

Содержание

Предисловие	5
От издательства	6
Глава 1. Естественная история генома человека	7
Общая характеристика генома	7
Что такое ген?	15
Повторяющаяся ДНК	21
Тандемно повторяющиеся последовательности	22
Диспергированные повторяющиеся последовательности	24
Хроматин	27
Экспрессия генов	29
Транскрипция	29
Созревание (процессинг) РНК	34
Трансляция	39
Краткое содержание главы 1	48
Глава 2. Механизмы мутаций: как изменяется геном человека	50
Спонтанные мутации	50
Мутации, возникающие при репликации ДНК	51
Влияние изменений в последовательности оснований на экспрессию генов	58
Мутации, вызываемые мобильными генетическими элементами	65
Мутации вследствие ошибок рекомбинации	66
Хромосомные мутации	69
Индуцированные мутации	77
Скорость (темпы) мутирования у человека	82
Метод 1. Полимеразная цепная реакция	84
Краткое содержание главы 2	87
Глава 3. Генетические болезни: последствия мутаций	89
Менделевское наследование аутосомных генов	89
Менделевское наследование генов в X- и Y-хромосомах	103
От Менделя к генам	109
Анализ сцепления	115
Позиционный анализ	120
За пределами менделизма: сложное наследование	122
Геномный импринтинг	131
Краткое содержание главы 3	137
Глава 4. Геном человека и практическая медицина	140
Генетический скрининг	141
Генетический скрининг новорожденных	141
Генетический скрининг взрослых	146
Этические и юридические аспекты генетического скрининга	155

Лечение генетических нарушений	157
Лечебное питание (диетотерапия)	157
Белковая заместительная терапия	159
Клеточная и тканевая заместительная терапия	165
Генная терапия	170
Фармакогеномика	178
Метод 2. Анализ генов и их экспрессии с помощью биочипов	181
Краткое содержание главы 4	183
Глава 5. Геном митохондрий и митохондриевые патологии	185
Митохондриевые генетические болезни	194
Болезни, приписываемые мутациям в мтДНК	194
Митохондриевые болезни и мутации в ядерной ДНК	201
Митохондрии и старение	205
Лечение митохондриевых болезней	206
Краткое содержание главы 5	207
Глава 6. Генетические основы рака	209
Онкогены	213
Гены–супрессоры опухолей	224
Нарушения регуляции роста	231
Клеточный цикл	231
Программируемая гибель клетки	233
Источники нестабильности генома	235
Колоректальный рак	236
Рак молочной железы	240
Рак с дефектами сохранности ДНК	242
Эпигенетические изменения при раке	244
Краткое содержание главы 6	247
Глава 7. Геном человека и биология развития	249
Генетика внутриутробного развития	254
Факторы транскрипции и раннее развитие	255
Органогенез	260
Сложное наследование пороков развития	265
Геномные нарушения	269
Генетика старения	271
Взгляд в будущее	274
Метод 3. Конструирование трансгенных мышей	276
Краткое содержание главы 7	278
Литература	280
Ресурсы Интернета	281
Приложение	283

Предисловие

«Геном человека» является краткой сводкой основных фактов о генах человека, о том, как они экспрессируются, как мутации приводят к простым или сложным нарушениям, и о том, как быстрые достижения в нашем понимании генома человека влияют на практическую медицину. Эта книга будет полезна людям, занимающимся науками о здоровье на всех уровнях, от студентов до состоявшихся профессионалов, которые хотели бы расширить свои знания о генетике человека без особых затрат времени. «Геном человека» окажется также хорошим дополнительным учебным пособием для тех разнообразных университетских курсов, где основной текст с недостаточной глубиной освещает генетику человека.

Эта книга предполагает, что вы уже прошли вводные университетские курсы по биологии или, по крайней мере, вы знаете основные факты о генах, ДНК, белках и т.п. Если вы не уверены, что эта книга именно то, что вам нужно, то попытайтесь проверить себя с помощью следующего короткого теста:

1. Законы Менделя были написаны 2500 лет тому назад мистиками Среднего Востока, которые жили в пещерах.
2. Уотсон и Крик — это пара известных комиков, которые снимались во многих кинокомедиях в 1940–1950-е годы.
3. АТФ — это ведущая телекоммуникационная компания.
4. Хромосомы различаются цветом, например, X-хромосома — розовая, а Y-хромосома — голубая.
5. Мейоз — это болезнь, передающаяся половым путем.

Если вы задумались на несколько миллисекунд, засомневались в ложности этих утверждений, то вы можете не читать эту книгу. Если же вы на все 100% справились с этим тестом, то вы можете заглянуть в конец каждой главы и прочитать их резюме. Если вы решите, что вам хорошо известны основные сведения о генах и их экспрессии, изложенные в главе 1, — не беда: просто пропустите то, с чем хорошо знакомы, и переходите к тем разделам, которые хотелось бы познать или освежить в памяти. Вполне возможно, что вы найдете здесь много полезного для себя.

От издательства

Автор, редакторы и издательство предприняли максимум усилий, чтобы представить по возможности корректную информацию. Однако они не несут ответственности за ошибки, упущения или любые другие недочеты, относящиеся к использованию читателями содержания данной книги и к употреблению описанных в книге продуктов. В тексте упоминаются и обсуждаются лекарственные препараты и другие медицинские средства, которые могут иметь регулируемое FDA (Food and Drug Administration)¹ ограниченное применение исключительно для лабораторных исследований и клинических испытаний. Информация о лекарствах, представленная в тексте, извлечена из литературных источников, недавно опубликованных данных и отчетов о фармакологических испытаниях. Исследователи, практики-клиницисты и органы государственного контроля и регулирования часто меняют принятые в данной области стандарты. Когда принимается решение использовать какое-либо лекарство в клинике, то за определение того, каков в настоящее время статус данного лекарства в FDA, ответственность несут врач или читатель-пользователь. Для этого они обязаны проверить, соответствует ли содержимое упаковки ее описанию, прочитать инструкцию по применению данного лекарства, рекомендации по дозировкам, предостережения и противопоказания и оценить приемлемость использования данного продукта. Это особенно важно в случаях с новыми или редко применяемыми лекарствами.

¹ В России контроль за лекарственными средствами осуществляют органы Министерства здравоохранения и социального развития Российской Федерации. См.: <http://www.mzsrif.ru/>. Здесь и далее прим. пер.

