

Федеральное агентство по образованию
Нижегородский государственный университет им. Н.И. Лобачевского

Национальный проект «Образование»
Инновационная образовательная программа ННГУ. Образовательно-научный центр
«Информационно-телекоммуникационные системы: физические основы и
математическое обеспечение»

В.В. Новиков

Хранение и реализация генетической информации в клетке

*Учебно-методические материалы по программе повышения
квалификации «Хранение и обработка информации в
биологических системах»*

Нижегород
2007

Учебно-методические материалы подготовлены в рамках инновационной образовательной программы ННГУ: Образовательно-научный центр «Информационно-телекоммуникационные системы: физические основы и математическое обеспечение»

Новиков В.В. Хранение и реализация генетической информации в клетке. Учебно-методический материал по программе повышения квалификации «Хранение и обработка информации в биологических системах». Нижний Новгород, 2007, 81 с.

В пособии представлены полученные в последние 10 лет сведения о строении и функционировании генома эукариот и прокариот. Приведены данные о строении генома человека.

Пособие предназначено для повышения квалификации преподавателей биологических факультетов классических университетов, научных сотрудников, студентов, магистров и аспирантов, изучающих молекулярную биологию и биоинформатику.

ОГЛАВЛЕНИЕ

Введение	4
1. Хранение генетической информации в клетке	4
1.1. Анатомия эукариотического генома	4
1.1.1. Ядерный геном эукариот	5
1.1.2. Геномы органелл	13
1.2. Анатомия прокариотического генома	15
1.3. Повторяющиеся последовательности геномной ДНК	23
1.4. Геном человека	28
2. ЭКСПРЕССИЯ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ	38
2.1. Этапы процессинга генетической информации	38
2.2. РНК, содержащаяся в клетке	40
2.3. Протеом клетки	45
2.4. Механизмы экспрессии генома	48
2.4.1. Обеспечение доступности генома	48
2.4.2. Инициация транскрипции	58
2.4.3. Дегградация матричной РНК	64
2.4.4. Инициация трансляции	67
2.4.5. Процессинг, фолдинг и дегградация белков	70
Литература	76

ВВЕДЕНИЕ

Три большие группы организмов (эукариоты, бактерии и археи) в настоящее время относят к трем доменам жизни. Каждый из доменов имеет свои особенности строения генома. Хотя геном архей исследован сравнительно меньше геномов эукариот и бактерий, тем не менее, уже давно ясно, что это отдельная ветвь жизни, отдельный домен, занимающий свое положение наряду с эукариотами и бактериями. Тем не менее, существует деление всего живого на эукариоты и прокариоты. У эукариот клетки содержат структуры, окруженные мембранами, включая ядро и органеллы (митохондрии, хлоропласты и др.). К эукариотам относят животных, растения, грибы и простейших. Прокариоты не содержат окруженных мембраной компартментов. Существует две сильно различающиеся группы прокариот, отличающиеся друг от друга генетически и биохимически: а) бактерии, которые включают основное число известных в настоящее время прокариот, например грам-отрицательные бактерии (*E. coli*), грам-положительные бактерии (*Bacillus subtilis*), цианобактерии (*Anabaena*); б) археи, которые встречаются реже бактерий и нередко существуют в экстремальных условиях, таких, например, как горячие источники, дно анаэробных озер.

В настоящем пособии автор постарался представить полученные в последние 5-10 лет данные о строении геномов и механизмах информационного процессинга, то есть о способах реализации генетической информации, особенностях экспрессии геномов. Основной акцент сделан на особенности строения и на недавно выявленные механизмы работы генома человека.

1. ХРАНЕНИЕ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ В КЛЕТКЕ

1.1. Анатомия эукариотического генома

Геном – совокупность генов и других нуклеотидных последовательностей, находящихся в клетке. Другими словами – это вся ДНК, присутствующая в клетке. Гены – единицы наследственной информации, занимающие определенные положения в геноме и контролирующие выполнение определенных функций в организме. В составе гена как единицы считывания (транскрипционной единицы) могут присутствовать так называемые кодирующие и некодирующие последовательности. Первые у эукариот и архей называют экзонами. Вторые – интронами. Экзоны кодируют белки и некоторые виды РНК (рибосомальная, транспортная). Интроны разделяют экзоны. Их принято считать некодирующими участками геномной последовательности. Однако исследования

последних лет показали, что в интронах заложена информация о строении микроРНК – небольших молекул РНК, обладающих регуляторными свойствами (Young-Kook Kim, Narry Kim, 2007). Наряду с генами в геномах присутствуют псевдогены и фрагменты генов, повторяющиеся последовательности и так называемая «бесмысленная» ДНК. У разных организмов приведенные компоненты генома представлены в разной степени.

