

ОГЛАВЛЕНИЕ

Авторский коллектив

Предисловие

Книга 1. Молекулярная цитология

Глава 1. Световая микроскопия. Строение и функция
клеточного ядра

Глава 2. Структурная организация эукариотической клетки.
Строение и функция плазматической мембраны

Глава 3. Закономерности существования клетки во времени

Глава 4. Половые клетки. Мейоз

Книга 2. Общая генетика

Принятые сокращения и условные обозначения207

Глава 5. Структура и экспрессия гена.....208

Глава 6. Закономерности наследования. Мобильные
генетические элементы.....257

Глава 7. Хромосомная теория наследственности.
Комбинативная и мутационная изменчивость.....327

Глава 8. Фенотипическая изменчивость. Эпигенетическая
модификация.....385

Толковый словарь терминов.....423

Литература.....447

Электронные справочные ресурсы учебника.....448

Предметный указатель.....450

Именной указатель.....452

Книга 3. Медицинская генетика

Глава 9. Клинико-генеалогический метод.
Косвенная ДНК-диагностика

Глава 10. Хромосомы человека. Цитогенетическая диагностика

Глава 11. Полиморфизм генов. Прямая ДНК-диагностика

Глава 12. Геном человека. Генная инженерия

Книга 4. Молекулярная биология развития

Глава 13. Общая эмбриология

Глава 14. Генетика раннего эмбриогенеза

Глава 15. Филогенетика живых систем

Глава 16. Генетика и антропология

Книга 5. Среда обитания человека

Глава 17. Неживая природа

Глава 18. Микроорганизмы (вирусы и прокариоты)
и их переносчики

Глава 19. Простейшие одноклеточные организмы и их переносчики

Глава 20. Грибы и грибоподобные организмы

Книга 6. Медицинская гельминтология

Глава 21. Эволюция червей и их симбиотических отношений
с человеком

Глава 22. Трематоды

Глава 23. Цестоды

Глава 24. Нематоды

Книга 7. Справочно-методические материалы

Глава 25. Объединённый толковый словарь терминов

Глава 26. Объединённая библиография

Глава 27. Именной указатель

Глава 28. Предметный указатель

Глава 29. Принятые сокращения и условные обозначения

Книга 8. Хрестоматия и дополнительные материалы

Глава 30. Теория биологии и медицины: предметная область
и создатели – исторический ракурс (от Аристотеля
до молекулярных биологов)

Глава 31. Методология и практика научной медицины и врачебного
искусства

Глава 32. Хронология научно-технических и методических
достижений в биологии и медицине

Глава 33. Список лауреатов Нобелевской премии по физиологии
или медицине (1901–2021)

Глава 34. Знаменитые умы о биологии, медицине и науке вообще

Глава 5

СТРУКТУРА И ЭКСПРЕССИЯ ГЕНА

ТЕОРЕТИЧЕСКИЙ ОБЗОР

Ген — единица наследственности [участок дезоксирибонуклеиновой кислоты (ДНК), с которого копируется рибонуклеиновая кислота (РНК)], занимающая специфическое место (локус) в хромосоме, способная к самовоспроизведению в клеточном цикле. Термин «ген» был предложен в 1909 г. датским ботаником Вильгельмом Йогансенем. В 1920 г. Ганс Винклер ввёл понятие генома — совокупности генов, заключённых в гаплоидном наборе хромосом организмов одного биологического вида. Геном человека длиной 3,2 млрд пар нуклеотидов (п.н.) расшифрован в 2003 г., содержит примерно от 20 000 до 25 000 генов. Окончательное количество генов не установлено.

Структура гена эукариот

Размеры генов варьируют от 250 п.н. (ген *IGF2* — инсулиноподобный фактор роста II, 67 аминокислот) до 2,2 млн п.н. (ген *DMD* — дистрофин, 3685 аминокислот). Общепринятая модель строения гена — экзон-интронная структура (рис. 5.1).

