США, в котором высказывали серьезные опасения о возможной опасности рекомбинантных ДНК для здоровья человека и окружающей среды. Прозвучал даже призыв к мораторию на некоторые виды генно-инженерных исследований. Два года спустя состоялась конференция в Асиломаре (США), где также шла речь об опасности генно-инженерных манипуляций и их ограничении. Один из основных исследователей структуры ДНК Э. Чаргафф, в частности, вопрошал: «Имеем ли мы право посягать необратимым образом на эволюционную мудрость миллионов лет только для того, чтобы удовлетворить амбицию и любопытство нескольких ученых?» Озабоченность ученых объяснялась теоретической возможностью создания в пробирке новых генетических структур, способных вызывать необычные формы инфекционных заболеваний человека, или созданием организмов, неблагоприятно воздействующих на окружающую среду. Однако как тогда, так и по прошествии трех десятков лет не существует никаких подтверждений тому, что рекомбинантные ДНК представляют опасность для человека. Современное состояние молекулярной генетики и как вершина ее развития – секвенирование генома человека – свидетельствуют о том, что генная инженерия полностью оправдала большие надежды ученых и поддерживающей ее общественности. Сейчас уже практически нет спора в отношении важности и необходимости использования генной инженерии для решения как научных, так и практических задач, стоящих перед человечеством.

Генная инженерия как принципиально новая технология возникла в значительной мере на базе достижений, которых ученые добились при изучении молекулярных процессов, которые происходят в бактериальной клетке. Было сделано несколько важных открытий, прежде чем исследователи научились легко и направленно манипулировать с ДНК, а не просто смешивать фрагменты ДНК с бактериями и отбирать среди них случайно возникающие варианты (реком-бинанты).

Бактерии имеют высокоэффективный механизм защиты от чужеродной ДНК. Исследование этой проблемы привело к обнаружению в бактериях специальных ферментов, получившие название **рестриктаз**. Первые из них получены еще в 1970 г. Сейчас уже выделено несколько сот разнообразных рестриктаз из клеток разных видов бактерий. Каждый из таких ферментов «узнает» строго определенные короткие (обычно 4—6 п. н.) нуклеотидные последовательности ДНК, которые чаще всего представляют собой палиндромы, и разрезает ДНК на фрагменты по этим участкам. В результате этих свойств рестриктазы стали незаменимым инструментом в исследовании генома, своеобразным скальпелем для ДНК.

Большой удачей для экспериментаторов было то, что в большинстве случаев при нарезании рестриктазами ДНК образующиеся фрагменты содержат «липкие» концы, т. е. способны в дальнейшем соединяться друг с другом. Для восстановления химической связи таких слипшихся фрагментов стали использовать еще один новый класс ферментов – **лигазы**. Так появились первые инструменты, позволяющие целенаправленно манипулировать с молекулами ДНК *in vitro*. Следующий важный этап – размножение сконструированных рекомбинантных ДНК, чтобы можно было с ними работать дальше. Для этой цели стали использовать специальные ДНК – **векторы**, которые представляют собой или бактериофаги, или внехромосом-ные кольцевые молекулы – **плазмиды**, которые обладают способностью размножаться в бактериальных клетках. Если любой фрагмент ДНК внести в состав ДНК бактериофагов или плазмид, то при переносе в бактериальные клетки он будет там размножаться вместе с вектором. Наконец, были разработаны специальные подходы для отбора бактериальных клеток, в которых содержатся необходимые ре-комбинантные ДНК. Все эти достижения позволили проводить **клонирование** ДНК, т. е. выделять из сложной смеси молекул определенные участки ДНК и размножать их.

Клонирование фрагментов ДНК дало в руки исследователей принципиально новый подход к выделению из сложной смеси ДНКовых фрагментов конкретных индивидуальных последовательностей. При этом они могли содержать осмысленные тексты (гены) или не содержать таковых. Традиционная биохимия не могла обеспечить подобную процедуру в силу того, что различные фрагменты ДНК химически очень сходны. Но генной инженерии это оказалось вполне под силу. Схематически процесс клонирования ДНК изображен на рис. 13.

После разрезания ДНК на фрагменты с помощью рестриктазы все они соединяются ковалентно с вектором (вектор разрезан той же рестриктазой) в результате «слипания» концов молекул и последующей обработки «гибридов» лигазой. Так в пробирке создается искусственная или рекомбинантная ДНК. Затем осуществляют перенос множества образовавшихся рекомбинантных молекул в бактериальные клетки – **трансформацию**. Особенность процесса трансформации заключается в том, что в одну бактериальную клетку попадает лишь одна рекомбинантная молекула ДНК. После этого, благодаря отбору, размножаются только бактерии с рекомбинантными ДНК, содержащими требуемые нуклеотидные последовательности.

Для изучения структуры генома необходимо не только клонировать все его участки в виде рекомбинантных молекул, но, главное, расшифровать в них последовательность нуклеотидов, т. е. «прочитать» ДНКовый текст.

Важным моментом во всей этой генно–инженерной «кухне» было создание принципиально новой технологии размножения индивидуальных фрагментов ДНК *in vitro*. Этот удивительно простой метод получения фрагментов ДНК в неограниченном количестве копий был придуман при необычных обстоятельствах – ночью, во время автомобильной поездки в горах Калифорнии. Автор этой идеи нобелевский лауреат Кэри Б. Мюллис так вспоминает момент озарения. «Иногда удачная идея приходит в голову совершенно неожиданно. Со мной, например, это случилось в одну из апрельских ночей в 1983 г., когда я, сидя за рулем автомобиля, пробирался по освещенной луной горной дороге в секвойные леса Северной Калифорнии. Мысль моя случайно натолкнулась на процесс, благодаря которому



Рис. 13. Схема процесса клонирования ДНК с использованием в качестве вектора бактериальной плазмидной ДНК

можно получать копии генов в неограниченных количествах. Теперь он называется полимеразной цепной реакцией». За создание новой технологии ее автор получил Нобелевскую премию.

Полимеразная цепная реакция (сокращенно ПЦР) дает возможность в течение дня из одной молекулы ДНК получать 100 миллиардов идентичных по структуре молекул. Эта реакция довольно проста в исполнении: нужны лишь ДНК, пробирка, несколько реагентов и источник нагревания и охлаждения. Препарат ДНК, который необходимо копировать, может

быть получен и из кусочка ткани, и из капли засохшей крови, и из засушенной мумии, и даже из тела мамонта, пролежавшего несколько тысяч лет в вечной мерзлоте. Суть этого метода изображена на рис. 14 на цветной вклейке.

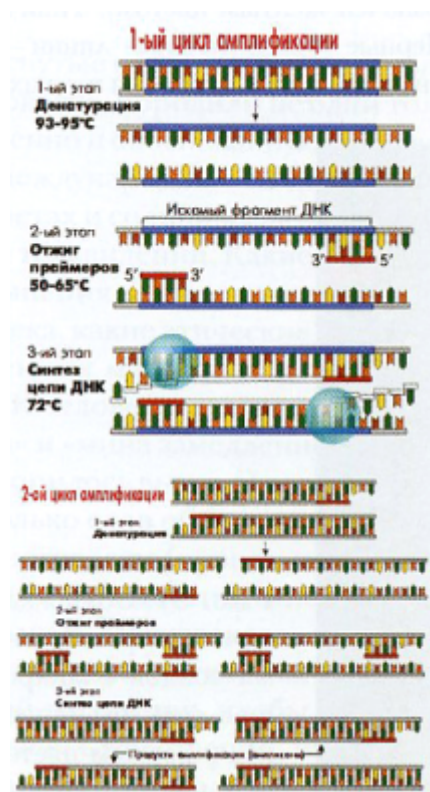


Рис. 14. Схема протекания полимеразной цепной реакции (ПЦР).

Сама процедура ПЦР заключается в следующем. Прежде всего, ДНК денатурируют, нагревая ее до 98°C. При этом две цепи ДНК, относительно слабо связанные между собой, разделяются друг от друга. Далее в ход идут так называемые затравки (праймеры) – короткие одонитевые фрагменты ДНК, которые комплементарны краям того участка генома, который собираются размножить (амплифицировать). Эти затравки гибридизуются с комплементарными участками одонитевых молекул ДНК, после чего специальный фермент (ДНК-полимераза) удлиняет затравки, строя молекулу ДНК по матрице (цикл 1). Затем продукты реакции вновь денатурируют и процедуру повторяют необходимое число раз. За 20—30 циклов из одной молекулы можно получить миллионы копий ДНК. Этот процесс подобен цепной ядерной реакции, но осуществляется в пробирке с помощью специального фермента – термостабильной ДНК-полимеразы. По этой причине он и получил название полимеразной цепной реакции.

Конечная цель – определение полной последовательности нуклеотидов в ДНК человека

Длительное время искали эффективные методы, позволяющие определять последовательность нуклеотидов в длинных молекулах ДНК (такое определение нуклеотидной последовательности получило название **секвенирование**, от англ. sequence – последовательность). Только в середине 70-х годов секвенирование протяженных участков ДНК стало возможным. Связано это было с изобретением принципиально новых подходов. В 1976 г. А. Максом и У. Гилбертом был разработан метод прямого секвенирования, основанный на химической деградации ДНК. В данном случае осуществляется специфическая химиче-

ская фрагментация длинной цепи ДНК (поли–нуклеотида), радиоактивно меченной с одного конца. Затем препарат меченой ДНК разделяют на четыре порции и каждую из них обрабатывают реагентом, модифицирующим одно или два из четырех оснований, содержащихся в ней. После разделения в специальном геле меченых фрагментов по размерам (с помощью электрофореза) на рентгеновских пленках смотрят, что при этом происходит с нуклеотидной последовательностью, и на основании этого делают вывод о порядке расположения нуклеотидов друг за другом в каждом фрагменте ДНК. Однако метод оказался довольно сложным. В становлении этого метода на его начальных этапах существенную роль сыграл российский ученый академик А. Д. Мирзабеков, работавший в лаборатории У. Гилберта.

Чуть раньше английский ученый Фред Сэнгер предложил иной способ расшифровки структуры ДНК, который в конечном итоге после разных модификаций стал основным. Согласно методу Сэнгера молекулу ДНК с помощью специальной обработки ферментами не только расщепляют на фрагменты, но и «расплетают» ее двойную спираль на две нити. Потом по каждому из полученных обрывков «нитей» с помощью специальных затравок восстанавливают недостающую вторую нить ДНК. При этом синтез второй нити осуществляется не полностью, а с помощью специального приема обеспечивается прекращение синтеза строго на одном из 4 нуклеотидов. В результате этого получался набор цепей ДНК с разной длиной – «лесенка». Эта «лесенка» становится видимой глазом после разделения всех фрагментов по размерам за счет того, что концы всех фрагментов содержат специфическую – или радиоактивную, или флуоресцентную метку. Следует отметить, что в разработку этого поистине революционизирующего метода существенный вклад был сделан и двумя российскими учеными: профессором С. К. Василенко и академиком Е. Д. Свердловым.

За разработку методов секвенирования Гилберт и Сэнгер получили Нобелевскую премию. Интересно, что для Сэнгера эта была уже вторая премия, первую он получил за расшифровку аминокислотной последовательности первого белка – инсулина.

Все автоматы–секвенаторы, которые появились позднее, были построены по принципу метода Сэнгера, поскольку он оказался более удобным для автоматизации процесса и компьютерной регистрации результатов. В настоящее время ряд крупных фирм занимается производством таких автоматов, а также стандартных наборов реактивов для анализа ДНК. Секвенирование стало достаточно рутинной лаборантской работой. А метод Максама–Гилберта имеет сейчас скорее историческое, чем практическое значение.

Во всех случаях основное правило заключается в том, чтобы в препарате ДНК, который собираются секвенировать, все фрагменты имели, по крайней мере, один одинаковый конец, т. е. одну точку отсчета. В этот фиксированный конец вводят радиоактивную (или флуоресцентную) метку. После этого проводят дальнейшие специфические расщепления молекул с помощью ферментов или химических агентов по строго определенным азотистым основаниям. Условия расщепления при этом таковы, что на каждый меченый фрагмент в среднем приходится один разрыв. В результате после расщепления образуется набор фрагментов разной длины, каждый из которых начинается в точке отсчета, а заканчивается на или перед определенным типом нуклеотида.

Полученные фрагменты разделяют по размерам с помощью электрофореза в гелях. Маленькие фрагменты движутся быстрее более крупных, и в результате такой процедуры все фрагменты выстраиваются друг за другом в зависимости от своей длины. Результат такого разделения изображен схематически на рис. 15.

Так осуществляют секвенирование небольших фрагментов ДНК (не более 1000 нуклеотидов). Но геномы даже самых примитивных организмов имеют размер свыше 500 000 п. н. Как же их секвенируют? Для этого разработано несколько стратегий. Не будем вдаваться в подробности, но укажем на некоторые основные из них.

Секвенирование простых геномов

Для определения нуклеотидной последовательности (т. е. первичной структуры) конкретного района ДНК в первую очередь необходимо упростить ее, что достигается путем разрезания ее на относительно короткие фрагменты. Сделать это можно, например, с помощью специальных «скальпелей» для ДНК – ферментов рестриктаз, о которых уже шла речь выше.

При секвенировании простейших организмов, у которых геном относительно невелик, обычно используют процедуру, называемую условно «сверху вниз». Всю ДНК «разрезают» на кусочки с помощью уже упоминавшихся выше ферментов рестриктаз, затем секвенируют эти кусочки по отдельности, а после «склеивают» из них полный геном. «Склеивание» оригинала осуществляется за счет того, что нуклеотидные последовательности разных кусочков перекрываются друг с другом, т. е. одинаковы по концам. Эта методология получила название «дробовика». Суть данной процедуры отражена на рис. 16.