Геном эукариот состоит из ядерного генома и генома органелл. Основная часть генетической информации заключена в ядерном геноме, содержащемся в хромосомах ядра клеток, и намного меньшая часть локализована в митохондриях и хлоропластах (в случае фотосинтезирующих организмов). Начнем рассмотрение геномов с ядерного генома.

1.1.1. Ядерные геномы эукариот

Все эукариотические ядерные геномы, которые в настоящее время изучены, разделены на несколько линейных молекул ДНК, содержащихся в разных хромосомах. Хотя основные физические характеристики эукариотических геномов сходны, одна важная черта значительно различается у разных организмов. Это размеры геномов. Наименьшие эукариотические геномы имеют длину менее 10 мегабаз, наибольшие – более 100 000 мегабаз. Одна мегабаза – это один миллион пар нуклеотидных оснований.

В самом общем приближении размеры геномов отражают сложность организма. Наиболее простые эукариоты, такие как грибы, имеют наименьшие геномы. Высшие эукариоты, такие как позвоночные и цветковые растения, имеют наибольшие геномы. Однако отсутствует корреляция между размерами геномов и количеством генов в его составе.

Гены менее сложных организмов гены упакованы теснее. У таких организмов нуклеотидные последовательности, разделяющие гены, меньше, чем у более сложных организмов. Различия в степени упаковки геномов хорошо видны при сравнении фрагмента третьей хромосомы дрожжей *S. cerevisiae*, расшифрованной в 1996 году, и фрагмента генома человека той же длины (Рис. 1).

Фрагмент третьей хромосомы дрожжевого генома обладает следующими чертами:

- Он содержит больше генов, чем человеческий сегмент. Этот регион третьей хромосомы дрожжей содержит 26 генов, кодирующих белки, и два гена, которые кодируют транспортные РНК – короткие не кодирующие белок молекулы РНК, участвующие в считывании генетического кода в процессе синтеза белка.

- Относительно небольшое количество дрожжевых генов прерывисто. В данном сегменте третьей хромосомы нет прерывистых генов. Суммарно в дрожжевом геноме, имеющем 5800 генов, обнаружено только 239 интронов, то есть участков некодирующей

ДНК, разделяющих кодирующие участки. Заметим, что в человеческом геноме обнаружено более 300 000 интронов.

- Имеется небольшое число геномных повторов. Рассматриваемая часть хромосомы дрожжей содержит один длинный концевой повторяющийся элемент (LTR элемент), называемый *Tu-2*, и четыре неполных LTR элемента, названных дельта-последовательностями. Эти пять геномных повторяющихся последовательностей занимают 13,5 % протяженности рассматриваемого сегмента третьей хромосомы дрожжей. Однако, такая ситуация не типична для дрожжевого генома в целом. После того, как была расшифрована нуклеотидная последовательность всех 16 хромосом дрожжей, подсчитали, что в расчете на весь геном дрожжей повторяющиеся последовательности составляют 3,4 %. У человека повторяющиеся последовательности занимают 44 % генома.

Таким образом, генетическая организация дрожжей более экономна, чем организация генома человека. Гены сами по себе более компактны, имеют очень мало интронов, расстояния между ними относительно короткие, меньше повторяющихся последовательностей и некодирующих бессмысленных последовательностей.