Экзон — последовательность ДНК, которая представлена в зрелой РНК. В составе гена должен присутствовать как минимум один экзон [например, в генах транспортной рибонуклеиновой кислоты (тРНК), гистонов]. Максимальное количество экзонов представлено в гене мышечного белка титина — 364 экзона. В среднем в гене содержится 8 экзонов.

- ▶ Фактор инициации транскрипции 5'-АСТТ(Т/С)ТГ-3' входит в состав первого экзона.
- ▶ Фактор терминации транскрипции (менее определённая последовательность) входит в состав последнего экзона.

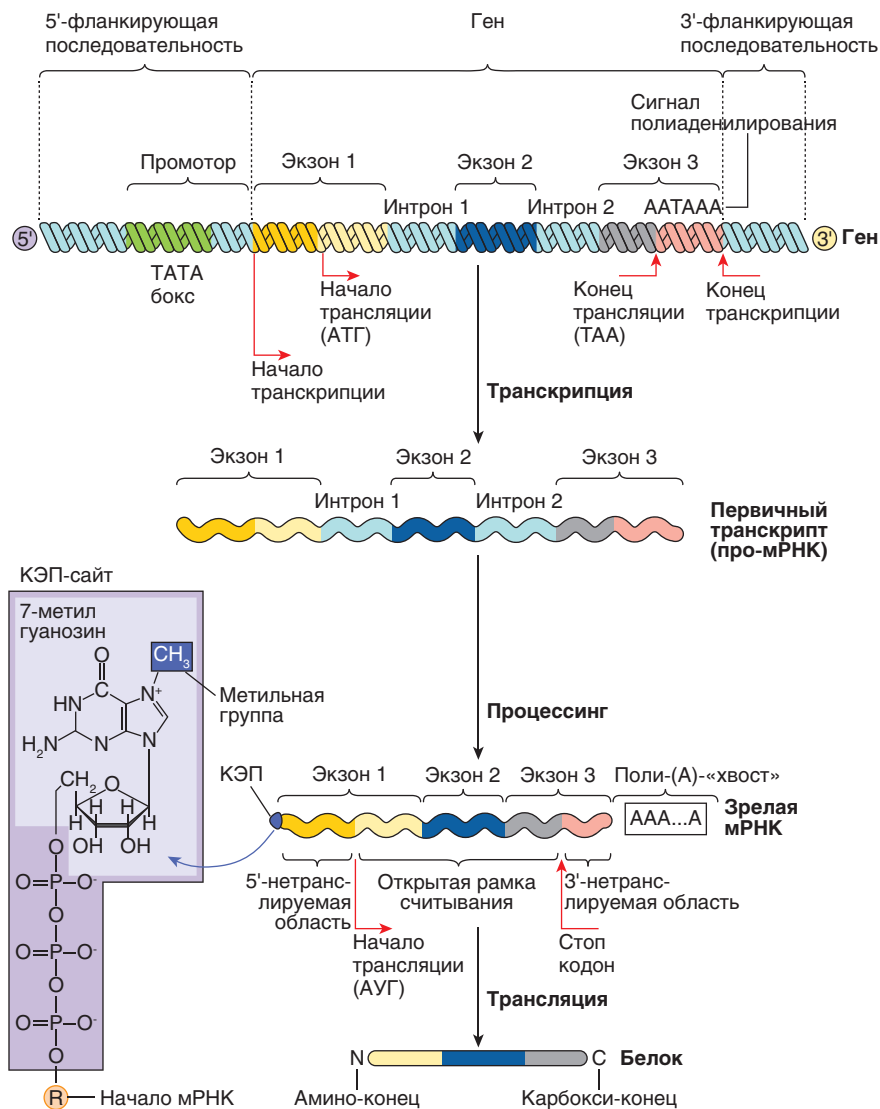


Рис. 5.1. Структура и экспрессия гена. КЭП-сайт (от англ. *cap* — шапка) — нуклеотид на 5'-конце транскриптов, синтезированных РНК-полимеразой II [по: Boron W.F., Boulpaep E.L., 2003]

Инtron — последовательность ДНК, включённая между экзонами, не входит в состав зрелой РНК. Интроны имеют определённые нуклеотидные последовательности, обуславливающие их границы с экзонами: на 5'-конце — GU последовательность, на 3'-конце — AG. Интроны содержат регуляторные элементы экспрессии гена и могут кодировать регуляторные РНК, например короткие интерферирующие РНК (см. ниже).