Специальные реакции в пробирках, приводящие к остановке синтеза на ДНК-матрице новых фрагментов ДНК на определенных нуклеотидах ДНК



Электрофоретическое разделение в геле по размерам вновь синтезируемых фрагментов ДНК



Авторадиограмма геля, с которого прямо читают ДНКовый текст

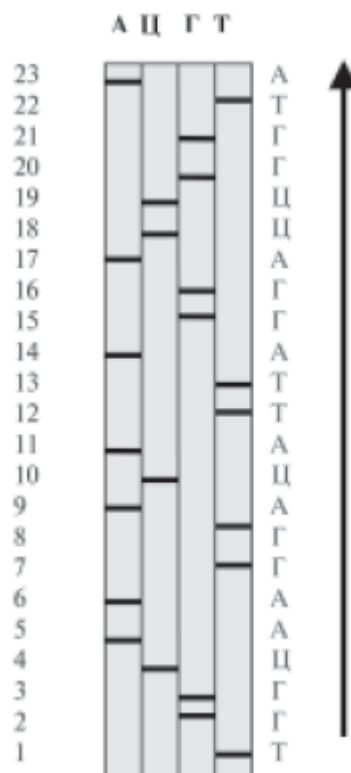


Рис. 15. Схема, поясняющая процесс подготовки и прочтения ДНКового текста

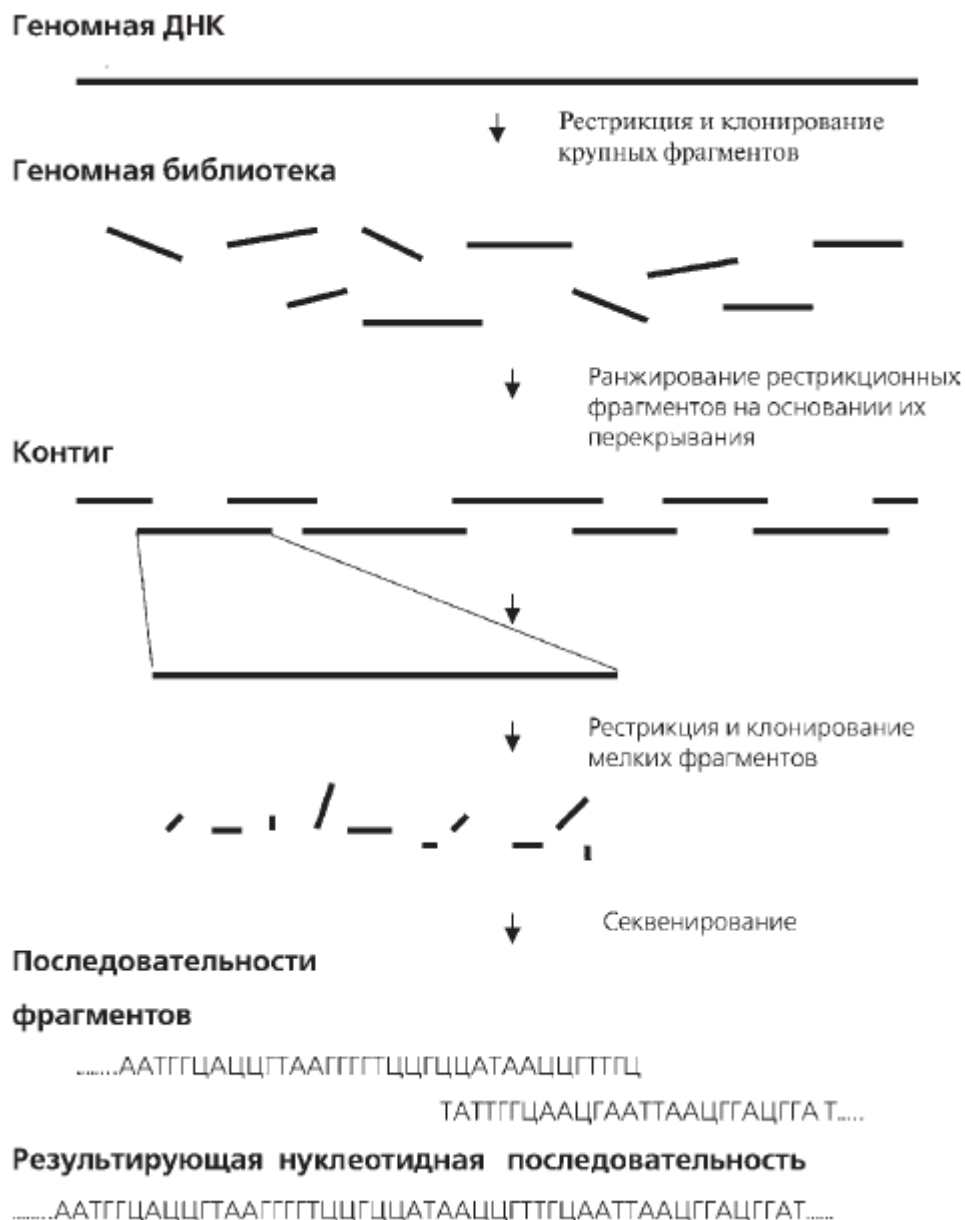


Рис. 16. Схема стратегии «дробовика», используемая для секвенирования больших молекул ДНК

Однако в случае такого очень сложного генома, как геном человека, начали с другого, а именно с определения положения известных генов и других генетических маркеров на отдельных хромосомах, то есть с генетического картирования генома. Подобную задачу генетическими экспериментами на более простых объектах пытался решить еще Т. Морган, который за свои работы получил Нобелевскую премию в 1933 году. Теперь появились более эффективные методы. Один из них, называемый методом «радиационных гибридов», заключается в следующем. Клетки человека, растущие вне организма в питательной среде, облучают рентгеном, что приводит к гибели клеток в результате разрыва хромосомной ДНК на куски, содержащие от 2,5 до 25 млн. п. н. Но прежде, чем убитые облучением клетки распадутся, их сливают с клетками хомяка, в результате чего в разные клетки хомячка попадают разные наборы фрагментов ДНК человека. Затем «гибридные» клетки размножают в культуре, при этом в них наряду с собственной ДНК реплицируются и фрагменты чужеродной ДНК. Затем определяют состав известных генов и других генетических маркеров в каждой

клеточной линии и, статистически обработав полученные данные, выводят наиболее вероятное их взаимное расположение в хромосомах. В качестве генетических маркеров использовали как гены, так и фрагменты ДНК с неизвестной функцией. Для картирования хромосом важным свойством маркеров являлся их полиморфизм, т. е. существование разных форм среди индивидуумов.

Так были построены первые генетические карты генома человека, на которых первоначально были отмечены различные генетические маркеры, отстоящие друг от друга на расстоянии не более 2 миллионов нуклеотидных пар (млн. н. п.).

Затем были составлены физические карты хромосом: первоначально с разрешением 0,1 млн. н. п., а затем 0,001 млн. н. п. Для этой цели на первом этапе применяли методы окрашивания хромосом и гибридизации с хромосомами *in situ*. И лишь позднее использовали рес-триктазы. С помощью этих удивительно точно работающих ферментов «дробили» ДНК по строго определенным участкам на миллионы перекрывающихся между собой по нуклеотидной последовательности фрагментов, «разбирали» каждый из них по отдельности, после чего из них «склеивали» оригинал. «Склеивание» проводили на основании перекрывания фрагментов по нуклеотидной последовательности. Так постепенно шли все выше и выше. Поэтому данная стратегия получила название «снизу вверх». Со всей очевидностью можно сказать, что решалась грандиозная по масштабам и сложности задача. И решить ее ученые смогли только с помощью суперкомпьютеров. В результате были созданы физические карты разных областей ДНК и целых хромосом, состоящие из последовательно перекрывающихся друг с другом фрагментов. Набор таких «родственных» фрагментов получил название **контиг** (рис. 16).

А далее наступила очередь секвенирования отдельных рестрик-ционных фрагментов. Это привело в конечном итоге к построению секвенсовых карт, на которых степень разрешения была доведена до своего максимального значения. Если 20 лет назад расшифровка нуклеотидной последовательности ДНК длиной 1000 п. н. считалась большим научным достижением (за это можно было сразу получить степень доктора наук), то уже к 1990 году секвенирование ДНК стало массовой технологией. А сейчас квалифицированный лаборант проделывает такую работу всего за несколько часов.

После появления эффективных методов секвенирования ДНК и нескольких стратегий использования этого метода стал стремительно нарастать вал расшифрованных нуклеотидных последовательностей, в первую очередь таких простых организмов, как вирусы, а также отдельных клонированных фрагментов ДНК различных видов высших организмов. Так, еще в конце 1970-х годов была расшифрована структура первого живого объекта – вируса бактерий – бактериофага, обозначаемого как фХ 174, имеющего длину 5386 п. н. Затем последовала очередь других.

В конце 80-х гг. были начаты крупные международные проекты, целью которых было полное секвенирование геномов бактерий, грибов, дрожжей, дрозофилы, мыши, пшеницы, человека и др. Первоначально была определена первичная структура ДНК микроорганизмов с размерами генома до 20 млн. п. н., к концу 1998 г. их число составило уже 18 (см. табл. 2). Наименьший размер генома у свободно живущих организмов (например, у бактерии *Mycoplasma genitalium*) составляет лишь около 600 000 п. н. У этой бактерии содержится всего 500 генов, причем 150 из них (если их удалять по одиночке) никак не влияют на её жизнеспособность. Так возникло предположение, что элементарная «машина жизни» теоретически должна работать при наличии всего лишь 350 генов.

Первые объекты для секвенирования были выбраны неслучайно. Большинство из этих микроорганизмов (археобактерии, спирохеты, хламидобактерии, кишечная палочка, возбудители пневмоний, сифилиса, гемофилии, метанобразующие бактерии, микоплазмы, риккетсии, цианобактерии) способны вызывать различные патологии у человека. В настоящее

время многие из этих проектов уже завершены; исследовано свыше 800 полных клеточных геномов микоплазм, археобактерий, кишечной палочки, возбудителей ряда болезней человека, а также пекарских дрожжей, маленького червя–нематоды, дрозофилы и весьма интересного в практическом плане растения арабидопсиса. Весьма вероятно, что истинное число секвенированных к настоящему времени геномов гораздо больше, потому что многие фармацевтические фирмы засекречивают свои результаты, не публикуя их в открытой печати.

Таблица 2. Некоторые микроорганизмы, геномы которых полностью секвенированы к настоящему времени

Организм (латинское видовое название)	Размер генома (млн. п. н.)	Количество генов
<i>Archaeoglobus fulgidus</i>	2,18	2436
<i>Aquifex aeolicus</i>	1,55	1512
<i>Bacillus subtilis</i>	4,20	4100
<i>Borrelia burgdorferi</i>	1,44	853
<i>Chlamydia trachomatis</i>	1,06	895
<i>Escherichia coli</i> K-12	4,60	4288
<i>Haemophilus influenzae</i>	1,83	1740
<i>Helicobacter pylori</i> 199	1,64	1495
<i>Helicobacter pylori</i> 26695	1,67	1590
<i>Methanobacterium thermoautotrophicum</i>	1,75	1855
<i>Methanococcus jannaschi</i>	1,66	1738
<i>Mycobacterium tuberculosis</i>	4,40	4033
<i>Mycoplasma genitalium</i>	0,58	468
<i>Mycoplasma pneumoniae</i>	0,82	716
<i>Pyrococcus horikoshii</i>	1,74	2061
<i>Rickettsia prowazekii</i>	1,12	834
<i>Synechocystis</i> sp.	3,57	3168
<i>Treponema pallidum</i>	1,14	1041

ТАЙНА ТАИН ЗА 6 МИЛЛИАРДОВ ДОЛЛАРОВ

Лучше изучить лишнее, чем ничего не изучить.
Сенека Старший

*Кто хочет много знать,
тому надо мало спать.*
Русская пословица

И вот дошла очередь до расшифровки сокровеннейшей тайны человека – ДНКового текста его генома, который некоторые называют Книгой жизни или Энциклопедией человека.

Напомним еще раз, как все это начиналось и происходило. В 1988 году был создан проект «Геном человека» Национального института здоровья США. Главой работ по полному секвенированию генома человека стал нобелевский лауреат Джеймс Уотсон. Одновременно в России с этой же идеей выступил академик А. А. Баев. Государственная программа «Геном человека» была принята в этом же году в Советском Союзе. Позднее к программе подключились другие страны, и широкомасштабные координированные исследования стали проводиться под эгидой международной организации Human Genom Organisation (HUGO). На работу по секвенированию генома человека, осуществленную этим мировым консорциумом, было потрачено в сумме более 3 миллиардов долларов. В России членами этой организации являются сегодня около 80 российских ученых (председатель Российского совета по геному человека – академик Л. Л. Киселев). Из 24-х хромосом генома человека ученые России в основном концентрировались на 3-й, 13-й и 19-й хромосомах. При этом непосредственно секвенированием ДНК они занимались очень мало, основное внимание было обращено на структурно-функциональные исследования генома. В материальном отношении наш вклад был настолько скромным, что сейчас при общих подсчетах его вообще не учитывают.

Первоначально огромный объем предстоящей работы очень пугал многих исследователей. И действительно ситуация на первых этапах не была слишком обнадеживающей. В первые 2 года работы проекта «Геном человека» скорости секвенирования были еще очень низкими. Для полного завершения работы такими темпами потребовалось бы около 100 лет. Это удручало одних исследователей, но не останавливало других. Последним стало очевидно, что решение данной задачи невыполнимо традиционными методами и необходим поиск новых подходов и технологий расшифровки нуклеотидных последовательностей фрагментов ДНК, а также создание новой вычислительной техники и оригинальных компьютерных программ. Все это, конечно, было невыполнимо в рамках отдельно взятого государства. По этой причине к проекту были привлечены около 20 различных стран, организован международный банк данных, куда поставлялась вся информация, полученная в процессе решения данной задачи. В результате технических новаций постепенно возрастала производительность аппаратуры, стали использовать промышленные роботы, многие процессы были автоматизированы. И вскоре скорость секвенирования достигла 10 млн. нуклеотидных пар в сутки! Теперь уже выполнение проекта стало реальностью. За 1995 г. длина участков ДНК человека с установленной последовательностью нуклеотидов увеличилась почти в 10 раз, но это составляло еще менее 0,001% от всего генома человека. К началу 1998 г. было секвенировано всего около 3% генома.

Примерно в это время к работе по секвенированию ДНК человека неожиданно присоединилась частная американская компания из штата Мериленд «Celera Genomics» под руко-

водством Крега Венте-ра, которая объявила, что закончит свою работу на 4 года раньше международного консорциума. Это сильно подхлестнуло всех участников.