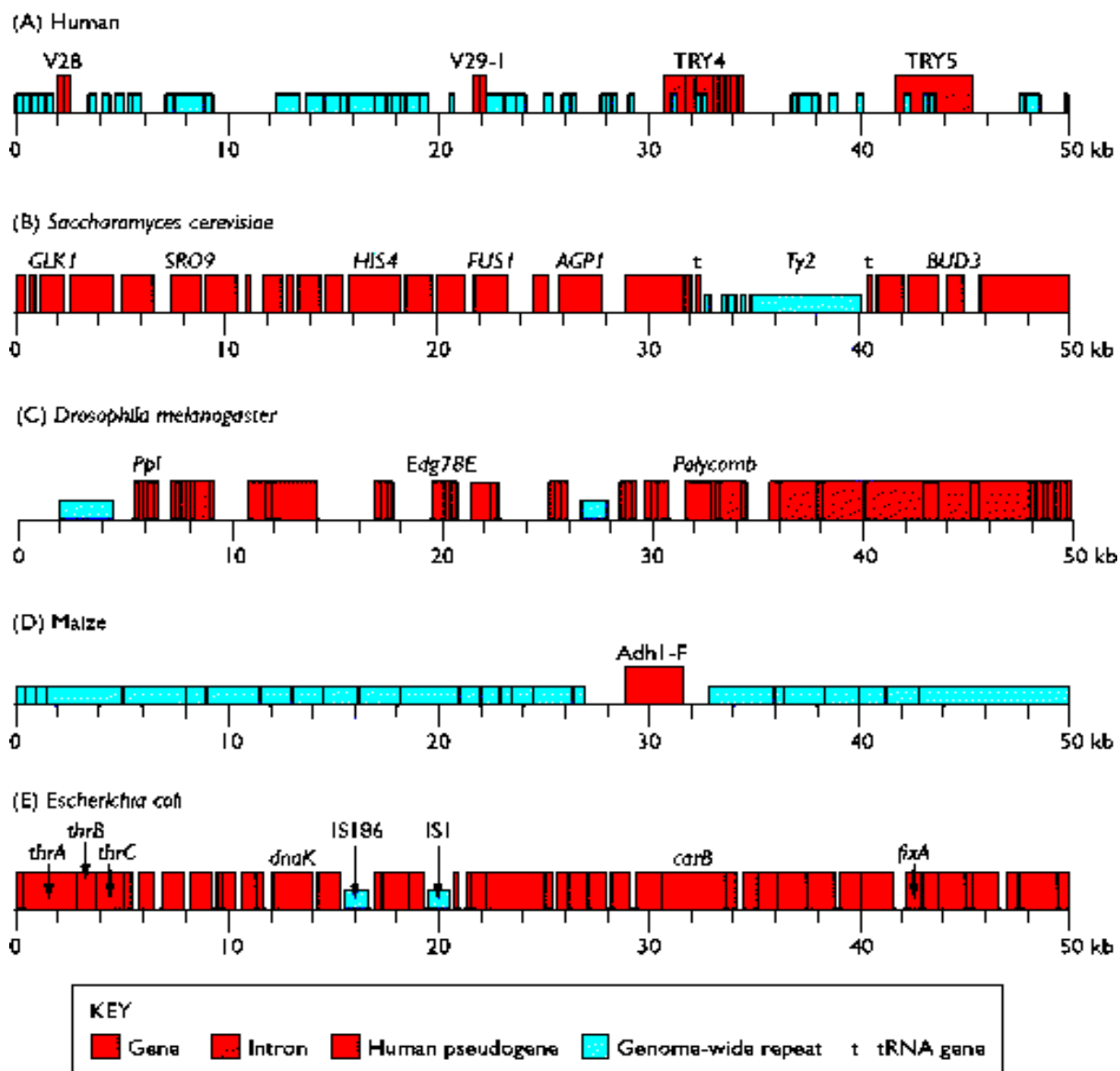


Рис. 1. Сравнительная картина фрагментов геномов человека (А), дрожжей (В), дрозофиллы (С), кукурузы (D) и кишечной палочки (Е), длиной 50 килобаз (Т.А.Браун, 2002).