ГЛАВА 1

ЕСТЕСТВЕННАЯ ИСТОРИЯ ГЕНОМА ЧЕЛОВЕКА

Мы живем в эру захватывающего дух прогресса в области генетики человека. Эта эра — эра геномики — время, когда последовательность ДНК в геноме человека определена почти полностью, время, когда анализируется роль тысяч генов человека в норме и при болезнях. Наступает время, когда изучение небольших вариаций во многих генах приведет нас к индивидуализированной медицине, время, когда будут выявлены генетические основы большинства врожденных аномалий, и время, когда сравнение генов человека и приматов выявит генетические основы человеческой уникальности. Это время постоянного эмоционального возбуждения и для тех, кто участвует в открытиях непосредственно, и для тех, кто участвует опосредованно, применяя бурный поток новых знаний к врачебным сферам деятельности.

Общая характеристика генома

Слово «геном» относится к общему содержанию ДНК у данного вида, включая гены и всю остальную ДНК. Однако *геном* есть нечто большее, чем выражение для обозначения количества ДНК; когда мы говорим о геноме, в действительности мы мыслим в терминах информационного содержания. Содержание ДНК в клетках человека является достаточно стандартным для млекопитающих и составляет примерно 3 миллиарда пар нуклеотидов в зародышевых (половых) клетках (*гаплоидный геном*) или примерно 6 миллиардов пар нуклеотидов в большинстве соматических клеток (*диплоидный геном*)¹. Последний эквивалентен примерно шести пикограммам ($6 \cdot 10^{-12}$ г) ДНК, что составляет примерно долю в 10^{-12} от массы соли в чайной ложке. Размер генома у слона и мыши примерно такой же, поэтому тот факт, что мы способны общаться и писать на тыся-

¹ За рубежом (а зачастую и у нас) длину последовательностей ДНК измеряют числом пар оснований (сокращенно: п.о. или bp).

чах языков, создавать компьютеры, посылать ракеты в глубины космоса и заниматься всякого рода познавательной деятельностью, чего не могут достичь другие млекопитающие, должен быть следствием различий в относительно небольшой части нашей ДНК. Идентификация генетических основ уникальности нашей человеческой анатомии, физиологии, поведения и познавательных способностей будет одним из величайших научных приключений XXI века. Но прежде всего мы должны завершить каталогизацию генов человека и понять, как они участвуют в нормальном развитии и обмене веществ (метаболизме). Эта информация, дополненная исчерпывающим каталогом нормальных и аномальных генетических вариантов, основательно повлияет на клиническую медицину.

Как генетический материал ДНК была идентифицирована в 1944 г., когда Эвери¹, МакЛеод² и МакКарти³ показали, что фенотип бактерии можно изменить, если клетки одного штамма обработать ДНК другого штамма, но не его белком или РНК. В 1953 г. Уотсон⁴ и Крик⁵ вывели двухспиральную структуру ДНК, в которой пары оснований А=Т (аденин-тимин) и Г≡С (гуанин-цитозин) расположены между нитями, и такая молекулярная модель сразу объяснила как кодирование генетической информации, так и репликацию генетического материала⁶.

Задолго до того, как ДНК была признана генетическим материалом, с помощью светового микроскопа были изучены хромосомы благодаря тому преимуществу, что в метафазе — центральной стадии митотического деления клетки — хромосомы сильно конденсированы и физически отделены друг от друга. На заре цитогенетики было множество недоразумений касательно общего

¹ Освальд Теодор Э[й]вери (Oswald Theodor Avery [21.10.1877 – 20.02.1955] — уроженец Канады, американский врач, бактериолог, один из основателей иммунохимии и один из первых молекулярных биологов.

² Колин Мунро МакЛеод (Colin Munro MacLeod [28.01.1909 – 11.02.1972] — канадско-американский генетик.

³ Маклин МакКарти (Maclyn McCarty [09.06.1911 – 03.01.2005]) — американский генетик.

⁴ Джеймс Дьюи Уотсон (James Dewey Watson [р. 06.04.1928] — один из наиболее влиятельных американских молекулярных биологов.

⁵ Френсис Харри Комптон Крик (Francis Harry Compton Crick [08.06.1916 – 28.07.2004]) — английский физик и молекулярный биолог.

⁶ Копии классических статей Эвери — МакЛеода — МакКарти и Уотсона — Крика свободно доступны на сайтах: <http://www.nature.com/nature/dna50/archive.html> и <http://www.exploratorium.edu/origins/coldspring/printit.html>.

числа хромосом в клетках человека. Не так-то легко было получить *метафазную пластинку* хорошего качества (т.е. такой разброс хромосом на предметном стекле, чтобы хромосомы из одной клетки были видны отдельно друг от друга)¹. В те времена хромосомы на препаратах выглядели как довольно бесформенные темные сгустки и часто перекрывали друг друга. Известный цитолог Теофилус Пейнтер², изучая сперматоциты, пришел в 1923 г. к выводу, что у человека 48 хромосом³. Эта цифра не подвергалась сомнению почти 30 лет. Много позже был разработан метод полу-

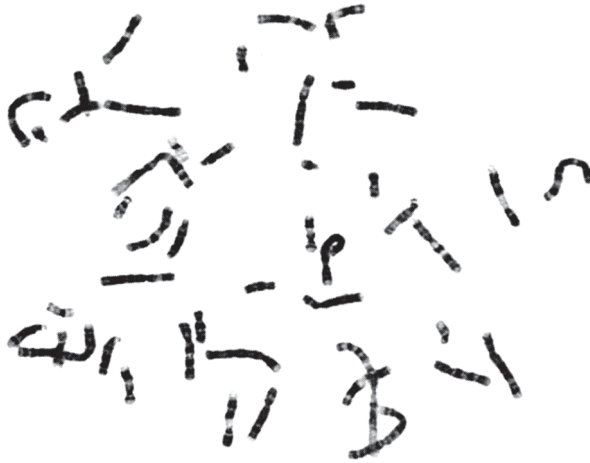


Рис. 1.1. Метафазная пластинка кариотипа нормального мужчины с окраской по Гимзе. (С любезного разрешения David Peakman, Reproductive Genetics Inc., Denver)

¹ В отечественной литературе термином *метафазная пластинка* (metaphase plate) обозначаются как скопление хромосом в плоскости, перпендикулярной оси деления (экваториальная плоскость) на стадии метафазы перед началом анафазного расхождения, так и скопление хромосом на цитогенетических препаратах, наблюдаемое под микроскопом после разрыва клетки (metaphase spread).

² Theophilus Shickel Painter [22.08.1889 – 05.10.1969] – американский зоолог и цитолог. Описал (1933 г.) «гигантские» (политенные) хромосомы в слюнных железах личинок дрозофилы и провел на них изучение хромосомных перестроек и конъюгации хромосом. Одним из первых провел анализ хромосом человека.

³ Удивительно не то, что Пейнтер ошибся, но что используя несовершенную технику, он оказался столь близок к правильному ответу.