Следует отметить, что два конкурирующих коллектива – HUGO и Celera – использовали в своей работе разные подходы к секвенированию ДНК. Эту разницу можно сравнить с процессом создания подробной маршрутной карты. Можно описать подробнейшим образом каждую горку, каждую канавку, каждый бугорок. А можно описать весь путь, но при этом отметить только наиболее важные участки. Международный консорциум применил в своей работе традиционный метод, который заключался в том, что «кусочки» молекул ДНК выстраивали друг за другом: сначала делали карты хромосом, а потом по ним «прогуливались». Таким образом, ученые теряли время на подготовке материала, картировании, но выигрывали в секвенировании – не нужно было помногу раз делать одно и то же. Руководитель компании «Селера» Крег Вентер предложил ускоренный вариант, так называемый «шрапнельный» метод расшифровки структуры целого генома. Этот метод предусматривает разделение всего генома на небольшие фрагменты ДНК без предварительного их картирования, а затем расшифровку последовательностей с использованием высокопроизводительных автоматов и «сборку» фрагментов в нужном порядке с помощью суперкомпьютера с учетом соответствия перекрывающихся нуклеотидных последовательностей на концах фрагментов. Все это напоминает сборку гигантского неизвестного паззла по отдельным совершенно абстрактным его фрагментам. Метод, конечно, казался очень простым, но при этом одни и те же «кусочки» неизбежно попадали в аппарат-секвенатор по 10, а то и по 20 раз. «Celera», конечно же, теряла на том, что по многу раз секвенировала одни и те же фрагменты, поскольку до начала их секвенирования ничего о них не знала. Это казалось многим непроизводительной тратой средств и времени. Но в конечном итоге в результате использования своего подхода «Селера» сильно сэкономила время и деньги на предварительной подготовке материала, так как нужно было только раздробить ДНК на фрагменты и не заниматься трудоемким процессом их предварительного картирования. Кроме того, «Селера» почти с самого начала использовала специальный робот, который способен автоматически расшифровывать ДНК. Этот робот обладал способностью метить разные «буквы» ДНКового текста различными флюоресцирующими красками, а затем с помощью лазера считывать последовательности цветов, как последовательность букв. Огромный объем информации обрабатывался с помощью уникального суперкомпьютера. Celera по-латински обозначает «быстрая». И компания оправдала свое название. Еще в 1995-м году на основе предложенного Вентером подхода было осуществлено секвенирование генома бактерии *Haemophilus influenzae*. А когда «Селера» приступила к секвенированию генома человека, то и здесь добилась большого успеха, быстро встав по числу секвенированных нуклеотидных последовательностей генома человека вровень с международным консорциумом.

Началась беспрецедентная в науке гонка. Два коллектива работали независимо, не жалея сил, чтобы прийти к финишу первыми. Технология все время совершенствовалась. В результате жесткой конкуренции была создана, можно сказать, новая индустрия по секвенированию ДНК. В Англии, США, Франции, Германии, Японии появились гигантские автоматизированные фабрики – геномотроны, – секвенирующие по несколько миллионов нуклеотидов в неделю.

И вот произошло долгожданное событие: в декабрьском номере журнала Nature за 1999 год появилась статья под названием «Нуклеотидные последовательности первой хромосомы человека». В этой статье коллектив, состоящий из более чем двухсот авторов, сообщил о полном секвенировании одной из самых малых хромосом человека – хромосомы под номером 22. Выход этой статьи окончательно подтвердил тот факт, что проект «Геном человека» будет выполнен. Произошедшее событие некоторые ученые сравнили с первым взглядом на ландшафт новой планеты.

Между разными участниками гонки, призом в которой была полная последовательность нуклеотидов генома человека, с самого начала возникло еще одно существенное противоречие. Суть его заключалась в том, что Вентер и его компания «Селера» настаивали на праве индивидуального владения всеми данными, полученными при установлении последовательности нуклеотидов, и возможности получать коммерческую выгоду от всех видов использования этих материалов. Руководители международного проекта считали, что результаты расшифровки генома должны быть доступны всем ученым и использоваться ими без всяких ограничений. Спор достиг такого накала, что в это вынуждены были вмешаться даже главы разных правительств. Весной 2000 года президент США Билл Клинтон и британский премьер-министр Тони Блэр выступили с совместным заявлением, в котором поддержали позицию руководителей международного проекта «Геном человека». В конечном итоге консенсус был достигнут, и в июне 2000 года два конкурирующих коллектива – «Celera Genomics» и международный консорциум, – объединив свои данные, официально объявили о том, что их совместными усилиями в целом завершено секвенирование генома человека, создан его черновой вариант.

Проект «Геном человека» оказался одним из наиболее дерзновенных, дорогостоящих и потенциально важных проектов в истории науки. Согласно первоначальным расчетам, работа по расшифровке генома человека должна была окончиться в 2005-м году и стоить в общей сложности 3 миллиарда долларов. Но жизнь внесла свои существенные поправки, и в конечном итоге проект реализовался на 5 лет быстрее, но на его реализацию потребовалось уже примерно 6 миллиардов долларов, т. е. чуть меньше чем по 2 доллара за каждую пару нуклеотидов. Дополнительные 3 миллиарда долларов были добавлены частным капиталом присоединившейся позднее компании Celera Genomics. Научные публикации чернового варианта структуры генома человека появились в феврале 2001 года. Качество секвенирования достаточно высокое и предполагает в среднем наличие только одного ошибочно определенного нуклеотида на протяжении более 50 тыс. п. н.

Итак, на сегодняшний день в основном геном человека секвенирован, т. е. определен порядок расположения нуклеотидов во всех молекулах ДНК на всех хромосомах. Текст этот огромен. По последним данным он состоит из 3,2 миллиардов «букв». Вся эта информация теперь содержится в сверхмощных компьютерах (книга с этим текстом не издана и едва ли будет издана когда-либо). Вокруг этого события сразу же поднялся невероятный шум и ажиотаж. Но только специалисты понимали до конца и тот огромный успех, который был достигнут в результате реализации первого этапа программы, и сверхтрудность тех задач, которые еще предстоит решить человечеству до достижения конечной цели. Важно понимать, что все произошедшее не означает автоматически, что геном человека расшифрован, что мы можем уже этот текст читать как любую другую книгу без всяких проблем. Секвенирование (определение последовательностей нуклеотидов) не приводит автоматически к пониманию того, что написано (т. е. собственно говоря, к расшифровке). С имеющимся текстом предстоит еще очень долго разбираться, в частности, правильно расставить знаки пунктуации, выверить опечатки и др. Где-то они уже намечены, а где-то еще предстоит их проставить. Расшифровать нуклеотидную последовательность – это все равно, что читать книгу, просто произнося названия букв подряд. Найти ген значит понять, как буквы складываются в слова. Но нужно еще понять и смысл фразы. И это теперь основная проблема. Кроме того, следует еще добавить, что ДНКовый текст написан на «мертвом» языке, в расшифровке которого науке помочь попросту некому – надеяться нужно только на себя. Как воображение кошки во дворце английской королевы не поднялось выше мышки, так, вероятно, обстоит дело и с нашим сегодняшним пониманием генома. Все богатство этой Энциклопедии нам еще предстоит осознать.

У читателя может возникнуть вполне естественный вопрос: геном какого же человека определен в результате этих титанических усилий, кто этот конкретный человек? Четкого ответа пока нет, но ясно, что это не один человек, а, по крайней мере, 7—10 различных людей. Согласно имеющимся данным, фирма Celera в основном ориентировалась на геном одного человека, о котором известно лишь, что это белый мужчина среднего возраста. Скорее всего, это был сам глава корпорации Крег Вентер, хотя другие источники утверждают, что Celera использовала 5 человек (3 мужчин и 2 женщины), относящихся к разным расам — афро-американской, монголоидной и европеоидной. Международный консорциум использовал в своей работе материал не менее семи различных людей. В любом случае каждый из двух соперничавших коллективов брал для анализа ДНК из своего источника. А в результате опубликована усредненная нуклеотидная последовательность одного человеческого генома.

На основании полученных данных мы сегодня имеем общую характеристику нуклеотидных последовательностей ДНК, образующих геном человека, которая не будет существенно пересмотрена в дальнейшем. Можно ожидать, что множество деталей еще будет уточняться и меняться, но общая характеристика безусловно останется неизменной. Это сродни догмату папской непогрешимости: *Roma locuta res finita est* (Рим сказал свое слово, дело окончено).

Теперь вспомним основные этапы, по которым двигалась наука на пути к достижению полного секвенирования генома человека.

ОСНОВНЫЕ ВЕХИ ГЕНЕТИКИ И ГЕНОМИКИ

Inventas vitamjuvat excoluisse per artes.

Изобретения улучшают жизнь, искусство украшает ее.

Надпись на нобелевской медали, фраза из «Энеиды» Вергилия

1865 год

Открытие Г. Менделем (1822—1884) факторов наследственности и разработка гибридологического метода, т. е. правил скрещивания организмов и учета признаков у их потомства.

1868 год

Ф. Мишер из спермы лосося выделил фосфорсодержащее вещество, происходящее из клеточных ядер, которое он назвал **нуклеином** (теперь его называют дезоксирибонуклеиновой кислотой).

1871 год

Ч. Дарвин публикует свою книгу «Происхождение человека и половой отбор».

1875 год

Ф. Гальтон демонстрирует возможность использования близнецов для изучения относительного влияния на организм наследственности и окружающей среды.

1900 год

Формальное рождение генетики как науки. Независимая публикация статей Г. де Фриза, К. Корренса и Э. Чермака с изложением основных законов наследования. «Переоткрыты» и стали известны широкой научной общественности исследования Г. Менделя.

1902 год

В. Саттон и Т. Бовери независимо создают хромосомную теорию наследственности.

1905 год

У. Бэтсон предлагает слово «генетика» (от греч. γέννωμαι – порождать) для нового направления науки.

1910 год

Т. Г. Морганом установлено, что гены расположены в хромосомах в линейном порядке, образуя группы сцепления. Морган установил также закономерности наследования признаков, сцепленных с полом (Нобелевская премия 1933 г. по физиологии и медицине за экспериментальное обоснование хромосомной теории наследственности).

А. Кёссель получил Нобелевскую премию по химии за установление того, что в состав ДНК входят четыре азотистых основания: аде-нин, гуанин, цитозин и тимин.

1922 год

Н. И. Вавилов сформулировал «закон гомологических рядов» – о параллелизме в изменчивости родственных групп растений, то есть о генетической близости этих групп. Закон Вавилова установил определенные правила формообразования и позволил предсказывать у данного вида еще не открытые, но возможные признаки (аналогия с системой Менделеева).

1925 год

Г. А. Надсон, Г. С. Филиппов, Г. Мюллер проводят первый цикл работ по радиационным методам индукции мутаций.

1926 год

С. С. Четвериков написал статью, заложившую основы популяци-онной генетики и синтеза генетики и теории эволюции.

1927 год

Г. Мюллер доказал мутационный эффект рентгеновских лучей, за что в 1946 г. получил Нобелевскую премию в области физиологии и медицины.

Н. К. Кольцов выдвинул идею матричного синтеза, которая и сегодня отвечает современным представлениям биологов: «В основе каждой хромосомы лежит тончайшая нить, которая представляет собой спиральный ряд огромных органических молекул – генов. Возможно, вся эта спираль является одной гигантской длины молекулой».

1928 год

Открытие явления трансформации у бактерий (Ф. Гриффит).

1929—1930 годы

А. С. Серебровский и Н. П. Дубинин впервые продемонстрировали сложную природу организации гена; первые реальные шаги на пути создания современного представления о тонкой структуре гена.

1931 год

Барбара Мак–Клинтон продемонстрировала наличие кроссинго–вера.

1934 год

Н. П. Дубинин и Б. Н. Сидоров открыли особый тип эффекта положения.

Б. Л. Астауров осуществил успешные опыты по получению у шелкопряда потомства из неоплодотворенных яиц (одно из самых интересных достижений в прикладной генетике того времени).

1935 год

Н. В. Тимофеев–Ресовский, К. Г. Циммер, М. Дельбрюк осуществили экспериментальное определение размеров гена. Ими дана трактовка гена с позиций квантовой механики, тем самым был создан фундамент для открытия структуры ДНК.

1940 год

Дж. Бидл и Э. Татум сформулировали теорию «один ген – один фермент». (Нобелевская премия по физиологии и медицине за 1958 г.).

1943 год

И. А. Рапопорт, Ш. Ауэрбах и Дж. Г. Робсон впервые показали индукцию мутаций химическими веществами.

1944 год

О. Эвери, К. Маклеод и М. Маккарти установили, что «веществом гена» служит ДНК. Начало «эры ДНК».

М. Дельбрюк, С. Лурия, А. Херши произвели пионерские исследования по генетике кишечной палочки и ее фагов, после чего эти объекты стали модельными для генетических исследований на многие десятилетия. (Нобелевская премия по физиологии и медицине за 1969 год за открытие цикла репродукции вирусов и развитие генетики бактерий и вирусов).

Л. А. Зильбер сформулировал вирусно–генетическую теорию рака.

1950 год

Э. Чаргафф сформулировал знаменитое «правило Чаргаффа», которое гласит: в ДНК число нуклеотидов А равно числу Т, а число Г – числу Ц.

Б. Мак–Клинтон показала существование перемещающихся генетических элементов. С большим опозданием (только в 1983 г.) она получила за это Нобелевскую премию в области физиологии и медицины.

1951 год

Р. Франклин и М. Уилкинсон получили первую рентгенограмму молекулы ДНК.

1953 год

Ф. Крик и Дж. Уотсон, опираясь на результаты опытов генетиков и биохимиков и на данные рентгеноструктурного анализа, создали структурную модель ДНК в форме двойной

спирали. В 1962 году им совместно с М. Х. Ф. Уилкинсом присуждена Нобелевская премия по физиологии и медицине.

1956 год

Ю. Тио и А. Леван установили, что диплоидный набор хромосом у человека равен 46.

А. Корнберн обнаружил первый фермент, способный синтезировать ДНК в пробирке – ДНК–полимеразу I. В 1959 году он совместно с С. Очоа получил Нобелевскую премию по физиологии и медицине за исследование механизма биологического синтеза РНК и ДНК.

1958 год

М. Мезельсон и Ф. Сталь доказали полуконсервативный механизм репликации ДНК.

1960 год

Открытие РНК–полимеразы С. Б. Вейсом, Дж. Гурвицем и А. Сти–венсом.

И. А. Рапопорт сообщил об открытии «супермутагенов».

1961 год

В работах М. У. Ниренберга, Р. У. Холли и Х. Г. Кораны начата расшифровка «языка жизни» – кода, которым в ДНК записана информация о структуре белковых молекул. В 1968 году все трое разделили Нобелевскую премию по физиологии и медицине, которая была присуждена им «за расшифровку генетического кода и его функционирования в синтезе белков».

Ф. Жакоб и Ж. Моно пришли к выводу о существовании двух групп генов – структурных, отвечающих за синтез специфических (ферментных) белков, и регуляторных, осуществляющих контроль за активностью структурных генов. В 1965 г. Нобелевская премия по физиологии и медицине присуждена А. М. Львову, Ф. Жакобу и Ж. Моно за открытие генетической регуляции синтеза ферментов и вирусов.

1962 год

Дж. Гёрдон осуществил первое клонирование животного организма (лягушка).