Сигнал полиаденилирования 5'-ААТAAA-3' входит в состав последнего экзона, начинается сразу после стоп-кодона (ТАА, TAG, TGA). Поли(А) сайты защищают матричную рибонуклеиновую кислоту (мРНК) от деградации.

5'- и 3'-фланкирующие последовательности. Копирование гена происходит в направлении 5' → 3'; на флангах (границах) находятся специфические сайты, ограничивающие ген и содержащие регуляторные элементы его транскрипции.

Регуляторные элементы — промотор, энхансеры, сайленсеры, инсуляторы — могут находиться за пределами сайта транскрипции и быть общими для нескольких генов.

- ▶ **Промотор** (от англ. *promoter* — «активатор», «ускоритель») — цис-регуляторная последовательность в 5'-области гена, определяющая место прикрепления РНК-полимеразы и интенсивность (частоту) транскрипции мРНК. Содержит ТАТА-боксы для связывания основного фактора транскрипции TFIID. Проксимальнее ТАТА-боксов содержатся GC-боксы (5'-GGGCGG-3') и СААТ-боксы (5'-ССААТ-3') для связывания дополнительных специфических белков, активирующих экспрессию генов. Активация только промотора недостаточна для экспрессии гена на физиологически значимом уровне.
- ▶ **Энхансеры** (от англ. *enhance* — «усиливать»), цис-позитивные регуляторные элементы, и **сайленсеры** (от англ. *silence* — «успокаивать» (гл.); «тишина» (сущ.)), цис-негативные регуляторные элементы, состоят из 6–12 нуклеотидов, специфически взаимодействующих с белками. Энхансеры связываются с белками-активаторами и усиливают экспрессию гена. Сайленсеры связываются с белками-репрессорами и блокируют экспрессию гена. Энхансеры и сайленсеры локализируются в 5'- или 3'- фланкирующих участках, или

интронах. Активность не зависит от их ориентации или локализации. Кроме того, они могут находиться на больших расстояниях от промотора (несколько сотен п.н.) и взаимодействуют с ним за счёт образования петель ДНК.

- ▶ **Инсуляторы** (англ. MAR — *matrix attachment regions*). Образуют дискретные функциональные домены — петли хромосом, ограничивающие влияние соседних регуляторных элементов. В состав петли могут входить специфические **последовательности, контролирующие локус** (англ. LCR — *locus-control region*) — позитивные цис-элементы, регулирующие активность нескольких генов (рис. 5.2).

Последовательность, контролирующая локус (LCR), впервые открыта в гене β-глобина. Семейство генов β-глобина включает в себя 5 генов (ε, γG, γA, δ, β) на 11-й хромосоме. Гены активируются