чения разбухших клеток в гипотонических растворах (растворах с низкой ионной силой), а другие исследователи открыли вещества, способные блокировать деление клеток (при выращивании их в культуре) на стадии метафазы, и таким образом сильно упростилось получение большого количества необходимого экспериментального материала. Эти два технических новшества резко повысили качество хромосомных препаратов, и с их использованием в 1956 г. ученые получили несомненные доказательства того, что истинное число хромосом у человека 46¹. Наш диплоидный геном оформлен в виде **23 пар хромосом**, размер которых варьирует от 45 миллионов пар нуклеотидов до примерно 280 миллионов пар нуклеотидов. Их них 22 пары — **аутосомы** и одна пара — **половые хромосомы**².

Более подробная классификация, которую разработали в 1940–1950-х гг., была основана на расположении **центромеры** — хромосомной структуры, к которой присоединяются нити митотического веретена. Центромеры могут располагаться либо примерно посередине хромосомы (**метацентрические**), либо заметно ближе к одному из концов хромосомы, чем к другому (**субметацентрические**), или очень близко к одному из концов (**аacroцентрические**). Хромосомы человека группировали на семь классов, согласно их размеру и расположению центромеры.

Значительный прогресс был достигнут в 1960 г., когда были разработаны методы окрашивания, которые позволяют выявлять чередующиеся светлые и окрашенные поперечные полосы (диски, «бэнды») в метафазных хромосомах. Наиболее часто используется система окрашивания, разработанная Гимзой³. Его имя теперь увековечено в названии самих окрашенных дисков, которые принято

¹ Авторами этого эпохального доказательства были яванец китайского происхождения цитогенетик Джо Хин Тджио (Joe Hin Tjio [1916–2001]) в период его работы в Университете Лунда (Швеция) и шведский ботаник и генетик Альберт Леван (Albert Levan [1905–1998]).

² Каждая пара аутосом представлена *гомологичными*, т.е. морфологически идентичными, экземплярами, и лишь пара половых хромосом является гетероморфной.

³ Густав Гимза (Gustav Giemsa [20.11.1867 — 10.06.1948]) — немецкий химик, бактериолог и фармацевт. Краситель Гимзы (раствор смеси азура, эозина и метиленового синего в метаноле и глицерине) есть незначительная модификация (добавка глицерина как стабилизатора) красителя, состав которого разработал русский терапевт, гематолог и маляриолог Дмитрий Леонидович Романовский [1861–1921]. Поэтому его называют также красителем (или раствором) Романовского — Гимзы.

сокращенно называть **G-дисками**, а сам метод — **G-методом** (дифференциального окрашивания хромосом). Рисунок («паттерн») окрашивания оказывается специфичным для каждой хромосомы, и это позволяет вполне однозначно идентифицировать все хромосомы в метафазной пластинке. Физическая природа темного и светлого окрашивания полос до конца не выяснена, но имеет место определенная корреляция с составом генов. Темные G-диски относительно более богаты парами оснований А=Т (аденин-тимин), в них довольно мало генов, и они позднее других реплицируются в клеточном цикле. Светлые G-диски более обогащены парами оснований G=C (гуанин-цитозин), содержат большую часть генов и реплицируются раньше — в S-фазе клеточного цикла (основная стадия репликации ДНК). Общее содержание GC в геноме человека составляет 41%, но оно явно неравномерно: темные G-диски содержат примерно 37% GC, а светлые — около 45%. Увы, мы до сих пор не знаем, имеют ли эти различия в составе эволюционное или функциональное значение.

В 1971 г. в Париже собрался комитет экспертов, который рекомендовал приписывать хромосомам номера, начиная с № 1 — для наибольшей хромосомы и кончая № 22 — для наименьшей хромосомы. Однако цитологическая техника все еще оставалась несовершенной, и потому была допущена небольшая ошибка: наибольший номер 22 был приписан предпоследней из самых малых хромосоме. В действительности же наименьшей является хромосома 21. *Парижская конференция ввела также систему нумерации дисков и буквенные обозначения: p — для короткого плеча каждой хромосомы и q — для длинного плеча.* Половые хромосомы не нумеруются; они обозначаются как X- и Y-хромосомы. Самки млекопитающих имеют две X-хромосомы, а самцы имеют одну X- и одну Y-хромосому. X-хромосома имеет средний размер и содержит среднее число генов. Y-хромосома довольно маленькая; она содержит совсем мало генов, некоторые из которых необходимы для развития организма мужского пола, но в основном она представлена **гетерохроматином** — протяженными участками повторяющихся и высоко конденсированных последовательностей, которые не кодируют никаких белков.

Полный набор хромосом человека называется **кариотипом**. Обычно нормальный кариотип человека принято обозначать как 46,XY — для мужчин и 46,XX — для женщин. С некоторыми аномальными кариотипами вы ознакомитесь в гл. 2. Результат анализа метафазной пластинки после окрашивания по G-методу (или какому-либо

иному методу) можно наглядно представить в виде упорядоченной диаграммы, которая называется *идиограммой* или *кариограммой*¹. Идиограмма кариотипа человека представлена ниже (рис. 1.2).

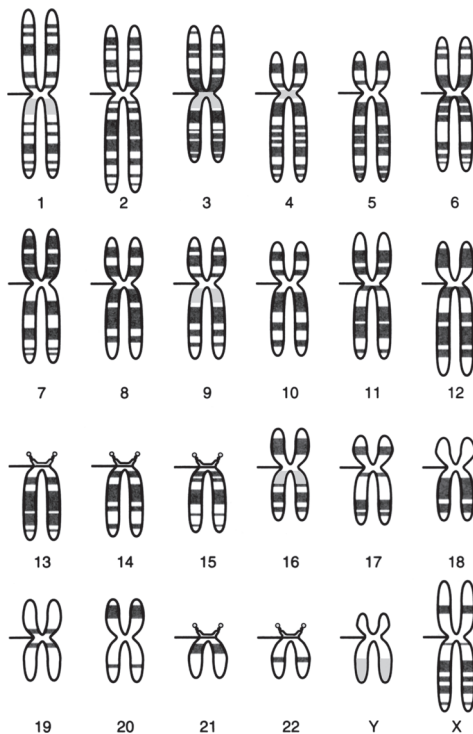


Рис. 1.2. Идиограмма хромосом человека. Показаны поперечные интенсивно окрашиваемые полосы (темные и светлые), наблюдаемые при использовании двух различных красителей. По: ISCN 1985: An International System for Human Cytogenetic Nomenclature (1985): Report of the Standing Committee on Human Cytogenetic Nomenclature. Eds: D. G. Harnden, et al. (с разрешения S. Karger SG)

¹ *Идиограмма* (idiogram), или *кариограмма* (karyogram), – графическое (схематическое) изображение («раскладка») диплоидного набора хромосом тестируемого объекта, систематизированное по микрофотографиям путем подбора гомологичных пар и распределения по морфологическим параметрам (размеры, соотношение длины хромосомных плеч, расположение вторичных перетяжек и спутничных элементов, локализация дифференциально окрашенных зон и другие признаки).