Дж. Кэндрию и М. Перутц были удостоены Нобелевской премии по химии за впервые осуществленную расшифровку трехмерной структуры белков миоглобина и гемоглобина.

1965 год

Р. Б. Хесин показал, что регуляция синтеза белка осуществляется путем включения и выключения транскрипции генов.

1966 год

Б. Вейс и С. Рихардсон открывают фермент ДНК–лигазу.

1969 год

Х. Г. Корана синтезировал химическим путем первый ген.

1970 год

Открытие обратной транскриптазы, фермента, синтезирующего ДНК с использованием комплементарной РНК в качестве матрицы. Это было сделано будущими Нобелевскими лауреатами по физиологии и медицине (1975) Г. Теминым и Д. Балтимором.

Выделена первая рестриктаза – фермент, разрезающий ДНК в строго определенных местах. За это открытие в 1978 году Нобелевская премия по физиологии и медицине была присуждена Д. Натансу, Х. Смиту и В. Арберу.

1972 год

В лаборатории Пола Берга получены первые рекомбинантные ДНК (Нобелевская премия по химии за 1980 г. вручена П. Бергу и Г. Бойеру). Заложены основы генной инженерии.

1973 год

С. Коэн и Г. Бойер разработали стратегию переноса генов в бактериальную клетку.

1974 год

С. Милстайн и Г. Келер создали технологию получения моноклональных антител. Ровно десять лет спустя они (вместе с Н. К. Эрне) получили за это Нобелевскую премию по физиологии и медицине.

Р. Д. Корнберг описывает структуру хроматина (нуклеосомы).

1975 год

С. Тонегава показал различное расположение генов, кодирующих вариабельную и константную часть иммуноглобулинов, в ДНК эмбриональных и миелоидных клеток, что дало основание для вывода о перегруппировках генов иммуноглобулинов при образовании клеток иммунной системы (Нобелевская премия по физиологии и медицине в 1987 г.).

Осуществлено первое клонирование кДНК.

Е. Саузерн описал метод переноса фрагментов ДНК на нитроцеллюлозные фильтры, метод получил название Саузерн-блот гибридизации.

1976 год

Открытие у животных (на примере дрозофилы) «прыгающих генов», сделанное Д. Хогнессом (США) и российскими учеными во главе с Г. П. Георгиевым и В. А. Гвоздевым.

Основана первая генно-инженерная компания (Genentech), использующая технологию рекомбинантных ДНК для производства различных ферментов и лекарственных средств.

Д. М. Бишоп и Г. Э. Вармус сообщили, что онкоген в вирусе представляет собою не истинный вирусный ген, а клеточный ген, который вирус «подхватил» когда-то давно в ходе репликации в клетках и теперь сохраняет в измененном мутациями виде. Было также показано, что его предшественник, клеточный протоонкоген, в здоровой клетке играет важнейшую роль—управляет ее ростом и делением. В 1989 г. оба этих ученых получили Нобелевскую премию по физиологии и медицине за фундаментальные исследования канцерогенных генов опухолей.

1977 год

Опубликованы быстрые методы определения (секвенирования) длинных нуклеотидных последовательностей ДНК (У. Гилберт и А. Максам; Ф. Сенгер с соавт.). Появилось реальное средство анализа структуры генов как основа для понимания их функций. В 1980 году У. Гилберт и Ф. Сенгер совместно с П. Бергом получили Нобелевскую премию по химии «за существенный вклад в установление первичной структуры ДНК; за фундаментальные исследования биохимических свойств нуклеиновых кислот, в том числе рекомбинантных ДНК».

Полностью секвенирован геном бактериофага φX174 (5386 п. н.).

Секвенирован первый ген человека – ген, кодирующий белок хорионный соматотропин.

П. Шарп и Р. Робертс показали, что гены у аденовирусов (позднее выяснилось, что и у эукариотических организмов) имеют мозаичную экзон-интронную структуру, и открыли явление сплайсинга (Нобелевская премия по физиологии и медицине в 1993 г.).

К. Итакура с соавт. синтезируют химически ген соматостатина человека и осуществляет искусственный синтез гормона соматостатина в клетках кишечной палочки *E. coli*.

1978 год

Компания Genentech осуществила перенос эукариотического гена инсулина в бактериальную клетку, где на нем синтезирован белок – проинсулин.

Определена полная последовательность нуклеотидов ДНК вируса SV40 и фага fd.

1979 год

Показано, что химически трансформированные клетки содержат активированный онкоген *BAS*.

1980 год

Дж. Гордоном с соавт. получена первая трансгенная мышь. В про-нуклеус оплодотворенного одноклеточного эмбриона микроинъекцией введен ген тимидин-киназы вируса простого герпеса и показано, что этот ген работает во всех соматических клетках мыши. С тех пор трансгеноз стал основным подходом как для фундаментальных исследований, так и для решения практических задач сельского хозяйства и медицины.

1981 год

Определена полная нуклеотидная последовательность митохондриальной ДНК человека.

Несколько независимых исследовательских групп сообщили об открытии человеческих онкогенов.

1982 год

Определена полная нуклеотидная последовательность бактериофага λ (48502 п. н.).

Показано, что РНК может обладать каталитическими свойствами, как и белок.

1983 год

С помощью биоинформатики найдена гомология фактора роста PDGF с известным онкобелком, кодируемым онкогеном SIS.

Показано, что разные онкогены кооперируют при опухолевой трансформации клеток.

Ген болезни Хантингтона локализован на хромосоме 4 человека.

1984 год

У. Мак-Гиннис открыл гомеотические (Нох) регуляторные гены, ответственные за построение общего плана тела животных.

А. Джеффрис создает метод геномной дактилоскопии, в котором нуклеотидные последовательности ДНК используются для идентификации личности.

1985 год

Создание К. Б. Мюллисом революционизирующей технологии – полимеразной цепной реакции, ПЦР – наиболее чувствительного до сих пор метода детектирования ДНК. Эта технология получила широкое распространение (Нобелевская премия по химии за 1993 г.).

Клонирование и определение нуклеотидной последовательности ДНК, выделенной из древней египетской мумии.

1986 год

Клонирование гена *RB* – первого антионкогена – супрессора опухолей. Начало эпохи массированного клонирования генов опухоле-образования.

1987 год

Созданы первые дрожжевые искусственные хромосомы – YAC (Yeast Artificial Chromosomes). Они сыграют большую роль как векторы для клонирования больших фрагментов геномов.

1988 год

Создание международного проекта «Геном человека», поставившего своей целью полное секвенирование ДНК человека.

Под эгидой Комитета по науке и технике в СССР начала работу программа «Геном человека», которую возглавил Научный совет по геномной программе во главе с академиком А. А. Баевым.

Показана возможность анализа митохондриальной ДНК из очень древних образцов при исследовании мозга человека давностью 7000 лет.

Предложен метод «нокаута» генов.

1989 год

Т. Р. Чех и С. Альтман получили Нобелевскую премию по химии за открытие каталитических свойств некоторых природных РНК (рибозимов).

1990 год

Создана Международная организация по изучению генома человека (HUGO), вице-президентом которой в течение нескольких лет был российский академик А. Д. Мирзабеков.

Ф. Коллинз и Л. – Ч. Тсуи идентифицировали первый ген человека (*CFTR*), ответственный за наследственное заболевание (кистозный фиброз), который расположен на хромосоме 7.

В. Андерсоном осуществлено первое успешное применение генной терапии для лечения больной с наследственным иммунодефицитом.

Определена полная последовательность генома вируса оспо-вакцины (192 т. п. н.).

1992 год

Э. Кребсу и Э. Фишеру присуждена Нобелевская премия по физиологии и медицине за открытие обратимого фосфорилирования белков как важного регулирующего механизма клеточного метаболизма.

1995 год

Определена полная последовательность генома первого самостоятельно существующего организма – бактерии *Haemophilus influenzae* (1 830 137 п. н.). Становление геномики как самостоятельного раздела генетики.

1997 год

Определена полная последовательность нуклеотидов геномов кишечной палочки *E. coli* и дрожжей *Saccharomyces cerevisiae*.

Нобелевская премия по физиологии и медицине присуждена американцу С. Прузинеру за вклад в изучение болезнетворного агента белковой природы, приона, вызывающего губчатую энцефалопатию, или «коровье бешенство» у крупного рогатого скота.

Я. Вильмут с сотрудниками впервые клонировали млекопитающее – овцу Долли.

1998 год

Определена полная нуклеотидная последовательность первого высшего организма – нематоды *Caenorhabditis elegans*.

У нематоды *C. elegans* обнаружен механизм РНК-интерференции.

1999 год

Полностью секвенирована ДНК хромосомы 22 человека.

Роберт Фурчготт, Луис Игнаро и Ферид Мурад получили Нобелевскую премию за открытие роли оксида азота в качестве сигнальной молекулы (то есть, регулятора и переносчика сигналов) сердечно-сосудистой системы.

Клонировали мышь и корову.

2000 год

Вчерне завершено полное секвенирование генома человека и дрозофилы.

Нобелевская премия по физиологии и медицине присуждена А. Карлссону, П. Грингарду и Э. Кенделу за открытие, касающееся «передачи сигналов в нервной системе».

Клонировали свинью.

2001 год

Нобелевская премия по физиологии и медицине присуждена Л. Хартвеллу, Т. Ханту и П. Нерсу за открытие ключевых регуляторов клеточного цикла.

2002 год

Полностью секвенирован геном мыши.

Нобелевская премия по физиологии и медицине присуждена С. Бреннеру, Р. Хорвитцу и Дж. Салстону за их открытия в области генетического регулирования развития органов и запрограммированной клеточной смерти.

ГЕНЫ – ОСНОВНОЙ ТЕКСТ ГЕНОМА

Гены – это атомы наследственности.

С. Бензер

Орлы летают одиноко, бараны пасутся стадами.

Ф. Сидни

Необходимо отметить, что термин «ген» многие, даже неспециалисты, воспринимают легко без всякого специального пояснения. Слишком много и давно об этом говорят и довольно часто используют этот термин в разных ситуациях. И всем вроде бы все ясно: ген – это участок ДНК, который кодирует определенный белок. Согласно центральной догме молекулярной биологии, информация, записанная в ДНК с помощью четырехбуквенного нуклеотидного алфавита, переводится простым механическим способом на другой, аминокислотный, алфавит из 20 букв, которым записывается строение белковых молекул. «Один ген – один белок» – вот каким был многие годы основополагающий тезис молекулярных генетиков. Для того, чтобы ген правильно работал, необходимы специальные знаки, обозначающие начало и конец инструкции для синтеза сначала мРНК, а потом белка. Для того, чтобы началась считывание (транскрипция) гена, в первую очередь необходим структурный элемент, получивший название промотора (старт–сигнал транскрипции). Для окончания транскрипции существует другой специальный сигнал. Обязательными элементами гена являются также сигналы, определяющие начало и конец трансляции мРНК на рибосомах. Все эти старт–и стоп–сигналы записаны в ДНК с помощью того же самого четырехбуквенного алфавита, что и информация для построения молекулы белка. Между двумя такими знаками препинания записаны инструкции для синтеза белка. В таком упрощенном виде ген определяют как последовательность нуклеотидов, расположенных между старт–сигналом и стоп–сигналом и кодирующих одну белковую цепь (полипептид). Однако на самом деле все обстоит намного сложнее. Кроме генов, кодирующих белки, существуют гены, кодирующие РНК, которые выполняют в клетке только структурную роль и не транслируются. Более того, в последние годы было обнаружено большое число ранее неизвестных РНК, которые выполняют в клетке специальные регуляторные функции. Обо всем этом мы поговорим далее. Современное понимание того, что же из себя представляет ген, складывалось на протяжении всех этих долгих лет, наполняясь постепенно конкретным и обобщенным содержанием, преодолевая сомнения, заблуждения и неожиданные повороты. В результате на сегодняшний день трудно дать однозначное определение термину «ген», хотя вроде бы этот вопрос в общем виде был решен Менделем почти 140 лет назад.

Конечно же, в секвенированном ДНКовом тексте генома человека в первую очередь внимание ученых было обращено на участки, кодирующие белки, то есть на белок–кодирующие гены. Отметим, что первый ген человека – ген, кодирующий белок со сложным названием хорионный соматомаммотропин, – был секвенирован еще в 1977 году. А что же выяснилось в результате секвенирования всего генома человека? Оказалось, что в геноме человека нуклеотидные последовательности, кодирующие белки (**экзоны**), занимают лишь 1,1–1,4% от длины всех молекул ДНК. Но именно эти мизерные проценты ДНКового текста человека и стали объектом пристального внимания. Для сравнения отметим, что у червя *C. elegans* доля таких последовательностей в геноме составляет 27%, у мухи дрозофилы – 20%, у дрожжей – 70%, а у бактерий – 86%. То есть, просматривается закономерность: по мере усложнения организмов доля кодирующих белки участков ДНК в их геномах резко падает.

Генам при их открытии, как людям при рождении, сразу же присваивают различные имена. Без этого невозможно было бы разобраться, о чем идет речь и что изучают разные ученые в каждом конкретном случае. Правда, путаница в этом вопросе пока еще существует, особенно если ген одновременно открывают в разных лабораториях и при этом называют по-разному. Известно, что в лингвистике существует целая наука об именах людей – ономастика. Эта наука давно уже отметила влияние разных факторов и причин на формирование имен, связь с историей, географией и многими другими науками, а также процессами, протекающими в обществе. Так, в постреволюционные годы появились имена Вилен (В. И. Ленин), Даздраперма (Да здравствует первое мая), при Б. Ельцине появились Борели, а при В. Путине уже официально зарегистрированы такие имена, как Путилла, Владпут и даже Затерсор (замочим террористов в сортире).

Конечно, и на имена, присваиваемые генам, также влияют разные обстоятельства. Но вообще-то этот процесс довольно стихийный. Нередко наименование гена связывают с выполняемой им функцией (например, дистрофин – ген, мутации в котором ответственны за дистрофию мышц). В других случаях используют особенности структуры гена (например, присутствие специфической последовательности в гене, называемой *homeo box*, автоматически приводит к имени *Нох*). Иногда авторы напрягают свою фантазию, и тогда появляются такие экзотические имена, как *харакири* (ген-самоубийца), *икарус*, *vang* (по имени художника ван Гога) или *indy* (первые буквы от слов *I'm not dead yet* – я еще не умер – название для гена, который влияет на продолжительность жизни). Ген *SHH* назван в честь популярной видеоигры Соника Хеджхога, (англ. hedgehog — ежик). Некоторые гены называются именами богов (например, *гефестин* – от греческого бога огня и кузнечного дела Гефеста). На практике генетики используют сокращенные обозначения генов, состоящие обычно из 3—5 латинских букв. Поскольку функции многих генов еще не известны, а удастся только определить, что какая-то конкретная область генома человека способна теоретически кодировать белок, то часто пока еще перед названиями генов появляется слово «гипотетический».