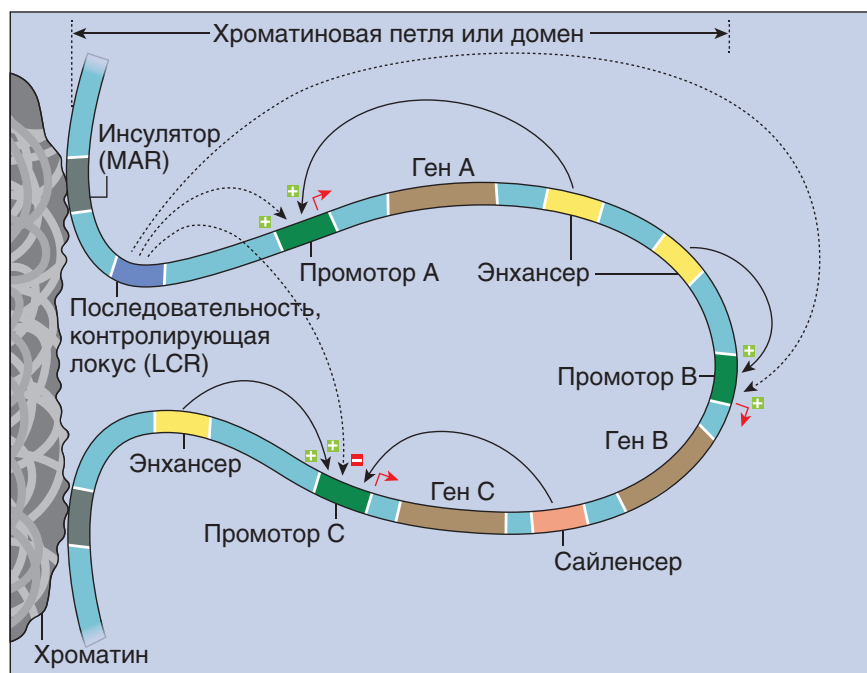


Рис. 5.2. Регуляторные элементы экспрессии гена [по: Boron W.F., Boulpaep E.L., 2003]

последовательно от 5'- к 3'-концу в онтогенезе в эритроидных клетках. β -Глобин экспрессируется в желточном мешке, γ G- и γ A-глобины — в печени плода, ϵ - и δ -глобины — в красном костном мозге в постнатальном периоде.

Последовательность, контролирующая локус, связывает факторы транскрипции GATA1 и NF-E2, изменяет форму хроматина, делая его доступным для факторов транскрипции, и выступает в роли энхансера.

Талассемия — гетерогенная группа генетических дефектов, характеризующихся отсутствием или сниженной экспрессией генов, кодирующих α - (α -талассемия) или β - (β -талассемия) цепи глобинов. При мутациях генов нарушается или полностью блокируется образование одной или нескольких цепей глобина, что приводит к развитию анемии. В случае β -талассемии мутации могут затрагивать последовательности, контролирующие локус.

Генетический код

В 1960-е гг. благодаря работам ряда учёных [М. Ниренберг, Х. Корана, Р. Холли (лауреаты Нобелевской премии по физиологии или медицине за работу по расшифровке генетического кода и его функции в синтезе белков)] при участии Ф. Крика были установлены четыре свойства генетического кода:

- ▶ три азотистых основания (триплет) кодируют одну аминокислоту (табл. 5.1);
- ▶ триплеты генетического кода не перекрываются;
- ▶ последовательности триплетов считываются с определённой начальной точки, знаки препинания внутри кодирующей последовательности отсутствуют;
- ▶ одна аминокислота может быть закодирована разными триплетами — вырожденность (избыточность) генетического кода. 61 кодон кодирует 20 аминокислот (табл. 5.2). УАА, УАГ и УГА — нонсенс-кодона (стоп-кодона, «бессмысленные» кодона).

Таблица 5.1. Генетический код. Представлены 1-е, 2-е и 3-е основания кодона. Названия оснований даются в двух вариантах. Без скобок указаны нуклеотиды РНК, в скобках — нуклеотиды ДНК

1-е основание	2-е основание				3-е основание
	У (А)	Ц (Г)	А (Т)	Г (Ц)	
У (А)	Фен Фен Лей Лей	Сер Сер Сер Сер	Тир Тир – –	Цис Цис – Три	У (А) Ц (Г) А (Т) Г (Ц)
Ц (Г)	Лей Лей Лей Лей	Про Про Про Про	Гис Гис Глн Глн	Арг Арг Арг Арг	У (А) Ц (Г) А (Т) Г (Ц)
А (Т)	Иле Иле Иле Мет	Тре Тре Тре Тре	Асн Асн Лиз Лиз	Сер Сер Арг Арг	У (А) Ц (Г) А (Т) Г (Ц)
Г (Ц)	Вал Вал Вал Вал	Ала Ала Ала Ала	Асп Асп Глу Глу	Гли Гли Гли Гли	У (А) Ц (Г) А (Т) Г (Ц)