В 1970—1980-е гг. в генетике преобладало построение и изучение *генетических карт*. Одним из плодотворных подходов было создание *гибридных клеток*, которые содержали полный геном грызуна (мыши или крысы) и одну или несколько хромосом человека. Когда у такого клеточного гибрида удавалось идентифицировать специфический продукт гена человека или ферментативную активность, то можно было заключить, какая из хромосом содержит соответствующий ген. Впоследствии *бурное развитие технологии рекомбинантных ДНК сделало возможным клонировать гены человека*. Это позволило производить в клетках микроорганизмов неограниченное количество небольших участков (фрагментов) генома человека. В сочетании с разнообразными новыми физическими методами и традиционными методами генетики клонирование невероятно повысило мощь генетического картирования и анализа.

Решающим достижением молекулярной генетики стала разработка методов секвенирования ДНК в 1977 г. До этого момента возможность интенсивного изучения генома человека и молекулярных основ генетических болезней казалась безнадежным делом. К середине 80-х гг. секвенирование ДНК было столь существенно улучшено, что стало возможным практическое осуществление крупномасштабных проектов по секвенированию геномов различных видов. Группа ученых, участвующих в реализации биологической программы Министерства энергетики США, нацеленной на изучение скорости мутирования у человека (см. гл. 2), осознали, что теперь появилась возможность изучать мутации на молекулярном уровне. Однако для поиска и изучения мутаций на уровне ДНК необходимо знать также исходную (нормальную) последовательность ДНК. Так было положено начало *проекту «Геном человека»*, который официально стартовал осенью 1990 г. благодаря совместным усилиям Министерства энергетики и Национальных институтов здоровья¹, которые, помимо других организационных мероприятий, учредили для этого специальный Национальный институт исследования генома человека². Проект сразу стал международным, в который, наравне с США, наиболее значительный вклад внесли также ученые из 16 институтов Великобритании, Франции, Германии, Японии и Ки-

¹ В России аналогами Национальных институтов здоровья являются научно-исследовательские институты Российской академии медицинских наук.

² См. <http://www.genome.gov>.

тая. Эти страны образовали *Международный консорциум по секвенированию генома человека*¹.

Главной целью проекта «Геном человека» было секвенировать все 3 миллиарда пар нуклеотидов в гаплоидном геноме человека и идентифицировать все гены. Отчет о главном прогрессе в этой области был опубликован в 2001 г. К этому времени было секвенировано примерно 90% генома, но значительная часть всей последовательности оставалась все еще в виде фрагментов длиной в несколько тысяч пар нуклеотидов. Очевидно, что это сильно ограничивает пользу от таких данных. Это аналогично тому, как если бы все предложения в этой книге были написаны на отдельных обрывках бумаги и, после того как в книжном магазине вы заплатили за книгу, вы получили мешок с этими обрывками, перемешанными в случайном порядке. Из такой «книги» вы вряд ли бы многое узнали о генетике человека!

В следующие два года почти все такие бреши были заполнены, и о завершении практически всей последовательности было объявлено в апреле 2003 г. — подходящая дата, чтобы отпраздновать 50-летний юбилей модели структуры ДНК Уотсона и Крика! Около 99% районов, содержащих гены, были секвенированы с точностью 99,99%. Пока еще остается менее 400 брешей, и в среднем непрерывные участки составляют более 27 миллионов пар нуклеотидов, а общий размер генома, как уже говорилось, составляет примерно 3,2 миллиарда пар нуклеотидов². Неожиданным результатом оказалось то, что *общее число генов у человека находится в пределах от 30 000 до 35 000*, что намного меньше,

¹ У нас с идеей секвенирования генома человека выступил в 1988 г. академик Александр Александрович Баев [10.01(28.12)1903 — 31.12.1994], и в 1989 г. в стране был организован Научный совет по программе «Геном человека». В 1990 г. была создана Международная организация по изучению генома человека (HUGO), вице-президентом которой в течение нескольких лет был академик Андрей Дарьевич Мирзабеков [19.10.1937 — 13.07.2003]. С самого начала работ по геномному проекту ученые договорились об открытости и доступности всей получаемой информации для его участников независимо от их вклада и государственной принадлежности. Работа над всеми 23 хромосомами человека была поделена между странами-участниками. Российские ученые должны были исследовать структуру 3-й и 19-й хромосом. Однако вскоре финансирование этих работ в нашей стране было урезано, и реального участия в секвенировании генома человека Россия, к сожалению, не принимала.

² Самой последней, лишь недавно, в 2006 г. была полностью секвенирована самая большая хромосома 1.

чем предполагалось ранее, и всего в два раза больше, чем число генов у плодовой мушки дрозофилы или у микроскопического круглого червя (почвенной нематоды)¹.

Завершение секвенирования генома человека является важной вехой в генетике человека, но это отнюдь не означает, что мы знаем функции всех этих генов. Анализ функции тысяч генов в норме и при болезнях будет гораздо более сложным проектом, чем секвенирование генома. Сейчас планируется использовать фундаментальные данные о последовательности генома в разнообразных сферах, о чем будет рассказано в следующих главах данной книги. Здесь же достаточно сказать, что наличие полной последовательности генома человека открывает возможность понять биологию человека и его болезни на таком уровне, о котором нельзя было даже вообразить всего лишь несколько лет.

Что такое ген?

Ген определяется как участок ДНК, который транскрибируется в РНК-копию одной из нитей ДНК (этот процесс будет описан далее в этой главе). Большинство генов являются участками ДНК, которые несут информацию о последовательности аминокислотных остатков в белке, однако некоторые гены кодируют только РНК. Фактически все метаболические функции живых клеток опосредуются белками, но в то же время другие белки образуют множество внутриклеточных и внеклеточных структур. Со всеми генами связаны регуляторные последовательности ДНК, которые являются такими участками, к которым присоединяются белки, определяющие, будет ли ген экспрессирован в данное время и в данном месте. Некоторые генетики называют такие регуляторные последовательности тоже генами.

Рис. 1.3 поможет освежить Вашу память относительно основных элементов структуры ДНК. В окнах *a* и *б* показаны структуры четырех оснований: двух пуринов – аденина (А) и гуанина (G) и двух пиримидинов – тимина (Т) и цитозина (С). В окне *в* показана общая структура нуклеотида, являющегося фундаментальной единицей, из которых собираются и ДНК и РНК. Заметьте, что символы 5' и 3', которые обозначают концы

¹ Плодовая мушка *Drosophila melanogaster* и нематода *Caenorhabditis elegans* – два модельных животных объекта, с помощью которых были получены и получаются крупнейшие достижения классической и молекулярной генетики.