Дополнительной иллюстрацией того, что более сложные организмы имеют менее компактные геномы, является сравнительный анализ фрагмента генома дрозофиллы, имеющего длину 50 килобаз, с фрагментами геномов той же длины у дрожжей и человека. Как видно из рис. 1, строение генома дрозофиллы занимает промежуточное положение между геномами человека и дрожжей. Во фрагменте длиной 50 килобаз имеется 11 генов. Это больше, чем во фрагменте генома человека, но меньше, чем во фрагменте генома дрожжей. Семь из одиннадцати генов имеют интроны. Представленная картина характерна и для всего генома дрозофиллы в целом. Плотность генов у дрозофилы

занимает промежуточное положение между дрожжами и человеком. Гены дрозофиллы имеют намного большее количество интронов, чем у генов дрожжей, но почти в три раза меньше, чем у генов человека. Повторяющиеся последовательности составляют 12 %.

В настоящее время ясно, что количество геномных повторов играет основную роль в определении степени компактности генома. Это можно проиллюстрировать на примере генома кукурузы, который имеет размеры равные 5000 мегабаз, что значительно больше человеческого генома, но относительно мало в сравнении с геномами других цветковых растений. Секвенирование генома кукурузы показало, что основную его часть составляют повторяющиеся элементы.

Как видно из рис. 1, выбранный для примера регион генома кукурузы размером 50 килобаз содержит всего один ген, кодирующий алкогольдегидрогеназу, и еще один ген с неизвестной функцией. Доминирующими последовательностями в представленном регионе являются не последовательности, кодирующие белки, а геномные повторы. Большинство из них принадлежит к LTR элементам, которые составляют почти все некодирующие участки представленного фрагмента и занимают по произведенным оценкам приблизительно 50 % от генома кукурузы. Подобным образом за счет большого количества геномных повторов может быть увеличен и размер геномов других организмов. Так, геном *Amoeba dubia* в соответствии с геномами других простейших должен иметь размеры 100-500 килобаз, но имеет размеры 200 000 мегабаз.

Ядерный геном расщеплен на несколько линейных молекул ДНК, содержащихся в отдельных хромосомах. Исключения из этого правила не известны. Все эукариоты, которые к настоящему времени изучены, имеют, по крайней мере, две хромосомы, а молекула ДНК всегда линейна. Вариабельность генома на этом уровне заключается в разном количестве хромосом у разных видов организмов. Так, геном дрожжей построен из 16 хромосом, а геном дрозофилы из 4 хромосом. Однако никакой связи между количеством хромосом и размером генома не обнаружено. Геномы некоторых саламандр в 30 раз больше человеческого, но построены из вдвое меньшего количества хромосом. Эти сопоставления интересны, но в настоящее время ничего нельзя сказать о смысловой нагрузке, которую несет то или иное количество хромосом.

Хромосомы намного короче, чем входящие в их состав молекулы ДНК. Средняя человеческая хромосома содержит молекулу ДНК длиной около 5 см. Для размещения столь длинной молекулы ДНК в хромосоме нужна сложно организованная система упаковки. Способ упаковки влияет на экспрессию отдельных генов, а значит, на механизмы реализации генетической информации.

Важный прорыв в понимании принципов упаковки ДНК был осуществлен еще в начале 70-х годов прошлого века при использовании комбинации биохимических методов и электронной микроскопии. Тогда уже было известно, что ДНК ассоциирована с ДНК-связывающими белками, называемыми гистонами, но истинная природа ассоциации была не понятна. В 1973-74 годах несколько групп исследователей осуществили эксперименты, в которых обрабатывали комплексы ДНК и гистонов нуклеазами. Последние режут ДНК в позициях, которые не защищены белками в составе ДНК-гистоновых комплексов. После ограниченной обработки очищенного хроматина нуклеазами основная масса получившихся фрагментов ДНК имела длину примерно 200 пар нуклеотидов или была кратна им, что предполагает регулярное расположение гистоновых белков вдоль молекулы ДНК.

Данные биохимиков были подтверждены электронно-микроскопическими наблюдениями очищенного хроматина. На нити ДНК были обнаружены белковые образования в виде зерен. Дальнейший биохимический анализ показал, что каждое зерно, или нуклеосома, содержит восемь молекул гистонов (по две молекулы гистонов H2A, H2B, H3, H4), которые образуют так называемый коровый октамер. Снаружи его дважды огибает нить ДНК. Стало понятно, что 140 или 150 пар нуклеотидов ассоциировано с нуклеосомной частицей. Нуклеосомы разделены участком связующей, линкерной ДНК длиной 50-70 пар оснований. В сумме получается 190-220 пар оснований на одну нуклеосому.