Таблица 5.2. Аминокислоты и соответствующие им сокращённые названия и кодоны

Однобуквенный код	Сокращённое название		Полное название	Кодоны мРНК
	Латиница	Кириллица		
A	Ala	Ала	Аланин	ГЦУ ГЦЦ ГЦА ГЦГ
C	Cys	Цис	Цистеин	УГУ УГЦ
D	Asp	Асп	Аспарагиновая кислота	ГАУ ГАЦ
E	Glu	Глу	Глутаминовая кислота	ГАГ ГАА
F	Phe	Фен	Фенилаланин	УУУ УУЦ
G	Gly	Гли	Глицин	ГГУ ГГЦ ГГА ГГГ
H	His	Гис	Гистидин	ЦАУ ЦАЦ
I	Ile	Иле	Изолейцин	АУУ АУЦ АУА

Окончание табл. 5.2

Однобуквенный код	Сокращённое название		Полное название	Кодоны мРНК
	Латиница	Кириллица		
K	Lys	Лиз	Лизин	AAA AAG
L	Leu	Лей	Лейцин	УУА УУГ ЦУУ ЦУЦ ЦУА ЦУГ
M	Met	Мет	Метионин	AUG
N	Asn	Асн	Аспарагин	AAU AAC
P	Pro	Про	Пролин	ЦЦУ ЦЦЦ ЦЦА ЦЦГ
Q	Gln	Глн	Глутамин	ЦАА ЦАГ
R	Arg	Арг	Аргинин	ЦГУ ЦГЦ ЦГА ЦГГ АГА АГГ
S	Ser	Сер	Серин	УЦУ УЦЦ УЦА УЦГ АГУ АГЦ
T	Thr	Тре	Треонин	АЦУ АЦЦ АЦА АЦГ
V	Val	Вал	Валин	ГУУ ГУЦ ГУА ГУГ
W	Trp	Три	Триптофан	УГГ
Y	Tyr	Тир	Тирозин	УАУ УАЦ

Классификация генов

Гены по компартментной локализации в клетке разделяются на ядерные и митохондриальные. Среди ядерных генов принято различать белок-кодирующие и белок-некодирующие гены.

Белок-кодирующие гены:

- ▶ гены «домашнего хозяйства» обеспечивают процессы репликации и репарации ДНК, синтеза РНК и белка, синтеза АТФ; другими словами, они контролируют поддержание важнейших жизненных функций организма и экспрессируются

практически во всех тканях на относительно постоянном уровне и на всех стадиях жизненного цикла организма;

- ▶ гены терминальной дифференцировки (кодируют специфические белки, в том числе ферменты, в конкретном клеточном типе и его фенотипах);
- ▶ гены факторов транскрипции, модулирующие активность генов «домашнего хозяйства» и генов клеточной дифференцировки.

Белок-некодирующие гены:

- ▶ обеспечивающие молекулярный механизм синтеза белка (рибосомная РНК, тРНК, малые ядерные РНК);
- ▶ регулирующие процесс синтеза белка (короткие интерферирующие РНК).

Экспрессия гена

Экспрессия гена — реализация наследственной информации, закодированной в гене, в функциональный продукт: РНК или белок. Экспрессия генов включает в себя последовательно следующие этапы: транскрипция, процессирование мРНК (кэпирование, сплайсинг, полиаденилирование), экспорт зрелой мРНК в цитоплазму, трансляция мРНК (синтез белка на рибосомах), деградация мРНК (см. рис 5.1).

Транскрипция (от лат. *transcriptio* — «переписывание») — синтез РНК на матрице ДНК (антисмысловой нити). Таким образом, транскрипт (РНК) является копией смысловой нити ДНК. Транскрипция состоит из стадий инициации, элонгации и терминации.

Инициация транскрипции. ДНК упакована в ядре в составе эухроматина или гетерохроматина. Чтобы белки-активаторы и ферменты транскрипции достигли своих сайтов связывания на ДНК, хроматин должен «раскрыться». Это достигается эпигенетическими процессами, например метилированием/деметилованием ДНК, ацетилизацией/деацетилизацией гистонов (см. Общая генетика, глава 8).