Анализ генома человека позволил на конец 2002 года обнаружить порядка 30 000—40 000 генов, кодирующих белки. Эти гены сильно отличаются друг от друга по размерам. Подсчитано, что средняя длина гена у человека составляет около 27 тыс. п. н. Такой усредненный ген содержит 9 экзонов (средний размер каждого около 150 п. н.) и 8 интронов (средний размер каждого чуть меньше 3400 п.н.). Но это лишь средние значения. Самые короткие гены содержат всего два десятка букв-нуклеотидов, например, гены эндорфинов – полипептидов, вызывающих ощущение удовольствия. В то же время самый длинный ген, кодирующий один из белков мышц, – миодистрофин содержит 2,4 млн. п. н.

Кроме того, гены не равномерно распределены между разными хромосомами. При средней плотности один ген на 100 т. п. н. генома их содержание в хромосоме 19 составляет около 2 на 100 т. п.н., а в хромосомах 2, 13 и Y – хромосоме – от 0,15 до 0,7. Если сравнить плотность генов с плотностью расселения людей, то Y – хромосома напоминает нашу Сибирь, а хромосома 19 – Европейскую часть России. Для сравнения: в геноме бактерий содержится свыше 1000 генов на 1,0 млн. п. н., у дрожжей около 450 генов на 1,0 млн. п. н., а у червя *C. elegans* – около 200. Следовательно, плотность расположения генов на единицу длины генома заметно падает по мере эволюционного усложнения организмов.

«Жи, ши – пиши через и»

После секвенирования генома человека и ряда других организмов был проведен детальный компьютерный анализ нуклеотидных последовательностей. Подсчитано, например, сколько в ДНК может идти подряд букв А, или как часто Г встречается после Ц. И тут выяснилось, что в генах эти сочетания подчиняются определенным правилам, тогда как

в промежутках между генами, там, где ничего существенного в ДНК не записано, частота сочетаний разных нуклеотидов близка к случайной. Где много генов, там много букв Г и Ц, а где генов нет, там много букв А и Т. Здесь опять можно провести параллель с обычным текстом. По правилам грамматики русского языка, которые мы учили в школе, «жи, ши – пиши через и», в словах после букв *ж* и *ш* буква *ы* не встречается. Нечто подобное характерно и для генетических текстов – в кодирующих участках некоторые сочетания нуклеотидов практически не встречаются, а распределение других сочетаний сильно отличается от случайного. Это был один из первых, но далеко не единственный результат анализа ДНК.

Знаки препинания (Генетическая пунктуация)

Как в письменном языке пунктуация (точки, запятые, тире и т.д.) используется для членения и графической организации текста, так и в ДНКовом тексте существуют всевозможные знаки–сигналы, обозначающие границы отдельных функциональных элементов. Без таких сигналов генетическая информация не может быть правильно считана с носителя. Классический пример из лингвистики – предложение «Убить нельзя помиловать». Если не поставить запятую, то остается непонятным, то ли убить, то ли помиловать. Так и в гене. Знаки препинания, в отличие от генетического кода, не так универсальны в живом мире. Часть из них у низших организмов (прокариот) существенно отличаются по сравнению с высшими (эукариота–ми), включая человека.

Вот как выглядит, например, одна из страниц ДНКового текста, содержащаяся в томе (хромосоме) 11 человека (рис. 17).

Глядя на этот рисунок, даже специалисту сразу трудно что–то понять. Однако простая компьютерная программа быстро позволяет найти здесь участок, кодирующий белок интерферон. Это осуществляется в результате того, что в генах имеются специальные сигналы–ярлычки, которые можно рассматривать как знаки препинания в ДНКовом тексте. И компьютер способен их различить. Рассмотрим главные из них.

Так, в генах человека, как и у других высших организмов, имеются специальные последовательности–сигналы, которые определяют начало и конец процесса транскрипции, то есть синтеза мРНК.

```

1  ГГЦЦЦЦЦЦТА ГГТАЦЦЦЦАА ААААТЦТАТТ ТЦААААААГ ТГЦЦЦЦТААГ ААТАТАГТТА
61  ТЦААГТТААГ ТААААТГТЦА АТАГЦЦТТТТ ААТТТААТТТ ТТААТГТТТТ ТАТЦАТЦЦТТ
121  ТГЦААТААТА АААЦАТТААЦ ТТТАТАЦТТТ ТТААТТТААТ ГТАТАГААТА ГАГАТАТАЦА
181  ТАГГАТАТГТ АААТАГАТАЦ АЦАГТТТТАТА ТТГГАТТААА АТАТААТГГГ АГАТЦЦААТЦ
241  АГААААААГТ ТЦТАААААГ ГЦЦЦТГГГТ ААААГАГГАА ГГАААЦААТА АТГАААААА
301  ТГТГТГГАГА ААААЦАГЦТГ ААААЦЦААТТ ТАААГАТГТТ АТАААГАААГ ЦААААГАГА
361  АГТАГАААГТ ААЦАЦАГГГТ ЦАТТТГГААА АТТТАААЦГА ГТАТГТЦЦЦ ТАТТТААГГЦ
421  ТАГГЦАЦААА ГЦААГТТЦТТ ЦАГАГААЦЦТ ГГАГЦЦТААГ ГТТТАГТЦЦЦ АЦЦАТТТЦА
481  АЦЦАГТЦТАГ ЦАГЦАТЦТГЦ ААЦАТЦТАЦА АТГГЦЦТТГА ЦЦТТТГЦТТТ АЦТГТТГЦЦ
541  ЦТЦЦТГТГЦ ТЦАГЦГЦАА ГТЦААГЦТГЦ ТЦТГТГТГЦ ГТГАТЦГЦЦ ТЦАААЦЦАЦ
601  АГЦЦТГТГТА ГЦАГГАГТАЦ ЦТТГАТГЦЦ ЦТГЦАЦАГА ГГАГГАГААТ ЦТЦТЦТТТЦ
661  ТЦЦТГЦТТГА АГТАЦАГАЦА ТГАЦТТТГГА ТТТЦЦЦААГ АГТАГТТТГ ЦААЦАГТТЦ
721  ЦААААГЦТТ АААЦАТЦЦЦ ТГТЦТЦЦАТ ГАГАТГАТЦЦ АГЦАГАТЦТТ ЦААТЦТЦТЦ
781  АГЦАААААГ АЦЦАТЦТГЦ ТГЦТГГГГАТ ГАГАЦЦЦТЦЦ ТАГАЦАААТ ЦТААЦТГАА
841  ЦТЦТАЦЦАЦ АГЦТГААТГА ЦЦТГГААГЦЦ ТГТГГАТАЦ АГТГТТТТГ ГТГАЦАГАГ
901  АЦТЦЦЦЦТГА ТГААГТАГГА ЦТЦАТТЦТГ ГЦТГТГАГГА ААТАЦТЦЦА ААГААТЦАЦ
961  ЦТЦТАТЦТГА ААГАГААГАА АТАААГЦЦЦТ ТГТЦЦТТТГ АГТТТГТЦАГ АГЦАГАААТЦ
1021 АГТАГАТЦТТ ТТЦТТТГТЦ ААЦАААЦТТ ЦААГАААГТТ ТААГААГТАА ГГААТГАААА
1081 ЦТГТТЦААЦ АТГГАААТГА ТТТЦАТТГА ТЦЦТАТГЦЦ АГЦТЦАЦЦТ ТТТАГТАТЦ
1141 ГЦЦАТТТЦАА АГАЦЦАТГТ ТЦТГЦТАТГ АЦЦАТГАЦАЦ ГАТТТАААТЦ ТТТТЦАААТГ
1201 ТТТТААГТАГ ТАГТААТЦАА ЦАТТГТАТЦ АГЦТЦТТААГ ГЦАЦТАГТЦЦ ЦТТАЦАГАГТ
1261 АЦЦАТГЦТГА ЦТГАТЦЦАТТ АТЦАТТТТАА АТАТТТТТАА ААТАТТАТТТ АТТТААЦТАТ
1321 ТТААААААЦА АЦТТАТТТТТ ГТЦАТАТТА ТГЦАТГТГЦ АЦЦТТТГЦАЦ АГТГТТТААГ
1381 ГТААТААААТ ГТГТЦТТТГ ТАТТГГТАА АТТТАТТТГ ГТТГТТЦАТ ГГААЦТТТТГ
1441 ЦТАГГГААЦТ ТТГТАЦТТГ ТТТАТЦТТТ ААААТГАААТ ТЦЦААГЦЦТА АТТГТГЦААЦ
1501 ЦТГАТТАЦАГ ААТААЦГГТТ АЦАТТЦАТТ ТГТЦАТЦАА ТАТТАТАТЦ ААГАТАААГ
1561 ТАААААТААА ЦТТТЦГТАА АЦЦААГТТТ АТГТТГТАЦ ЦААГАТААЦА ГГГТГААЦЦТ
1621 ААЦАААТАЦА АТТЦТГЦЦЦ ЦТГТГТААТ ГТАТТТТТТ АТГААААААА ЦТААААААТГ

```

Рис. 17. Страница Энциклопедии человека, на которой записана последовательность нуклеотидов ДНК человека, кодирующая белок альфа–интерферон. Ген интерферона выделен курсивом. Подчеркнут участок, кодирующий белок, состоящий из 188 аминокислот. Жирным шрифтом выделены знаки пунктуации в этом гене, обеспечивающие его правильную транскрипцию с образованием мРНК и последующую правильную трансляцию этой мРНК

Фермент, называемый РНК–полимеразой, узнает эти сигналы и осуществляет синтез «правильных» молекул мРНК (со строго запрограммированными в геноме концами). Одним из таких знаков препинания в ДНКовом тексте, отмечающим начало процесса транскрипции, служит короткая последовательность нуклеотидов: ТАТА (как пример, см. рис. 17).

Для процесса трансляции, то есть синтеза белка на мРНК, также имеются свои специальные знаки препинания. Таковыми являются отдельные тройки нуклеотидов (триплеты). На рис. 17 они выделены жирным шрифтом. Место, с которого начинается кодирование белка, – старт–сигнал – триплет АУТ. Далее следуют триплеты, которые кодируют аминокислоты в белке. А в конце текста имеется специальный стоп–сигнал (триплет УГА).

Но это еще не все. ДНК иногда претерпевает некоторые химические изменения. Известна такая химическая модификация ДНК, как метилирование, которая происходит строго по динуклеотидам–сигналам, представляющим собой пару нуклеотидов, в которой первым идет Г, а за ним Ц. Наконец, важными сигналами в генах являются участки, определяющие такой важный процесс, как **сплайсинг**, который происходит уже после образования РНК на гене. Об этом и пойдет речь дальше.

Гены построены из кусков

Успехи в изучении генома человека, а еще ранее – проведенный анализ отдельных генов других высших организмов, привели к ряду принципиально новых и совершенно неожиданных для генетиков результатов в отношении устройства генов, особенно тех, которые кодируют белки. Со времени появления центральной догмы молекулярной биологии думали, что непрерывная линейная последовательность нуклеотидов в ДНК с помощью триплетов нуклеотидов кодирует непрерывную последовательность нуклеотидов в мРНК, а последняя как посредник передает эту информацию на рибосомы, которые, считывая ее, осуществляют синтез линейных молекул белков. Это соответствовало ситуации, которую первоначально наблюдали при изучении генов у бактерий. Однако у высших организмов, включая человека, (и даже у некоторых низших) картина оказалась более сложной.

Когда сравнивали мРНК и соответствующий белок, то действительно обнаруживали непрерывность их нуклеотидных и аминокислотных последовательностей. То же наблюдали и при сопоставлении последовательностей аминокислот в белках и нуклеотидов в геномах вирусов. Однако с появлением возможности секвенировать протяженные участки ДНК (т. е. непосредственно генов) как у некоторых вирусов, так и у высших организмов получили совершенно неожиданные результаты. Внутри нуклеотидных последовательностей ДНК, кодирующих белок, были обнаружены какие-то иные, ничего не кодирующие участки. Чудеса, да и только! В чем же дело? Оказалось, что многие гены «расчленены» на отдельные куски. Одни из этих кусков, как и положено, кодируют белок (их называли **экзонами**), а другие никакие белки кодировать не способны (их называли **интронами**) и расположены между экзонами. Таким образом, в гене кодирующие и не кодирующие участки перемежаются друг с другом (рис. 18). Так устроено большинство белок-кодирующих генов. Хотя имеются и исключения. В этих случаях говорят, что ген устроен из одного экзона и интронов не содержит.

Такое разорванное (экзон-интронное) строение оказалось характерным для подавляющего большинства (хотя и не для всех) генов человека. Вместе с тем у большинства простейших организмов инт



Рис. 18. Схематическое изображение экзон–интронного строения некоторых генов человека. Экзоны – прямоугольники, интроны – линии между прямоугольниками. Темные прямоугольники – участки генов, кодирующие белок, белые прямоугольники в первом и последнем экзонах – специальные участки гена, кодирующие мРНК, но не кодирующие белок

роны не обнаружены. Размеры интронов часто значительно превышают размеры экзонов, что существенно увеличивает общую длину нуклеотидных последовательностей ДНК, образующих ген. При этом в геноме человека интроны в целом существенно длиннее, чем у других организмов.

Каким же образом разорванный ДНКовый текст реализуется в клетках в виде неразорванного РНКового, а затем белкового? Выяснилось, что первоначально при транскрипции гена синтезируется большой РНК–предшественник, копия ДНКового текста (с экзонами и интронами вместе). Далее в результате работы специфических ферментов происходит нарезание РНК–предшественника на куски. Те из них, которые ничего не кодируют (интроны), отбрасываются, а кодирующие куски (экзоны) соединяются между собой, обычно в том же порядке, как они располагались исходно в ДНК. В результате формируется зрелая молекула мРНК. Этот сложный процесс и получил название **сплайсинга** (рис. 19).

В интронах содержатся особые сигналы, которые узнаются специальными ферментами, осуществляющими сплайсинг. Так, в 98% случаев интроны начинаются с динуклеотида ГТ, а заканчиваются динуклеотидом АГ. Они–то и служат главными сигналами для правильного осуществления сплайсинга.