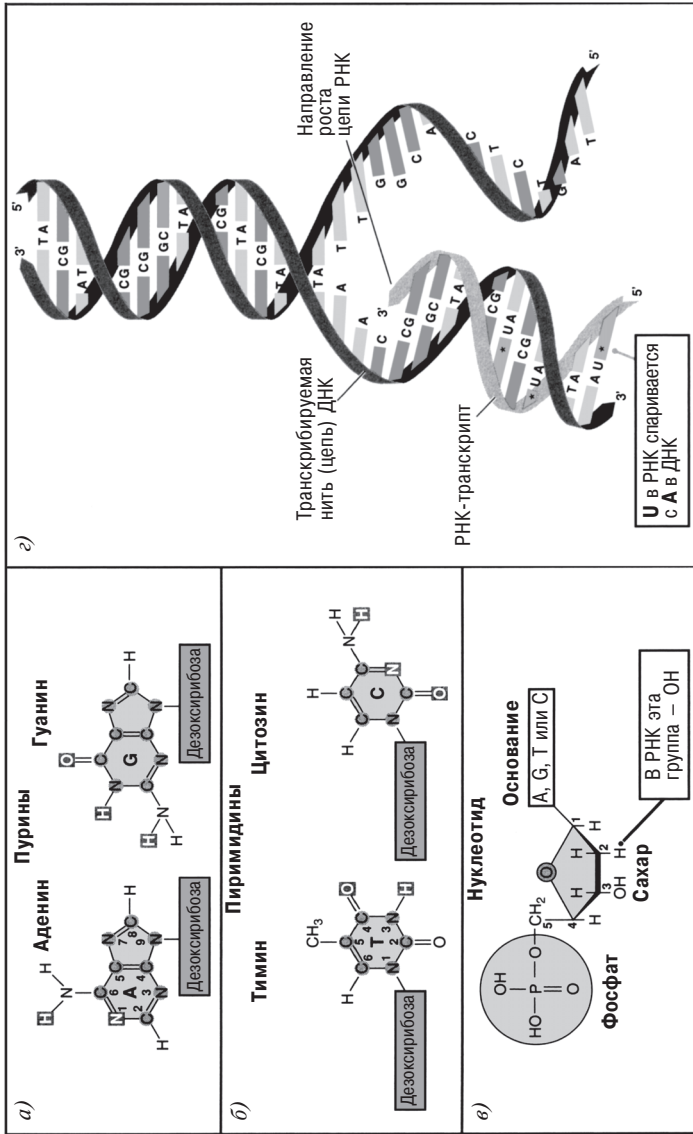


Рис. 1.3. Структура ДНК. Белыми буквами на темном фоне выделены те атомы вололота (Н), кислорода (О) и азота (N), которые участвуют в образовании водородных связей между парами оснований А=Т и G≡C в двойной спирали ДНК

полинуклеотидной цепи, являются номерами атомов углерода в остатке сахара в составе нуклеотида. Первый нуклеотид в начале полинуклеотидной цепи имеет 5'-фосфатный остаток, а последний нуклеотид в конце цепи имеет 3'-гидроксильную группу. В окне ε показана двухспиральная структура ДНК, нити которой удерживаются вместе водородными связями между парами оснований (A=T или G≡C). Лентообразные линии серого цвета символизируют сахарофосфатный остов, к которому присоединены основания. На этом же рисунке схематически показан также процесс транскрипции (синтеза комплементарной нити РНК), который мы обсудим далее в этой главе.

Прежде чем углубиться в структуру гена, следует указать на некоторые терминологические и аббревиатурные несуразности. Поскольку элементарными единицами ДНК (или РНК) являются нуклеотиды, то длину двунитевых последовательностей ДНК (или РНК) логично выражать в *парах нуклеотидов*, сокращенно *п.н.* Но иногда, следуя англоязычной традиции, используют словосочетание *пары оснований* и соответственно сокращение – п.о. Тысячу пар нуклеотидов обычно обозначают как *т.п.н.* Но иногда неоправданно используется лабораторный жаргон, являющийся калькой с английского: «килобаза» (kilobase) и сокращение kb – вместо т.п.н., а также Mb – «мегабаза» (megabase) – вместо миллиона пар нуклеотидов и Gb – «гигабаза» (gigabase) – вместо миллиарда пар нуклеотидов. Длину одонитевых ДНК и РНК измеряют, естественно, не парами, а просто числом нуклеотидов (или оснований). «Начало» каждой последовательности ДНК или РНК обозначается как 5' (читается как «пять-штрих»), а окончание – как 3' («три-штрих»). Эти номера соответствуют структуре нуклеотидов (см. рис. 1.3в) и тому, каким путем они собираются в нити ДНК или РНК (см. гл. 2).

Типичный ген человека представляет собой чередование экзонов и интронов. Экзоны являются участками ДНК, которые будут представлены в зрелой матричной РНК (мРНК), которая образуется в процессе экспрессии гена. Большинство экзонов содержат информацию о последовательности *аминокислот* – элементарных единиц белков. Кроме этого, в начале и в конце мРНК находятся такие экзоны, которые не кодируют последовательность аминокислот, но могут содержать различные типы регуляторной информации. *Интроны же являются такими участками генов, которые расположены между экзонами и отсутствуют в зрелых мРНК.*

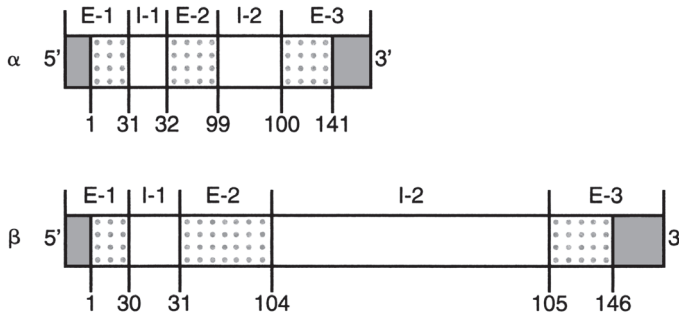


Рис. 1.4. Структура глوبيновых генов человека. E-1, E-2 и E-3 являются экзонами; I-1 и I-2 – интроны. Закрашенные области на концах войдут в состав РНК-транскрипта (мРНК – матричной или информационной РНК), но не будут транслироваться в белковую последовательность. Числа на границах между интронами и экзонами обозначают номера кодируемых аминокислот

Отношение числа экзонов к числу интронов варьирует достаточно широко. Лишь небольшое количество генов не содержат интронов, в то же время есть гены, в которых интроны составляют более 95% их длины. Функция интронов и их эволюционное возникновение до сих пор не до конца поняты, но принято считать, что наличие генов, сконструированных из ряда коротких кодирующих последовательностей (экзонов), обеспечивает эволюционную пластичность. На рис. 1.4 схематически представлена экзон-интронная структура двух хорошо изученных генов, которые кодируют α - и β -полипептидные цепи глобиновой части гемоглобина.