РНК-полимераза II связывается с промотором в точке инициации транскрипции через основные факторы транскрипции (рис. 5.3). TFIID — первый белок, прикрепляющийся к ТАТА-боксу промотора гена и определяющий правильную точку начала транскрипции. TFIIA и TFIIB, РНК-полимераза II и TFIIF, TFIIЕ, TFIIN и TFIIF последовательно связываются с TFIID, обеспечивая связь РНК-полимеразы II с ДНК.

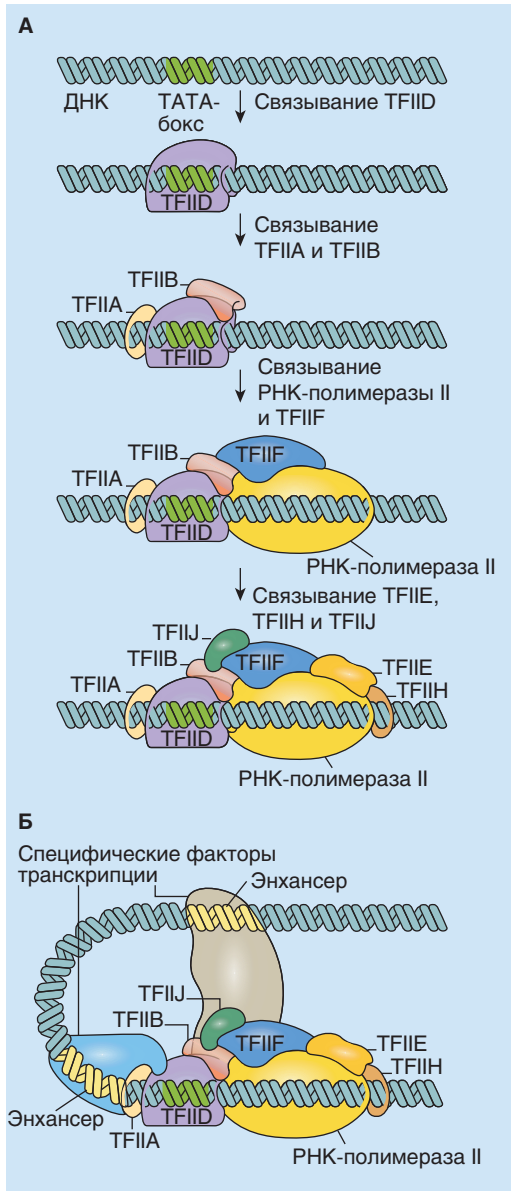


Рис. 5.3. Инициация транскрипции. А — сборка основных факторов транскрипции; Б — активация энхансеров специфическими факторами транскрипции [по: Boron W.F., Boulpaer E.L., 2003]

Специфические факторы транскрипции требуются для активации транскрипционной машины. Регуляторные белки взаимодействуют с цис-элементами энхансеров, расположенных за пределами промотора. Благодаря гибкости молекулы ДНК образуются петли ДНК, позволяющие взаимодействовать энхансеру и промотору с последующей активацией транскрипции гена. К этим факторам относятся рецепторы стероидных и тиреоидных гормонов, ретиноидов, витамина D, CREB (белок, взаимодействующий с цАМФ-чувствительным элементом), c-Myc, AP-1 (c-Jun/c-Fos).

Элонгация, или синтез первичного транскрипта (про-мРНК), осуществляется с помощью РНК-полимеразы II, которая движется по ДНК-матрице в направлении 3' → 5'. Синтез РНК протекает по правилу Уотсона–Крика (комплементарного спаривания нуклеотидов) на матрице ДНК, при этом вместо тимина в молекулу РНК встраивается урацил. Для того чтобы получить копию смысловой цепи ДНК, синтез РНК происходит с антисмысловой цепи. Расположение смысловой и антисмысловой последовательностей для каждого гена между двумя антипараллельными цепями полинуклеотидов в двухцепочечной молекуле ДНК индивидуально.

Терминация. Синтез про-мРНК прекращается в точке терминатии последнего экзона.

Процессинг (процессирование, посттранскрипционная модификация мРНК). Процессирование про-мРНК (созревание мРНК) включает в себя цепь событий: сплайсинг, экпирование 5'-конца РНК, удаление нуклеотидов на 3'-конце, образование полиаденинового хвоста.