Во всем этом и состоит основное ноу–хау в организации генов человека и других эукариотических генов. «Лоскутное» устройство большинства генов – чрезвычайно важное эволюционное приобретение высших организмов. За счет сплайсинга в РНК может происходить соединение не только между соседними экзонными нуклеотидными последовательностями, но и между другими, отстоящими порой в гене на значительном расстоянии. Это называют **альтернативным сплайсингом** (рис. 20). В геноме человека альтернативный сплайсинг характерен для более трети генов. Этому можно привести следующую аналогию: из слова «администрация» за счет удаления отдельных букв и слогов можно образовать множество совершенно разных по смыслу слов: ад, министр, амнистия, астра, нация. Существование множества альтернативных вариантов сплайсинга в конечном итоге обеспечивает появление более чем одного белка при экспрессии одного единственного гена. В частности, на гене белка тро–понины человека, содержащем 18 экзонов, за счет альтернативного сплайсинга может образовываться 64 различных продукта! Сейчас подсчитано, что в среднем один ген у человека способен кодировать около 3–х разных белков. Альтернативный сплайсинг зависит от множества внутриклеточных и внешних факторов, в том числе, как показано недавно, даже от вирусного воздействия на клетку.

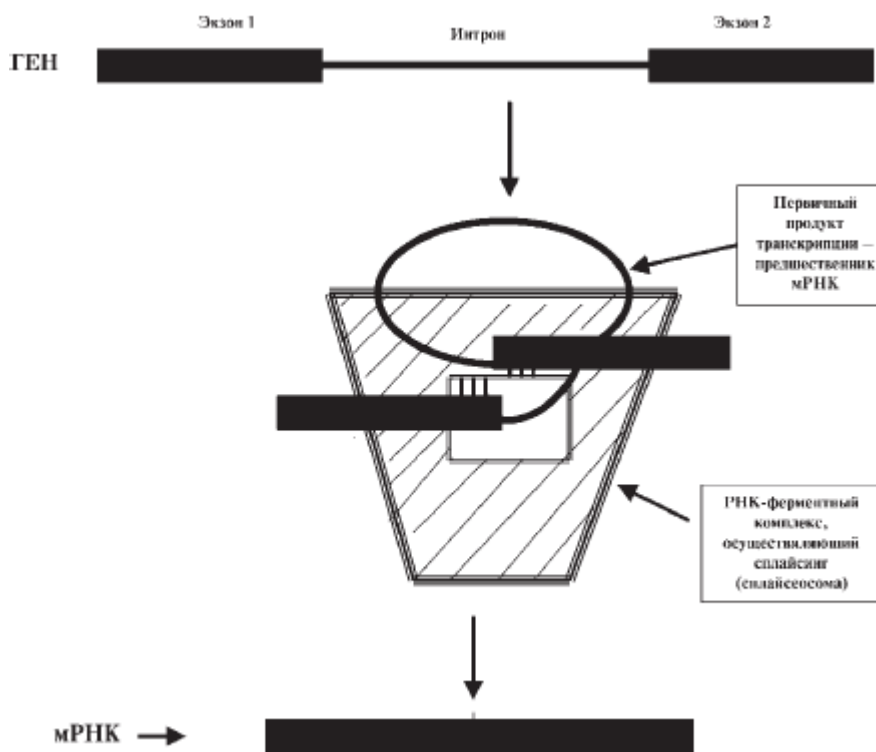


Рис. 19. Образование молекул мРНК на РНК–предшественнике в результате сплайсинга. Из РНК–предшественника вырезаются фрагменты, синтезированные с интронных последовательностей гена, а участки, синтезированные на экзонах, соединяются друг с другом, что в конечном итоге приводит к формированию зрелой функционально активной мРНК

Экзон–интронное строение генов дало принципиально новую возможность для эволюции генов – комбинировать отдельные элементы при формировании генов. В результате этого в эволюции возникали новые гены, кодирующие разные белки за счет разного сочетания экзонов одного и того же гена. То есть из уже готовых «слов» составляются совершенно разные «предложения». И такой подход, используемый природой, оказался весьма продуктивным.

Следует отметить, что и в зрелой мРНК, которая образуется уже после удаления из первичного транскрипта интронных нуклеотидных последовательностей, также не все участки кодируют белок. Однако, в отличие от интронов, такие последовательности нуклеотидов расположены в начале и в конце молекулы (см. рис. 18). Их размеры также существенно различаются. Так, в разных мРНК чело

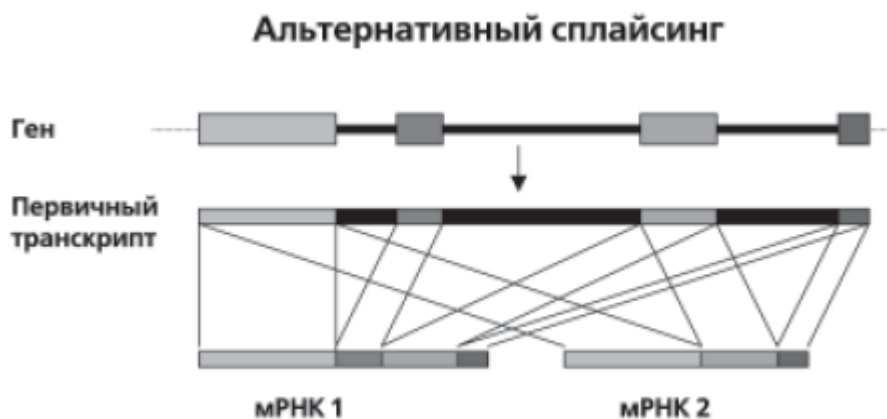


Рис. 20. Схематическое изображение альтернативного сплайсинга. Прямоугольниками разного оттенка обозначены экзоны, черными линиями – интроны. Показаны лишь две из множества возможных мРНК, образующихся на одном гене

века не кодирующая область, расположенная в начале молекулы, имеет длину от 2800 до 18 нуклеотидов (в среднем 200 нуклеотидов), а та, которая расположена в конце, составляет от 8500 до 20 нуклеотидов (в среднем 1000 нуклеотидов). Эти не кодирующие белок области гена очень важны для нормального функционирования мРНК, ее трансляции, стабильности и локализации в клетке. Как уже говорилось, на одном гене за счет альтернативного сплайсинга может образовываться несколько мРНК. Следует подчеркнуть, что и на одной уже сформировавшейся мРНК могут образовываться разные белки. Это происходит за счет наличия в 50% мРНК человека не одного, а двух АУГ-кодонов, являющихся старт-сигналами для синтеза белка на мРНК. За счет этого, в разных ситуациях клетка может начать синтез белка с разных мест мРНК, в результате чего образуются разные по длине, а порой и по своим свойствам, белки. Опять же пример из лингвистики. Если читать текст не с первой, а со второй буквы, смысл зачастую совершенно меняется (сравните: удочка и дочка).

В связи со всем этим понятно, что классическая формула генетиков «один ген – один белок», предложенная лауреатами Нобелевской премии Джоржем Бидлом и Эдуардом Татумом, оказывается в реальности справедливой лишь для небольшой группы генов, а многие гены (возможно, большинство) кодируют семейства родственных, но существенно отличающихся белков, то есть нередко действует принцип «один ген – много белков». За счет этого в геноме осуществляется принцип экономии генетического материала (компактная запись) и достигается большой набор генных продуктов – белков – на относительно небольшом числе генов.

В таблице 3 приведена общая характеристика белок-кодирующих генов, выявленных в результате секвенирования генома человека.

Таблица 3. Усредненные данные о структуре белок-кодирующих генов генома человека

Характеристика генов	Средние значения
Число экзонов	8,8
Размер интронов	3365 п. н.
Длина кодирующей последовательности	1340 п. н.
Размер	27000 п. н.

В реальности размеры разных генов, а также их экзонов и интронов сильно отличаются. И основной вклад в эту вариабельность вносят интроны. Уже говорилось, что максимальный по размерам ген дистрофина занимает в геноме человека 2,4 млн. п. н. Подавляющая часть занята именно интронами (их средний размер составляет несколько десятков тысяч п. н.). Наибольшую кодирующую последовательность в геноме человека имеет ген по имени *титин*. Его размер равен около 81000 п. н. Этот ген – чемпион и по числу интронов (178 штук!), и по длине единичного экзона (17106 п. н.).

Важно отметить, что разбивка генов на отдельные экзоны неслучайна. Индивидуальный экзон часто кодирует не просто какую-то часть белковой молекулы, а определенный фрагмент белка, выполняющий в целом белке отдельную функцию. Такой отдельный структурно-функциональный фрагмент, входящий в состав целого белка, называют **доменом**.

Перекрытие генных текстов

Принцип компактности записи информации в ДНК проявляется еще в одном. Важный факт, обнаруженный первоначально при анализе последовательностей генов у вирусов, заключается в том, что в одном и том же участке ДНК иногда может быть записана информация о двух совершенно разных РНК (и, соответственно, о совершенно разных белках). Такое случается и в геноме высших организмов, включая человека. Как же это осуществляется? Вспомним, что записанный в ДНК текст явно не разбит на слова (буквы – нуклеотиды – следуют в молекуле ДНК одна за другой). Но слова в этом тексте есть, просто знаки пунктуации и разрывы между словами скрыты от глаза и узнаются клеткой после перекодировки информации из ДНК в РНК. Для большей наглядности приведем широко известный пример такого текста:

НАПОЛЕОНКОСИЛТРАВУПОЛЯКИПЕЛИСОЛОВЬЯМИ.

В зависимости от разбивки этого текста на слова получается два предложения с совершенно разным значением:

НА ПОЛЕ ОН КОСИЛ ТРАВУ ПОЛЯ КИПЕЛИ
СОЛОВЬЯМИ
НАПОЛЕОН КОСИЛ ТРАВУ ПОЛЯКИ ПЕЛИ
СОЛОВЬЯМИ.

Теоретически на двух цепях ДНК с учетом того, что код триплетный, можно записать 6 разных текстов: три на одной цепи и три на другой. Это может быть достигнуто за счет сдвига точки начала считывания информации, что называют сдвигом рамки считывания (начало считывания возможно с любого из 3-х разных рядом лежащих нуклеотидов) (см. рис. 21). Такое явление получило название перекрытия генов. Впервые оно было обнаружено у вирусов. И здесь было понятно, зачем это потребовалось. Вирусные геномы обычно очень маленькие. Перекрытие генов диктовалось необходимостью как можно компактнее разместить информацию на отно

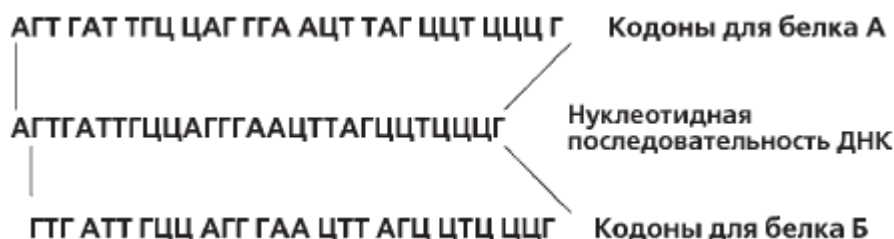


Рис. 21. Схематическое изображение перекрывания двух ДНК-текстов, записанных на одном участке ДНК. Белок А образуется на мРНК, которая транслируется с одних кодонов, а белок Б образуется на мРНК, читаемой (транслируемой) со сдвигом в один нуклеотид. В результате кодоны (они разделены пробелами) совершенно разные, и, соответственно, при трансляции мРНК образуются совершенно разные белки сительно коротких молекулах ДНК. Большой геном человека теоретически позволяет обойтись без такого перекрывания. Однако и в геноме человека, хотя и относительно редко, перекрывание генов также имеется.

Каков может быть смысл в сохранении этого феномена в геноме человека? Пока еще не все ясно, но кое-какие моменты уже проясняются. Например, в ряде случаев обнаружили считывание РНК с разных цепей одного участка ДНК. Это приводит к образованию таких РНК, которые в силу комплементарности могут взаимодействовать друг с другом в клетке, образуя в результате двунитевые РНК. А в дву-нитевом виде мРНК не способна к такому важному процессу, как трансляция. Таков реальный механизм специфической регуляции (инактивации) экспрессии генов человека, число которых, по последним оценкам, составляет ~ 1600.

Не все гены кодируют белки

Прежде всего, следует отметить, что кроме генов, кодирующих белки, в геноме имеются еще гены, на которых синтезируется РНК, которые не являются мРНК (то есть, не кодируют белок), но выполняют ряд самостоятельных важных функций в клетках. В результате получается ситуация, что традиционное привычное определение гена надо расширять, включив в него гены, кодирующие белки, и гены, не кодирующие белки, но кодирующие функционально значимые РНК.

Давно уже известно, что кроме мРНК на ДНК синтезируются разнообразные вспомогательные РНК, которые сами не транслируются (т. е. они не кодируют белки), но участвуют в разных клеточных процессах. В первую очередь это РНК «домашнего хозяйства» — ри-босомные РНК (рРНК), транспортные РНК (тРНК) и др., которые участвуют непосредственно в синтезе белка на мРНК. Эти РНК кодируются участками, которые также называются генами, и составляют основную массу РНК в клетках. рРНК входят в состав рибосом, являясь важным компонентом их структурной организации. У человека присутствует два основных типа рРНК размером около 1900 нуклеотидов и более 5000 нуклеотидов в малой и большой субчастицах соответственно. Еще одни не кодирующие белок РНК — тРНК, которые обеспечивают аппарат трансляции, подтаскивая к рибосомам различные аминокислоты, вступающие в реакцию друг с другом. В геноме человека гены, кодирующие рРНК и тРНК, представлены многочисленными копиями. Так, в секвенированной ДНК человека обнаружено около 500 генов, кодирующих тРНК. Генов для рРНК в геноме человека выявлено около 200 копий, которые расположены на 5-ти разных хромосомах (13, 14, 15, 21 и 22).

Изучение свойств РНК привело к тому, что представление об исключительности белков в катализе биохимических реакций пришел конец. Выяснилось, что в природе име-

ются виды РНК, которые, подобно белкам, обладают высокоспецифической каталитической активностью.

Очень важным стало обнаружение в геноме человека множества других генов, также производящих РНК, но не способных кодировать белок. Постепенно выяснилось, что некоторые из таких не кодирующих белки РНК принимают участие в важнейших процессах, происходящих в клетке: регуляции транскрипции ДНК, сплайсинга и трансляции мРНК, модулировании функций белков и их пространственного распределения в клетке. По этой причине их назвали **риборегуляторами**. И примеров таких риборегуляторов уже сейчас можно привести немало. Так, установлено, что не кодирующий белок участок гена *H19* имеет отношение к ряду процессов, протекающих в клетках, и, в частности, к их злокачественному перерождению. Другой РНК-кодирующий ген контролирует работу белок-кодирующего гена *HFE*, вовлеченного в метаболизм железа и связанного с наследственным заболеванием хемохроматозом. В последнем случае РНК-продукт кодируется тем же самым геном, который он регулирует, но его образование осуществляется на другой нити ДНК. В результате так называемый антисмысловой (комплементарный) РНК-продукт способен взаимодействовать с мРНК, образуя гибриды, неспособные транслироваться в рибосомах с образованием белка. Еще один интересный ген – ген РНК-активатора стероидного рецептора. Он обеспечивает активность стероидных рецепторов за счет образования комплекса с этим белком.