Согласно отчету Консорциума по секвенированию *в среднем ген человека содержит 27 т.п.н.* Если мы умножим 27 т.п.н. на 30000 генов, то получим, что *гены человека занимают 0,8 млн.п.н., т.е.* примерно лишь одну четверть от всего генома. Консорциум по секвенированию сообщает нам еще некоторые интересные количественные данные о генах человека. *Среднее число экзонов, приходящееся на один ген, равно примерно 8* (соответственно среднее число интронов в гене должно быть равно 7). Средний размер экзона составляет 145 п.н., а средний размер интрона – 3365 п.н. Легко сосчитать, что в среднем экзоны составляют менее 5% от общей длины гена. В среднем суммарная длина кодирующих экзонов в одном гене составляет 1340 п.н.; этого достаточно, чтобы образовать белок длиной в 447 аминокислотных остатков. Однако имеет

место громадная вариабельность в размере генов, числе интронов, размере кодируемых белков и т.п. Размер наибольшего из известных генов превышает 2,4 млн.п.н., выявлены интроны длиной более 30 т.п.н., а некоторые белки содержат более 3000 аминокислотных остатков.

Упомянутая оценка числа генов в пределах 30 000—35 000 была получена путем компьютерного анализа геномных последовательностей. Сначала подсчитали число *известных генов* (что, естественно, было достаточно легко сделать) и к нему добавили число генов, наличие которых можно было предсказать из оценок числа возможных экзонов, числа сопряженных экзон-интронных последовательностей и некоторых других характеристик. Понятно, что число *предсказанных генов* довольно неопределенно, поскольку, с одной стороны, некоторые гены компьютер может не распознать, а с другой стороны, компьютер может ошибочно предсказать гены, которые в реальности не существуют.

Один из способов обнаруживать неизвестные гены заключается в установлении сходства с последовательностью известного гена. *Многие белок-кодирующие гены образуют семейства, которые представляют собой группы генов, имеющих значительное сходство в своих последовательностях.* Основным событием, приводящим к появлению семейств генов, является дупликация гена, которая может случайно возникать вследствие ошибок репликации и рекомбинации ДНК (см. гл. 2). Когда образуются две копии гена, одна из копий может мутировать таким образом, что образуется несколько измененный ген, который будет кодировать белок со свойствами, слегка отличными от оригинала. Если различия, приобретенные новым белком, придадут ему некоторые преимущества, то процесс отбора может увековечить их в последующих поколениях. Вследствие очередных ошибок в ходе репликации или рекомбинации число таких по-разному измененных копий генов может разрастись, и в результате получится мультигенное семейство и, кроме того, они могут рассредоточиться по геному на отдаленные расстояния.

Классическим примером генных семейств у человека является кластер генов β -глобина на хромосоме 11 и генов α -глобина на хромосоме 16. Их схема представлена на рис. 1.5.

Обратите внимание на наличие *псевдогенов* в каждом кластере (они обозначены греческой буквой ψ). Псевдогены образуются из дуплицированных генов, одна из копий которых приобретает такие мутации, которые делают невозможным ее экспрессию. Известны многие другие генные семейства, число членов в кото-

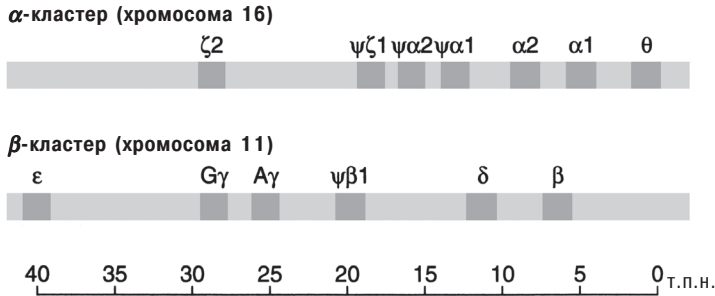


Рис. 1.5. Кластеры глобиновых генов

($\psi\zeta 1$, $\psi\alpha 2$, $\psi\alpha 1$ и $\psi\beta$ – псевдогены. Каждый из глобинов синтезируется в определенном органе и на определенной стадии развития. Глобины ζ и ϵ вырабатываются на ранних стадиях развития эмбриона клетками желточного мешка, и гемоглобин представляет собой тетрамер $\zeta 2\epsilon 2$. К концу третьего месяца его выработка прекращается и начинается синтез фетальных (утробных) глобинов $A\gamma$ и $G\gamma$ в клетках печени и селезенки. Образуется тетрамер фетального гемоглобина $A\gamma 2G\gamma 2$ (Hb-F). В период постэмбрионного развития синтезируются в основном глобины α и β клетками костного мозга. Большая часть гемоглобина взрослого человека состоит из тетрамера $\alpha 2\beta 2$ (Hb-A). – Прим. пер.)

рых исчисляется десятками; примерами являются семейства генов для актинов, миозинов, аполипептопротеинов, гистонов и иммуноглобулинов. Когда анализируются удаленные друг от друга семейства, то многие из них можно назвать сверхсемействами, потому что они насчитывают сотни членов.

Давайте теперь вычислим еще одно интересное число. Какая доля генома человека содержит информацию для кодирования белков? Умножив 1340 п.н. на 30 000 генов, мы получаем, что за кодирование белков отвечают 40 200 000 п.н. Разделив это число на размер гаплоидного генома человека (3,2 млрд.п.н.), мы приходим к выводу, что *только 1,25% нашего генома несут информацию о кодировании белков*. Эти числа приблизительные, так что не удивляйтесь, если в других источниках вы встретите несколько отличные от этих числа. *Важно то, что лишь очень малая доля человеческой ДНК кодирует белки.*

Что же представляет собой остальная часть генома? Мы знаем, что 20–25% занимают интроны, но большая часть остальной ДНК является *межгенной ДНК*. Значительную часть межгенной

ДНК составляют **регуляторные последовательности**, которые мы сейчас обсудим. Существует несколько групп **генов, которые не кодируют белки**; продуктами таких генов являются РНК, которые играют важную роль во многих клеточных процессах и структурах. Иногда из них образуются нуклеопротеиновые структуры, иногда они нацеливают ферменты на другие РНК. В этой книге я буду неоднократно упоминать специфические классы функционально активных **некодирующих РНК**. Кроме того, некоторые части генома играют структурную роль. Тем не менее у нас нет никакого очевидного объяснения, почему так много ДНК не участвует непосредственно ни в качестве структуры для генов, ни в любых других функциях. Некоторое понимание проблемы «избытка» ДНК можно получить, взглянув на геном с несколько иной стороны. Что мы сейчас и сделаем.