Сплайсинг (от англ. *splice* — «соединять концы чего-либо») — процесс созревания РНК, в котором происходит удаление из про-мРНК интронов и объединение экзонов друг с другом.

Малые ядерные РНК (*small nuclear RNA, snRNA*) — класс РНК. Образуют комплексы со специфическими белками (малые ядерные рибонуклеопротеины), обладают ферментативной активностью (рибозимы), вырезают интроны и соединяют вместе экзоны.

Сплайсосома формируется в месте вырезания интрона, состоит из последовательно присоединяющихся малых ядерных РНК (U1, U2...U12) и белков. U1 узнает 5'-место сплайсинга

и 3'-место сплайсинга по специальным фланкирующим последовательностям нуклеотидов в интроне: на 5'-конце — GU-последовательность, на 3'-конце — AG-последовательность.

Альтернативный сплайсинг позволяет получить различные белковые продукты с одного гена (рис. 5.4). Ген кальцитонина может кодировать гормон кальцитонин и относящийся к кальцитониновому гену пептид альфа. В С-клетках щитовидной железы в результате сплайсинга объединяются экзоны 1–4, которые кодируют кальцитонин. В чувствительных нейронах процессируется мРНК, в которой отсутствует экзон 4, но прибавляются экзоны 5 и 6. В результате вместо кальцитонина клетки синтезируют относящийся к кальцитониновому гену пептид альфа.

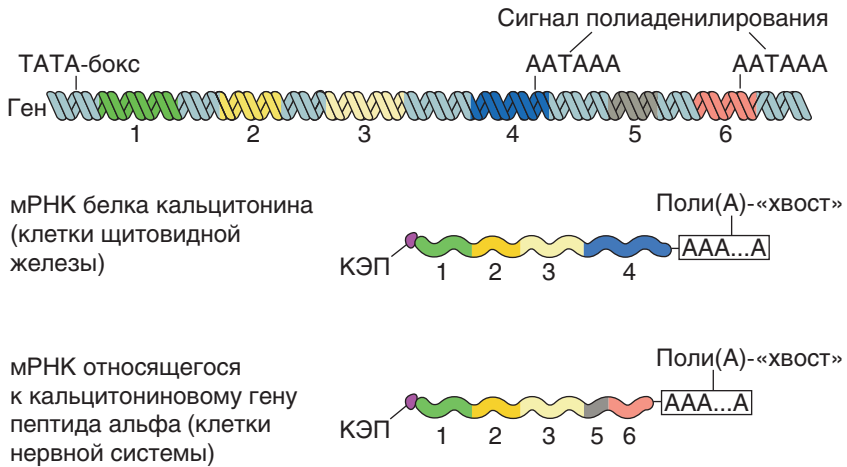


Рис. 5.4. Альтернативный сплайсинг. Экзоны, выделенные разным цветом, обозначены цифрами 1–6. Интроны между экзонами выделены голубым цветом

Создание КЭП-сайта — присоединение к 5'-концу про-мРНК 7-метилгуанозина. Кэпирование защищает транскрипт от деградации экзонуклеазами, способствует экспорту мРНК из ядра в цитоплазму и обеспечивает связывание мРНК с рибосомой.

Полиаденилирование. На первом этапе происходит расщепление: с помощью экзонуклеазы удаляются 20 нуклеотидов на 3'-конце про-мРНК до сайта инициации полиаденилирования

(5'-AAUAAA-3'). На втором этапе с помощью поли(А)полимеразы к 3'-концу присоединяются адениновые основания (100–200 оснований), образуется полиадениновый «хвост», защищающий мРНК от деградации экзонуклеазами.

Зрелая мРНК состоит из экзонов, содержит 5'- и 3'-нетранслируемые концы. 5'-нетранслируемый конец кэпирован, 3'-нетранслируемый конец защищён полиадениновым «хвостом». Экспорт зрелой мРНК в цитоплазму происходит через комплекс ядерной поры с участием экспортинов и гуанозинтрифосфатазы Рап.