В клетках человека, как и у других организмов, выявлены короткие двунитевые РНК (микроРНК), отдельные из которых, по-видимому, могут участвовать в процессе регуляции экспрессии генов через механизм, названный **РНК-интерференцией**. Этот механизм впервые был обнаружен в 1998 году у низших организмов. Различные микроРНК в клетках червя *C. elegans* оказались в состоянии «приглушать» работу строго определенных генов путем воздействия на процесс синтеза кодируемых ими белков. Вполне вероятно, и у человека микроРНК играют подобную роль.

Таким образом, мы видим, что продукты генов, кодирующих только РНК, вмешиваются в различные клеточные процессы, используя при этом совершенно разные механизмы. Гены риборегуляторов составляют, по-видимому, заметную часть генома человека. Примерные оценки говорят о величине на порядок больше, чем доля белок-кодирующих нуклеотидных последовательностей. Сам факт существования таких генов, неспособных кодировать белок, но реально проявляющих себя в производстве функционирующих в клетках РНК, ставит большой вопрос перед исследователями генома. И, в первую очередь, что следует после этого считать собственно геном?

Ген в гене (генная матрешка)

Изредка обнаруживаются варианты, когда внутри одного гена целиком содержится другой, меньший по размерам ген. Этакая своеобразная «матрешка», построенная из генов. Такая организация генов весьма редка. Так, в хромосоме 22 имеется лишь 2 таких случая. Чаще всего белок-кодирующий ген располагается в интроне другого белок-кодирующего гена. Но встречаются и другие варианты. В качестве примера можно привести ситуацию, имеющую место для ми-тохондриального гена одной из рибосомных РНК. Ген, кодирующий эту рРНК, обеспечивает ею рибосомы митохондрий в качестве структурного компонента (т. е. не кодирует белок). Однако вместе с тем небольшой участок, расположенный внутри этого гена, кодирует короткий белок (полипептид), получивший название гуманин (от англ. human – человек), который принимает участие в процессе программированной клеточной гибели. То есть в РНК-кодирующем гене может содержаться белок-кодирующий ген. Другой вариант—уже упоминавшийся выше ген *H19*. Здесь, наоборот, ген, кодирующий

белок, содержит внутри своей кодирующей части другой более короткий ген, кодирующий только РНК, которая принимает участие в регуляции работы этого гена.

Генные семейства

Как у людей имеются семьи, так и у генов нередко существуют семейства. Семейством генов называют набор из двух или более генов, чьи экзоны родственны между собой, т. е. похожи (хотя и не идентичны) по нуклеотидной последовательности. В геноме человека присутствует около полутора тысяч таких семейств генов. Причем только около сотни из них специфичны для человека и других позвоночных животных, тогда как основная масса генных семейств имеется и у человека, и у червя.

Разные члены одного семейства генов возникали в эволюции из одного гена–предшественника (условно – от одного прапрапрадеда, как и в человеческой популяции). Процесс, в результате которого появлялись «копии», получил название **дупликации**, т. е. удвоения. Такие дупликации в некоторых случаях происходят не одноразово, а многократно. Сейчас стало ясно, что дупликации играют очень важную роль в эволюции. Подсчитано, что в геноме человека в сумме дуплицировано около 3,6% нуклеотидных последовательностей размером в 1000 и более п. н.

Разные копии одного семейства генов могут располагаться в геноме рядом и следовать друг за другом (это называют тандемной дупликацией). Рассмотрим несколько примеров. Так, гены, кодирующие рибосомную РНК (рРНК) и белки гистоны, расположены в геноме человека в виде кластеров, построенных из одних и тех же последовательностей, следующих друг за другом (рис. 22).

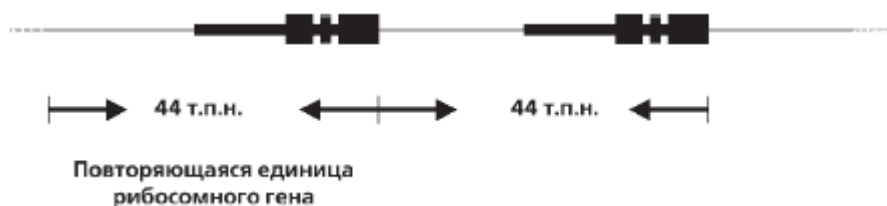


Рис. 22. Строение участка генома человека, состоящего из множества генов, кодирующих рибосомные РНК. Утолщенные линии – транскрибируемые участки

Такие «гроздьи» одинаковых генов обычно требуются организму для того, чтобы набирать большие количества определенного продукта. Действительно, рРНК составляет по массе основную часть всей клеточной РНК, входя в состав таких важных «машин» клетки, как рибосомы. Огромная масса гистонов нужна клетке для упаковки гигантских молекул ДНК и плотной укладки их в хромосомы. Во всех таких случаях говорят о большой «дозе» гена в геноме, которая в конечном итоге и обеспечивает клетку большим количеством продукта.

Однако в большинстве других случаев возникающие в результате дупликации «копии» генов постепенно претерпевают изменения (мутации), что приводит к возникновению определенных различий между ними, хотя формальное родство при этом сохраняется. Такой процесс был назван **дивергенцией**, т. е. расхождением. Дивергенцию обычно выражает в процентах различий между двумя родственными последовательностями нуклеотидов в ДНК или последовательностями аминокислот в белках.

Однако даже после того, как разные «копии» генов одного семейства становятся не совсем похожими друг на друга, они чаще всего выполняют сходную функцию. Они оста-

ются очень дружной семейкой! Но экспрессируются разные члены одного такого семейства обычно в разное время или в разных типах клеток. Так, в геноме человека обнаружено тридцать генов, кодирующих родственные факторы роста клеток кожи (фибробластов). А вот у низших организмов число таких генов существенно меньше (у дрозофилы и червя их всего 2). Другой пример – гены, кодирующие кератины – белки наружного слоя кожи и ее производных (волосы, ногти). Их у человека 111 штук! Но и это не предел. Геном человека, как ни удивительно, имеет около 1 000 копий генов–рецепторов обоняния.

Дивергировавшие в ходе эволюции человека гены семейства иногда оказываются разбросанными в разных местах на одной или даже разных хромосомах. При этом ген может быть удвоен целиком, а может происходить удвоение небольшого набора его экзонов или даже всего лишь одного из его экзонов.

Как и в семье не без урода, так и в семействах генов встречаются иногда такие члены, которые не утруждают себя никакой работой. Нарушение их способности функционировать в клетке обычно связано с накоплением опечаток в их ДНК-тексте. Такие гены–тунеядцы получили название «псевдогенов». Перед их названием ставят греческую букву ψ . Не совсем ясно, зачем геному нужны такие гены, почему он сохранил их в эволюции, не избавился от них. Но факт остается фактом. По имеющимся оценкам, в геноме человека имеется около 20000 таких, подобных вымершим реликтам, псевдогенов. В частности, в огромной семье генов–рецепторов обоняния, состоящем из примерно 1000 генов, около 60% являются псевдогенами. Считается, что массивная потеря функциональных генов произошла за последние 10 млрд. лет, и связано это со снижением роли обоняния у человека по сравнению с другими позвоночными организмами. Здесь, по-видимому, действует принцип отмирания из-за неупотребления. Установлено, что существует, по крайней мере, два механизма, приводящих к образованию псевдогенов. Один из них заключается в удвоении уже существующего гена (дупликация), а потом дополнительная копия накапливает мутации, которые препятствуют его работе. Другой механизм связан с процессом сплайсинга и обратной транскрипции. На сформировавшейся мРНК синтезируется ее ДНК–копия, которая в дальнейшем встраивается в геном. Такие псевдогены не содержат ни интроны, ни промоторные участки. Кроме того, в них со временем накапливаются многочисленные мутации, не позволяющие этим генам обеспечивать синтез нормального белка.

Все вышесказанное можно проиллюстрировать на примере семейства генов, кодирующих белки крови человека (рис. 23). У человека имеется два семейства глобиновых генов (альфа и бета). Эти семейства расположены на разных хромосомах. Каждое семейство состоит из нескольких членов, имеются здесь и псевдогены. Нормальные гены глобиновых семейств экспрессируются в клетках крови на разных стадиях развития человека. Одни у эмбрионов (ξ , ϵ), другие у плода (G_γ , A_γ , $a1$), а третьи – у взрослого организма ($\alpha2$, δ , β). Функция у всех глобиновых генов одинакова – синтез глобинов, участвующих в переносе кислорода клетками крови. И место их работы одно – клетки крови эритроциты. Отличаются они лишь тем, что включаются в разное время, в результате чего выполняют одну и ту же роль, но на разных стадиях развития организма.



Рис. 23. Строение двух локусов генома, состоящих из кластеров родственных глобиновых генов человека. Экспрессирующиеся гены указаны стрелками (направление транскрипции), молчащие гены (псевдогены) – квадратиками

Организация кластеров глобиновых генов у человека совершенно одинакова с организацией этих генов у гориллы и павиана. Это указывает на то, что такая организация полностью сформировалась еще от 20 до 40 млн. лет назад и с тех пор не изменялась. Анализ мутаций в межгенных областях позволил сделать вывод об эволюции не отдельных генов, собранных в такой кластер, а кластера как единого целого **супергена**.

По-иному ведут себя многочисленные гены семейства актинов, которые кодируют сократительные белки клетки. Эти гены экспрессируются почти все время, и их экспрессия происходит как в мышечных клетках, так и во многих других немышечных. При этом функция актинов везде одинакова.

Новые предложения из фрагментов старых текстов

Экзон–интронная организация генов способствует еще одному механизму возникновения новых генов – созданию новых ДНКовых текстов из фрагментов старых. Это осуществляется в результате «тасования» уже предсуществующих экзонов, что может приводить к появлению нового их сочетания и новых белков с новыми функциями. Чаще всего это происходит за счет процесса присоединения новых экзонов к старым. Примером этому может служить случай, представленный на рис. 24.

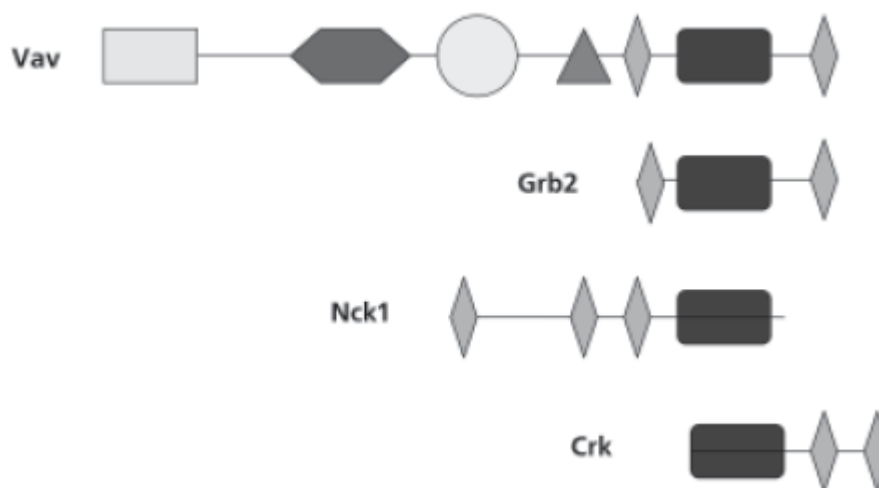


Рис. 24. Пример формирования новых генов за счет различного сочетания уже имеющихся экзонов. Изображены четыре разных белка, состоящих из сходных блоков–доменов (обозначены разными символами), которые кодируются разными экзонами

В процессе эволюции из общего предшественника образовались гены дрозофилы *trx* и ген человека *alr*. Ген человека *hrx* возник в результате присоединения к гену *trx* четырех новых экзонов. Это приводит к появлению в белке, кодируемом этим геном, четырех новых функциональных доменов, что существенно меняет его свойства по сравнению с предшественником. Такая ситуация вообще–то характерна для всех высших организмов, но у человека она встречается в 2—5 раз чаще, чем, например, у дрозофилы или червя.

После получения всех этих новых многочисленных сведений об особенностях организации и эволюции генов у человека (сходным образом обстоит дело и у других высших организмов) стало ясно, что дать строгое определение того, что же такое собой представляет ген, довольно сложно. Но даже если бы такое определение сегодня имелось, его, как и всякое другое, не следовало бы считать незыблемым. Не зря говорил Спиноза: «Всякое определение есть ограничение». Однако в целом базовое представление о гене осталось в силе. Ген – это физическая (определенный участок ДНК) и функциональная (кодирует белок или РНК) единица наследственности.

Сколько же всего генов у человека?

Это наиболее интересный вопрос, ради которого собственно и затевалось полное секвенирование генома человека. После получения основной информации о структуре генома человека в первую очередь были произведены различные анализы по поиску генов и определению их числа. Однако задача оказалась не простой. Это может показаться странным для читателя, но однозначного ответа на поставленный вопрос до сих пор нет.

Сколько же всего генов в ДНК человека? Еще несколько лет назад полагали, что их около 100 тыс., затем решили, что не более 80 тыс. В конце 1998 г. пришли к выводу, что в геноме человека не более 50—60 тыс. генов и на их долю приходится около 3% общей длины ДНК.

Последние подсчеты общего числа генов в геноме человека проводили несколько международных команд ученых. Уже упоминавшаяся компания «Celera» провела собственные исследования, результаты которого изложены в журнале «Science» в 2001 году. По ее оценкам общее число генов в геноме человека составляет от 26383 до 39114. Средний размер гена оценивается равным примерно 3000 п. н. Если принять, что число генов у человека порядка

30 тысяч генов и на каждый ген приходится примерно 3 тыс. п. н., то нетрудно подсчитать, что в кодировании белков принимает участие менее 1,5% хромосомной ДНК. Таким образом, генетические инструкции по формированию человеческой личности занимают меньше 3 сантиметров на двухметровой молекуле ДНК. Удивляет и малое количество генов, несущих эти инструкции, – их всего в пять раз больше, чем, например, у такого на наш взгляд совершенно примитивного организма, как муха дрозофила.