Кроме коровых гистонов в ядре имеется группа дополнительных гистонов, тесно связанных друг с другом и вместе названных линкерными гистонами. У позвоночных они включают гистоны H1a-e, H1^o, H1t и H5. Отдельные линкерные гистоны связываются с каждой нуклеосомой с образованием хроматосомы, но истинное расположение линкерных белков пока не известно. Структурные исследования подтверждают традиционную модель, в соответствии с которой линкерные гистоны действуют как защелка, предотвращающая отсоединение свернутой ДНК от нуклеосомы. Однако, результаты других исследований свидетельствуют о том, что, по крайней мере, у некоторых организмов, связующие гистоны не локализуются на поверхности нуклеосом, как это предполагается в соответствии с первой моделью, а расположены между коровым октамером и ДНК (Pruss 1995).

Выявлена конденсированная форма нуклеопротеидного комплекса, названная 30-нанометровыми нитями (фибриллами), шириной примерно 30 нм. Истинный механизм, с помощью которого нуклеосомы ассоциируют в нити, неизвестен, но предложено несколько моделей. Наиболее популярна соленоидная модель. Отдельные нуклеосомы в

составе нитей могут соединяться вместе в единую структуру с помощью линкерных гистонов или коровых гистонов, чьи хвосты экспонированы из нуклеосом. Модель подтверждается тем, что биохимическая модификация хвостов приводит к разворачиванию 30-нанометровых нитей и появлению способности генов к активации.

30-нанометровые нити составляют основной тип хроматина в ядре в интерфазе, то есть в период между делениями ядер клеток. При делении ядра ДНК принимает более компактную форму упаковки, что приводит к образованию высоко конденсированных метафазных хромосом, видимых в световой микроскоп.

Метафазные хромосомы образуются на стадии клеточного цикла после репликации ДНК, когда каждая хромосома несет по две копии ДНК. Две копии ДНК удерживаются вместе в районе центромеры, которая находится в специфической позиции у каждой хромосомы. Имеется множество различных методов окраски хромосом. С их помощью выявлено характерное распределение полос, индивидуальное для каждой хромосомы. Набор хромосом называют кариограммой, его можно рассматривать как характерную черту организма того или иного вида.

ДНК, расположенная в районе центромер, и присоединенные к ней белки также имеют специальные характеристики. Нуклеотидная последовательность центромерного участка ДНК очень хорошо изучена у арабидопсис (*Arabidopsis thaliana*). Центромеры арабидопсис имеют протяженность 0,9-1,2 мегабаз и все они построены из повторяющихся последовательностей размером 180 пар оснований. У человека эквивалентные последовательности имеют протяженность, равную 171 паре оснований. Они названы альфоидной ДНК. Наряду с повторяющимися последовательностями у арабидопсис в этом районе имеется также небольшое количество генов. Их плотность составляет 7-9 генов на 100 килобаз в сравнении с 25 генами на 100 килобаз в нецентромерных последовательностях хромосом арабидопсис. Обнаружение того факта, что центромерная ДНК содержит гены, было большим сюрпризом, поскольку считалось, что этот регион генетически не активен.

Известно, по крайней мере, 7 специальных центромерных белков. Один из них, CENP-A, похож на гистон H3 и даже способен заменять его в центромерных нуклеосомах. Предполагается, что специфические центромерные белки имеют отношение к функциям этого региона хромосом и участвуют в прикреплении веретена к метафазным хромосомам. Часть кинетохора построена из альфоидной ДНК, CENP-A и других белков, но их структура пока не описана в деталях.

Второй важной частью хромосомы является ее терминальный регион или теломера. Теломеры маркируют концы хромосом и тем самым дают возможность репарационным

механизмам клетки отличать разрывы хромосом от натуральных концов хромосом. У человека теломерная ДНК построена из сотен повторяющихся мотивов 5'-TTAGGG-3', с которыми связываются два специальных белка TRF-1 и TRF-2, участвующих в регуляции длины теломеры. Еще один теломерный белок образует связь между теломерой и периферией ядра и определяет место локализации конца хромосомы в ядре клетки. Несколько других белков регулируют ферментативную активность, поддерживающую длину каждой теломеры в процессе репликации ДНК. Эта активность является критически важной для сохранения хромосомы.