Трансляция (от лат. *translatio* — «перенос», «перемещение»; «перевод» — на другой язык) — преобразование последовательности нуклеотидов мРНК в последовательность аминокислот (синтез белка) на рибосомах с помощью тРНК.

Транспортная РНК (тРНК) несёт второй генетический код (антикодон). тРНК содержит 70–90 нуклеотидов, формирующих вторичную структуру — «клеверный лист» (рис. 5.5). тРНК доставляет аминокислоты к рибосоме, где они присоединяются к растущей полипептидной цепи. 3'-конец тРНК (акцептор) содержит сайт для прикрепления аминокислоты. Аминоацил-тРНК-синтетаза присоединяет специфичную аминокислоту к молекуле тРНК с образованием аминоацил-тРНК. Для каждой аминокислоты существует своя аминоацил-тРНК-синтетаза. Специфическое

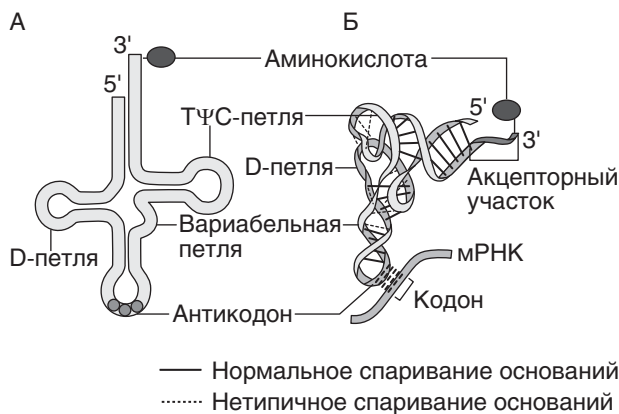


Рис. 5.5. Строение транспортной РНК. А — структурная модель тРНК; Б — конформационная модель тРНК [из: Колман Я., Рем К.-Г., 2000]

связывание тРНК с мРНК обеспечивает участок тРНК, который содержит антикодон из трёх нуклеотидов, соответствующий кодону мРНК и спаривающийся с ним в А-участке рибосомы. Избыточность генетического кода предполагает существование для одной аминокислоты нескольких тРНК. Поэтому тРНК несут 48 разных антикодонов.

Деградация мРНК. Регуляция экспрессии генов на транскрипционном и посттранскрипционном уровнях происходит с помощью механизма РНК-интерференции. РНК-интерференция — подавление экспрессии генов (деградация мРНК), индуцированное короткими интерферирующими РНК. К коротким интерферирующим РНК относятся малая интерферирующая РНК (*siRNA* — *small interfering RNA*) и микроРНК (*miRNA* — *microRNA*).

Малая интерферирующая РНК (siRNA). РНК-интерференция с помощью siRNA является защитным механизмом против РНК вирусов и мобильных генетических элементов (транспозонов).

Механизм посттранскрипционного ингибирования экспрессии гена на уровне блокады трансляции белка-мишени путём деградации его мРНК начинается с транскрипции длинной двухцепочечной РНК (обычно >200 п.н.) вирусного или эндогенного (транспозоны) происхождения (рис. 5.6). Эндогенная рибонуклеаза III (*Dicer*) процессирует длинную двухцепочечную РНК с образованием siRNA длиной в 19–23 п.н. Далее двухцепочечная siRNA попадает в протеиновый комплекс RISC (*Ribonucleic Acid Induced Silencing Complex*). RISC «раскручивает» двойную спираль siRNA и освобождается от смысловой нити siRNA. Антисмысловая нить в составе белкового комплекса прикрепляется к комплементарному участку мРНК-мишени, и RISC разрезает мРНК. Клеточные нуклеазы заканчивают деградацию мРНК, прекращая синтез белка. Реакция RISC — каталитическая, она повторяется несколько раз, обеспечивая очень эффективное и долговременное ингибирование экспрессии определённого гена. Кроме того, siRNA может функционировать как праймер для синтеза дополнительных длинных двухцепочечных молекул РНК, усиливая эффект siRNA, участвующей в ингибировании экспрессии гена.