Вторая команда исследователей из Национального института геномных исследований США во главе с Френсисом Коллинзом, подсчитав число генов у человека независимым способом и на основе своих данных, получила сходный результат – около 32000 генов содержится в геноме каждой клетки человека.

Разнобой в окончательные оценки пока вносят два других коллектива ученых. Доктор Вильям Хезелтайн (руководитель фирмы «Хьюмэн Геном Сайенс») продолжает настаивать, что в их банке содержится приватизированная информация на 120 тыс. генов. Этой информацией он не собирается пока делиться с мировой общественностью. Фирма вложила деньги в патенты и собирается заработать на полученной информации, поскольку она относится к генам широко распространенных болезней человека. Фирма «Инсайт» сообщила о том, что имеет в настоящее время каталог, состоящий из 140 тысяч идентифицированных ей генов человека, и также настаивает на этом количестве общего числа генов у человека.

Очевидно, что наспех приватизированная генетическая информация будет еще тщательно анализироваться и проверяться в ближайшие годы, пока точное число генов станет окончательно «канонизировано». Дело в том, что устройство генов весьма многообразно и до конца еще не поняты все возможные варианты. Вот мы прочитали последовательность нуклеотидов ДНК. Определено, что она способна кодировать белок. Но один ли? Выше уже говорилось о том, как транскрипция и последующие модификации РНК, а затем трансляция и модификации полипептидов, способны обеспечить огромное многообразие белков, кодируемых одним участком ДНК. И понять это исходя только из нуклеотидной последовательности ДНК очень часто просто невозможно. Тем не менее структура генома представляет собой единственную базу для осмысления данных, получаемых такими новыми направлениями, рожденными геномикой, как транскриптомика (исследует совокупность РНК–транскриптов организма), протеомика (исследует совокупность белков организма), метаболомика (исследует обмен веществ – метаболизм – в организме). Эти направления призваны дополнить лежащий в основе структурной геномики метод геномного секвенирования, дать возможность выйти за пределы его разрешающей способности.

Выше уже также говорилось об альтернативном сплайсинге. Сейчас хорошо известно, что за счет этого процесса с одних и тех же генов могут считываться разные белки, которые затем взаимодействуют друг с другом, образуя неповторимую смесь, как из основных цветов в живописи – желтого, красного и голубого можно получить мириады оттенков. Такой сплайсинг характерен не менее чем для половины генов человека. Считается, что в среднем с одного гена человека за счет альтернативного сплайсинга может образовываться три разных пептида. Но некоторые гены имеют до 10 альтернативно сплайсируемых экзонов, что позволяет теоретически получать более 1000 различных вариантов белков всего лишь на одном гене. В реальности число разных белков, кодируемых одним геном, достигает 10. Кроме того, существуют еще и альтернативные промоторы, альтернативные кодоны инициации трансляции, редактирование РНК (превращение Ц в У или А в аналог Г – инозин). Все вышесказанное пока еще невозможно учесть при оценке общего числа генов у человека.

Но и это не все. Кроме генов, кодирующих белки, имеются еще гены, конечным продуктом которых являются РНК. Вспомним об упоминавшихся выше генах–рибосегуляторах – они не кодируют белки, но производят функционирующую в клетках РНК. Так что скорее всего окончательная оценка числа генов у человека будет сделана еще нескоро.

На сегодняшний день ученым известны функции всего лишь около восьми–десяти тысяч из них. А детальные сведения о механизмах их регуляции еще более скудны. Тем не менее, приведенные выше данные о строении и функционировании генов человека свидетельствуют о том, что у человека, царствующего в природе, в отличие от других существующих на нашей планете организмов, очень высока сложность **протеома** – полного набора функциональных белков в клетке, которая обеспечивается не просто за счет крупного размера генома или большого числа генов, а благодаря всевозможным инновациям, связанным с функционированием генов и формированием белков: большее число доменов–модулей, более высокая комбинаторика (перемешивание) этих модулей в белках, активное использование альтернативного сплайсинга и многое другое, о чем мы поговорим дальше.

ОСНОВНАЯ ЧАСТЬ ГЕНОМА – ТЕРРА ИНКОГНИТА

*Прогресс в биологии – это переход
от ложного знания к истинному незнанию.*
В. Я. Александров

В природе нет ничего бесполезного.
М. Монтень

Теперь стало возможным оценить, что РНК синтезируется лишь максимум на 25—28% нуклеотидных последовательностей генома человека. А на кодирование белков используется вообще всего лишь чуть более 1% генома. Остальные транскрибируемые участки ДНК – это интроны белок-кодирующих генов и гены для РНК, не кодирующих белки. Таким образом, большая часть генома (свыше 70%) не имеет, по-видимому, прямого отношения ни к каким генам и инертна в плане транскрипции. Как это ни странно на первый взгляд, но факт остается фактом. Большая часть генома не кодирует ни белки, ни какую-нибудь РНК вообще, т. е. не относится к генам или каким-то другим, функционирующим на уровне транскрипции (РНК), нуклеотидным последовательностям.

В ДНК, на которой не образуется никакая РНК и которая, таким образом, для кодирования информации совершенно нейтральна, записаны длинные тексты, которые сегодня представляются для ученых полной абракадаброй, смысл и происхождение которой совершенно не ясен. Зачем нужны клетке эти «бессмысленные» и никак, на первый взгляд, не работающие участки ДНК, пока никто не знает, хотя априори ясно, что ничего лишнего в природе не должно быть. При полном секвенировании генома человека, естественно, была определена последовательность не только генов, но и этих пока ничего не обозначающих для ученых участков. Но из чего-то они все-таки состоят. Когда провели их детальный анализ, то выяснилось, что здесь присутствует скопище всех мыслимых и немыслимых вариантов расположения букв, бессмысленных слов и предложений, различных повторов одного и того же «текста», палиндромов и так далее и так далее. Рассмотрим подробнее хотя бы основные из них.

ДНКовый текст человека перенасыщен повторами

Как уже было показано в ранних экспериментах, описанных выше, в ДНК человека содержатся многочисленные элементы, которые не кодируют никакие белки, но при этом многократно встречаются в ДНКовом тексте. В «догеномную» эру считалось, что такие повторы составляют около 25% генома человека. Секвенирование ДНК показало, что их существенно больше: повторяющиеся последовательности в сумме занимают около 50% ДНКового текста. А общее число таких повторов в геноме человека равно приблизительно 5 миллионам. Для сравнения: у дрозофилы доля повторов в геноме составляет менее 5%, у червя – 6—7%, а у других млекопитающих – чуть меньше, чем у человека. Повторы напоминают бессмысленные фразы, которые с упорным постоянством в разных вариациях повторяются в тексте, но от этого смысл в них все равно не возникает.

На рис. 25 приведены некоторые основные варианты расположения повторов в геноме человека, а далее мы подробнее остановимся на описании отдельных из них.

Перевертыши (обращенные повторы)

В ДНКовых текстах человека, как, впрочем, и в ДНК других организмов, присутствуют уже упоминавшиеся обращенные повторы, **перевертыши** или, как их еще часто называют, **палиндромы**. Уже говорилось о том, что перевертыши–палиндромы хорошо известны из литературы – это предложения, которые читаются одинаково слева направо и справа налево без учета знаков препинания и интервалов между словами. Считается, что палиндромы известны еще со времен Древней Греции, т. е. свыше двух тысяч лет. На Руси палиндромы когда-то назывались «рачьими стихами». В качестве примера приведем две державинские строки:

Я разуму уму заря,
Я иду с мечем судия.



Рис. 25. Основные варианты расположения повторяющихся последовательностей в геноме человека. Повторы обозначены стрелками, неповторяющиеся участки – серыми линиями

Особое свойство таких участков ДНК заключается в том, что при определенных состояниях ДНК они способны образовывать структуры, называемые крестами. На рис. 26 дан пример одного из таких возможных крестов.

В реальной ситуации в клетке такие структуры действительно могут формироваться. Но зачем это нужно? Окончательного ответа на этот вопрос нет. Одно из предположений сводится к тому, что кресты могут быть местами связывания с какими-то специфическими белками, обеспечивающими регуляцию или процессов транскрипции, или репликации ДНК. С другой стороны, при транскрипции таких структур в РНК должны формироваться «шпильки», а такие структуры в РНК часто являются значимыми (то есть выполняют определенные функции – или способствуют транскрипции, или, наоборот, препятствуют ей).

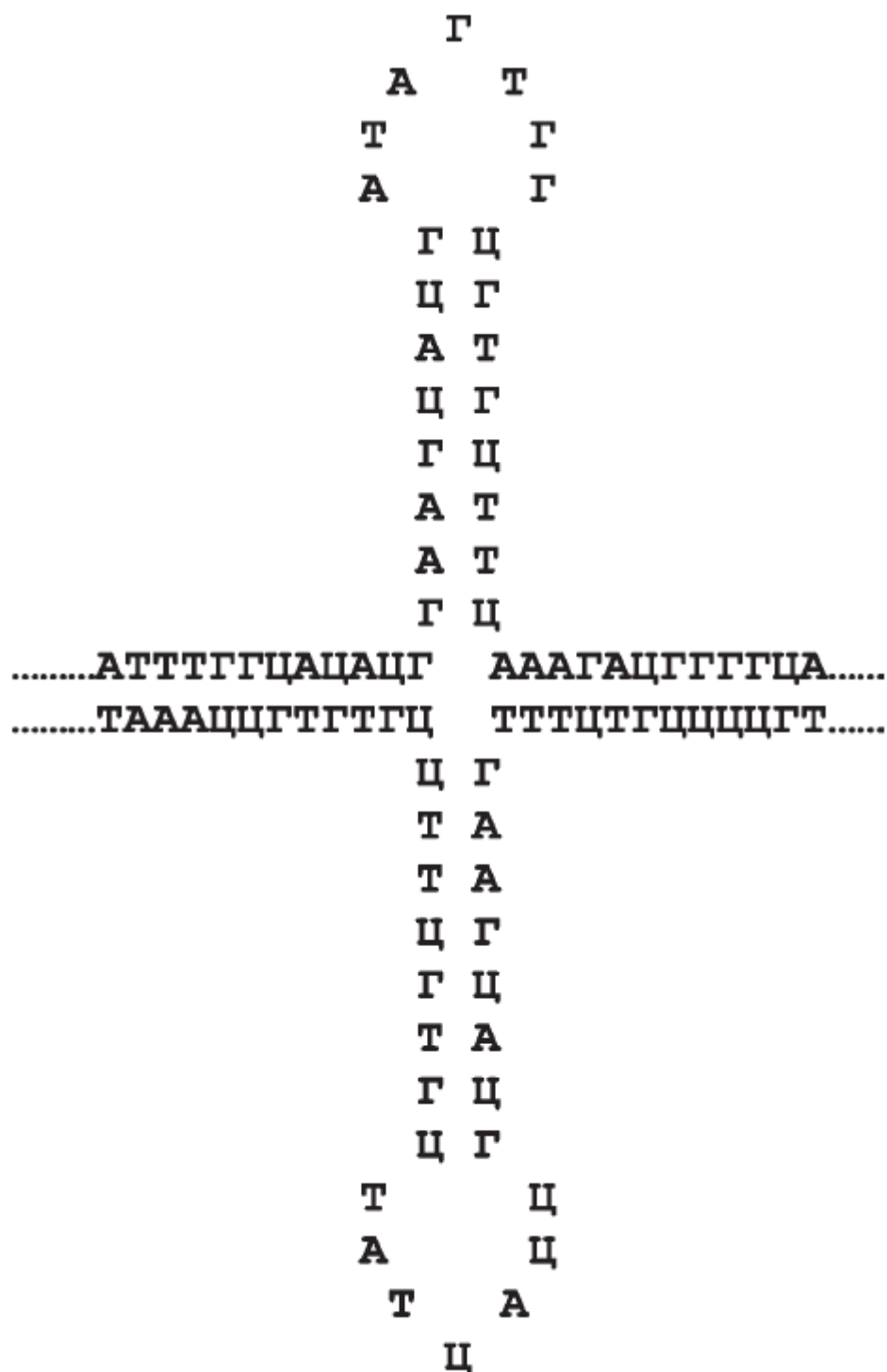


Рис. 26. Крест, способный образовываться последовательностью ДНК, состоящей из обращенных повторов

Простые тандемные повторы (сателлиты)

Как и у других организмов, в геноме человека присутствуют короткие повторяющиеся нуклеотидные последовательности, которые расположены вплотную друг за другом (тандемно). Их часто называют **сателлитными ДНК**. Мы уже говорили о сателлитных ДНК ранее. Это те участки ДНК, которые при центрифугировании в градиенте солей цезия осаждаются отдельно вне области основной массы ДНК. В разных типах сателлитной ДНК раз-

личные копии представлены неодинаково – их численность варьируется от сотен до миллионов. Если записать их структуру текстом, то это выглядит, как, например, предложение из одного и того же слова: столстолстолстол–столстолстолстолстол. Ясно, что такие предложения не могут кодировать никакую серьезную информацию. Тем не менее выяснилось, что с сателлитными нуклеотидными последовательностями в некоторых случаях могут специфически взаимодействовать отдельные белки. А это уже некоторое довольно серьезное указание на возможное участие таких повторов в каких-то клеточных процессах.

К настоящему времени в геноме человека обнаружено по крайней мере шесть видов сателлитных повторов. Это в первую очередь три типа так называемых классических сателлитов. Сателлит 1 состоит из элементарной повторяющейся последовательности нуклеотидов длиной 42 п. н. (располагается на семи разных хромосомах); сателлиты 2 и 3 состоят из элементарных повторяющихся единиц длиной 5 п. н. (первый из них обнаружен на четырех, а второй – на двенадцати разных хромосомах).

Кроме того, во всех хромосомах человека присутствуют альфа–сателлитные повторы. Они имеют элементарную повторяющуюся нуклеотидную последовательность длиной 171 п. н. Существуют еще бета–сателлитные повторы с элементарной повторяющейся единицей длиной 68 п. н., которые встречаются лишь на девяти хромосомах. Еще реже представлены в геноме человека гамма–сателлитные повторы, имеющие элементарную повторяющуюся единицу длиной 220 п. н. Такие повторы присутствуют только на двух хромосомах человека (8 и X).

Распределение всех сателлитных ДНК вдоль хромосом не выглядит случайным. Так, некоторые из них «гнездятся» в районах, прилежащих к концам хромосом, и в области **центромеры** (перетяжка на хромосомах, которая не всегда расположена в самом центре, см. рис. 7). В частности, центромеры всех хромосом человека, размер которых варьирует от 0,1 до 4 млн. п.н., состоят преимущественно из альфа–сателлитных ДНК.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.