Повторяющаяся ДНК

*В любом сложном геноме ДНК можно подразделить на два типа: **однокопийную ДНК** (т.е. последовательности, представленные в гаплоидном геноме единственными экземплярами) и **повторяющуюся ДНК** (последовательности, представленные в гаплоидном геноме многократно). Примерно 50% генома человека представляют собой повторяющуюся ДНК. Популярный термин для большинства повторяющихся ДНК и для некоторых однокопийных ДНК, которые не являются частью генов, — это «мусорная (junk) ДНК». Да, скорее всего в настоящее время наши геномы содержат некоторое количество ДНК, которые не несут никакой функции и могут вполне законно считаться «мусорными», но чем больше мы узнаем о геномах и о регуляции экспрессии генов, тем больше мы открываем новых функций для ДНК, о которых раньше мы не имели никакого представления. Кроме того, значительная часть ДНК, которая в настоящее время явно не используется, определенно является запасным материалом для эволюции генома. Таким образом, если мыслить о биологических видах как о динамических сущностях, изменяющихся (эволюционирующих) во времени, то вполне возможно, что «мусора» в их геномах не так уж и много.*

*Различают два класса повторяющейся ДНК: (1) **тандемно повторяющиеся последовательности (повторы)**, которые расположены друг за другом «голова к хвосту», и (2) **диспергированные повторы**, которые разбросаны по всему геному, причем чаще всего они бывают представлены одной копией в данном месте (сайте).*

Тандемно повторяющиеся последовательности

Основным классом тандемно повторяющихся последовательностей является *центромерная ДНК*. Наиболее распространенный тип центромерной ДНК называется альфоидной (alphoid) ДНК¹, повторяющиеся единицы которой имеют длину примерно в 170 п.н. Эти единицы образуют ряды, длина которых варьирует от 250 т.п.н. до 5 млн.п.н., и они составляют не менее 3% генома. Внутри одного ряда повторы неидентичны, они немного различаются, и между хромосомами эти различия еще больше. Центромерная ДНК образует центромеры, сложные структуры, которые кроме ДНК содержат белки нескольких типов, к которым присоединяются нити веретена в процессе деления клетки.

Тандемно повторяющиеся последовательности найдены также в *теломерах*, которые расположены на концах каждой хромосомы. У человека теломерными последовательностями являются GGGTTA; в разных хромосомах они повторяются от 250 до 1500 раз. В последние годы *теломерная ДНК* стала предметом пристального внимания исследователей из-за того, что была обнаружена связь между укорочением теломер и старением. Я не буду здесь детально обсуждать этот вопрос, скажу только, что основная идея состоит в том, что укорочение теломер является одним из последствий процесса репликации ДНК, который происходит перед каждым актом клеточного деления². Если такое продолжается достаточно долго, то теломерные последовательности элиминируются и станут повреждаться прилежащие к ним гены. В результате клетка может умереть или перестать делиться. Такое

¹ *Альфоидная* (от слов «альфа-сателлитная») ДНК впервые описана у зеленых мартышек рода *Cercopithecus* в виде повторов последовательности размером 172 пары нуклеотидов. Она включает значительное число нуклеотидных замен, что в ряде случаев делает ее хромосомоспецифичной. В связи с этим соответствующие зонды широко используются для маркирования отдельных хромосом методом гибридизации *in situ* в интерфазных клетках, в частности, в случаях моно- или трисомии — например, в пренатальной диагностике синдрома Дауна.

² Гипотезу о связи укорочения теломер и старения впервые сформулировал в 1971 г. российский биолог Алексей Матвеевич Оловников. Лишь в 1998 г. его вывод экспериментально подтвердили американские исследователи. В 2003 г. автор выдвинул «гипотезу редусомного старения» и идею о том, что старение есть «универсальная хроническая болезнь количественных признаков». См.: <http://www.medline.ru/public/art/tom4/art28.phtml>. См.: сайт автора: <http://www.chronos.msu.ru/nameindex/olovnikov.html>.

происходит в большинстве нормальных соматических клеток, и поэтому в организме или в клеточной культуре они имеют ограниченную способность к размножению. Однако в зародышевых (половых) клетках, стволовых клетках и разнообразных раковых клетках имеется фермент *теломераза*, способный восстанавливать теломерные последовательности, которые в его отсутствие укорачиваются при каждом акте репликации ДНК. Теломераза является необычным ферментом, у которого имеются два компонента: РНК и белок; при этом РНК служит матрицей для восстановления утраченных теломерных повторов. Рис. 1.6 в общих чертах показывает, как осуществляется этот процесс.

Некоторые другие классы tandemно повторяющихся последовательностей являются генами для некодирующих РНК. Наиболее известными из них являются гены для *рибосомной РНК (рРНК)*. В геноме человека имеются пять групп таких генов, каждая из кото-

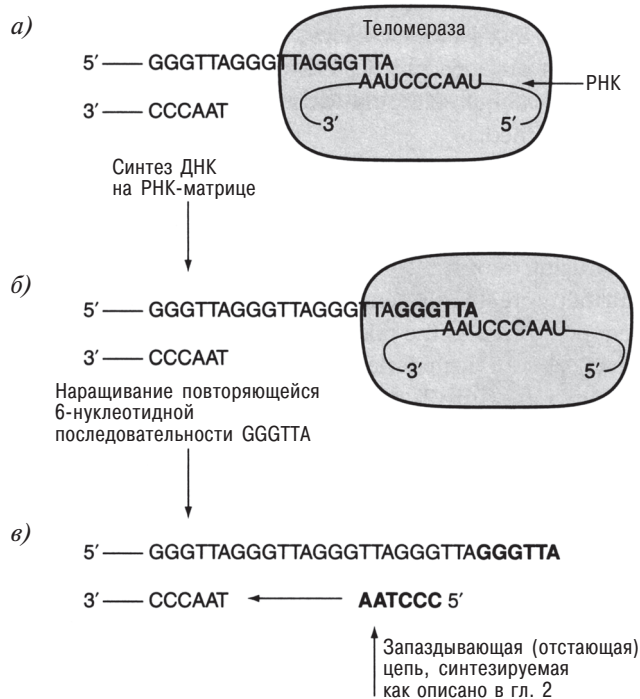


Рис. 1.6. Поддержание структуры теломер

рых содержит около 60 копий. Они расположены в коротких плечах акроцентрических хромосом 13, 14, 15, 21 и 22. Эти кластеры для рРНК вместе с некоторыми дополнительными участками ДНК называются *ядрышковыми организаторами*, потому что ядрышко может быть образовано каждым из них. Ядрышки расположены в клеточном ядре и являются фабриками по сборке рибосом. Существует более 80 типов рибосомных белков. Они синтезируются в цитоплазме и мигрируют в ядро, где они связываются с рРНК, которая синтезируется в ядрышке. В каждом ядрышке накапливается также множество других белков, где они принимают участие в сборке рибосом. Кроме того, в сборке рибосом участвуют небольшие некодирующие РНК нескольких типов. В хромосоме 1 расположен еще один кластер tandemно повторяющихся генов для *5S-рРНК*, который также является важным компонентом рибосом.

Диспергированные повторяющиеся последовательности

Диспергированные повторяющиеся последовательности чаще всего разбросаны по геному по отдельности, а не кластерами. Согласно размеру их подразделяют на две группы: длинные диспергированные элементы, обозначаемые как LINE (Long INterspersed Elements), и короткие диспергированные элементы, обозначаемые как SINE (Short INterspersed Elements). Оба класса являются подвижными (мобильными) генетическими элементами, которые называются *ретротранспозонами*. Полностью функционирующий ретротранспозон способен размножить либо сам себя, либо родственные последовательности, как это будет описано в следующем абзаце. Ретротранспозоны могут возникать как ретровирусы, которые мы рассмотрим в гл. 4 и 6.

Одним из важнейших классов диспергированных повторяющихся последовательностей является LINE-1 или группа L1, которая представлена в геноме человека в количестве до 500 000 копий и составляет примерно 15% от всего генома. Большинство элементов L1 являются укороченными копиями полноценных единиц, длина которых составляет около 5000 п.н., но несколько тысяч элементов L1 имеют полную длину. Только 40–50 из них функционально активны, т.е. они кодируют несколько белков, которые способны вызывать транспозицию либо самого элемента L1, либо некоторых других мобильных элементов. Элементы L1 несут два гена (называемых открытыми рамками считывания – ORF) (см. рис. 1.7). ORF1 кодирует белок, связывающийся с нуклеиновыми кислотами (p40), ORF2

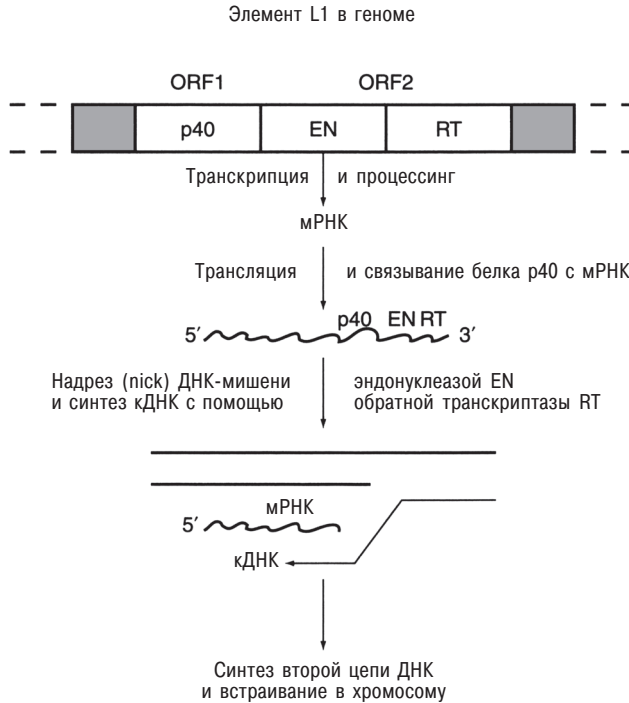


Рис. 1.7. Ретротранспозиция элемента L1

(ORF1 и ORF2 — гены, называемые в данном случае *открытыми рамками считывания* (open reading frame). ORF1 кодирует один белок р40, а ORF2 кодирует два белка: EN — эндонуклеазу и RT — обратную транскриптазу. Белок р40 имеет высокое сродство для связывания с РНК и образует рибонуклеопротеиновую частицу с РНК элемента L1. Одновременно он является шапероном для нуклеиновой кислоты, и эта его активность необходима для ретротранспозиции L1. Эндонуклеаза EN осуществляет надрез геномной ДНК, вследствие чего становится возможным синтез кДНК по матрице мРНК с помощью обратной транскриптазы RT. — *Прим. пер.*)

кодирует и *обратную транскриптазу* (фермент, который использует мРНК как матрицу и создает комплементарную ей одноцепочечную ДНК-копию, кДНК) и *эндонуклеазу*, которая производит надрезы в геномной ДНК, куда может встраиваться новая кДНК.

Когда мРНК из элемента L1 транслируется, то образующиеся белки обычно связываются непосредственно со своей мРНК. Та-

кой комплекс белка с РНК перемещается в ядро, где эндонуклеаза разрезает одну из цепей ДНК, и в результате образуется свободный конец. Обратная транскриптаза использует этот свободный конец в качестве затравки и создает ДНК-копию мРНК элемента L1. В конце концов образуется вторая цепь кДНК и двуниевая молекула встраивается в хромосому на место одностевого разрыва. Мы до сих пор не знаем, почему только часть элемента L1 является наиболее частым продуктом ретротранспозиции.

Считается, что обратная транскриптаза ответственна также за образование *процессированных псевдогенов*, которые являются ДНК-копиями мРНК и которые встраиваются в места, неродственные (негомологичные) исходному гену, из которого произошла скопированная мРНК. Процессированные псевдогены не содержат интронов и обычно неспособны экспрессироваться в виде полипептидов (хотя иногда и случаются исключения). Происходит это либо из-за того, что они не имеют регуляторных последовательностей, либо потому, что они содержат мутации. Псевдогены (как процессированные, так и обычные) довольно распространены и в геноме человека составляют 0,5–1%. Например, секвенирование хромосомы 22 выявило 134 псевдогена.

Наибольший класс элементов SINE состоит из последовательностей Alu (название происходит от названия фермента рестрикции Alu I — эндонуклеазы, которая расщепляет ДНК в местах нахождения специфичных коротких последовательностей и может использоваться для вырезания последовательностей Alu из геномной ДНК)¹. В геноме человека находится примерно миллион последовательностей Alu, которые равняются примерно 10–12% всей ДНК. Длина основной единицы составляет примерно 300 п.н., но в классе Alu существует много различных последовательностей. В основном они находятся между генами и внутри интронов, но изредка они могут быть включены в мРНК. Последовательности Alu не кодируют белков, и поэтому они неспособны сами перемещаться из одного места в другое. Однако многие последовательности Alu транскрибируются, и на их концах существуют некоторые короткие последовательности, подобные РНК у элементов L1. Поэтому распространено мнение, что ферменты, которые производятся элементами L1, участвуют в *ретротранспозиции последовательностей Alu*, хотя окончательные доказательства этого все еще не получены.

¹ В свою очередь Alu обозначает источник рестриктазы — азотфиксирующие бактерии *Arthrobacter luteus*.