В геноме большинства организмов гены располагаются неупорядоченно. Существуют вариации в плотности генов на разных участках хромосом. Как уже упоминалось, средняя плотность расположения генов у арабидопсис составляет 25 генов на 100 тысяч пар нуклеотидов. Разброс частоты встречаемости генов в разных участках генома этого растения составляет от 1 до 38 генов на 100 тысяч пар оснований. Плотность генов в геноме человека варьирует от 0 до 64 генов на сто тысяч пар нуклеотидов.

Важной общей чертой всех геномов является наличие мультигенных семейств – групп генов, имеющих идентичную или близкую последовательность. Например, для всех эукариот, у которых определена нуклеотидная последовательность ДНК, обнаружено множество копий генов, кодирующих рибосомальные РНК. В частности, человеческий геном содержит приблизительно 2000 генов 5S РНК, локализованных в едином кластере первой хромосомы. Имеется также 280 копий повторяющихся единиц, содержащих гены 28S, 5,8S, 18S рибосомальных РНК. Предполагается, что такое большое число копий нужно клетке в период деления, для обеспечения быстрого синтеза белка в этот период.

Гены транспортных РНК также являются примерами классических, или простых, мультигенных семейств, в которых все члены семейства идентичны или имеют почти идентичные последовательности. Такие семейства возникают вследствие дупликаций генов и сохраняют исходную нуклеотидную последовательность на протяжении значительных отрезков времени в процессе эволюции.

Другие мультигенные семейства, более характерные для высших эукариот, называются сложными. Индивидуальные члены таких семейств имеют сходные, но, тем не менее, значительно различающиеся последовательности. Продукты таких генов различаются. Один из самых известных примеров сложного мультигенного семейства – это глобиновые гены млекопитающих. Глобины являются белками крови. Каждая молекула гемоглобина состоит из двух глобинов альфа-типа и двух глобинов бета-типа. У человека глобины альфа-типа кодируются небольшим мультигенным семейством,

расположенным на 16 хромосоме, а глобины бета-типа – вторым семейством, локализованным на хромосоме 11.

Гены глобинов были среди первых, для которых в конце 70-х годов была определена нуклеотидная последовательность. Результаты секвенирования показали, что в каждом семействе гены похожи друг на друга, но не идентичны. Нуклеотидные последовательности наиболее отличающихся друг от друга генов внутри семейства бета-глобинов (бета и эpsilon глобины) идентичны только на 79,1%. Этого сходства было достаточно, чтобы отнести оба белка к одному семейству. Подобные различия обнаружены и в генах альфа-кластера.

Почему члены семейства генов глобинов отличаются друг от друга? Ответ на этот вопрос был получен при исследовании экспрессии отдельных генов на разных стадиях развития организма человека. Так, гены эpsilon, G_γ и A_γ , относящиеся к бета-кластеру, экспрессируются на ранних стадиях эмбрионального развития. Сигма и бета гены экспрессируются во взрослом организме. Все эти гены в разной степени отличаются друг от друга. Белковые продукты этих генов выполняют различающиеся функции. Различия в биохимических свойствах глобиновых белков отражают сравнительно небольшие, но существенные изменения в физиологической роли, которую они играют на разных стадиях развития человека.

В некоторых мультигенных семействах индивидуальные представители кластеризованы, то есть, расположены сравнительно компактно, в одном участке хромосомы. В других семействах гены могут быть разбросаны по всему геному. Пример такого диспергированного семейства – пять генов альдолазы, фермента энергетического обмена. Они локализованы на 3, 9, 10, 16 и 17 хромосомах. Важно, что даже если гены диспергированы, то все равно представители этого семейства имеют сходство в своих последовательностях, что отражает общность эволюционного происхождения.

При сравнении этих последовательностей можно проанализировать взаимоотношения не только внутри генов данного семейства, но также и между разными семействами. Например, гены семейств альфа и бета глобинов имеют некоторое сходство в последовательностях, исходя из чего, было сделано заключение о существовании у них единого эволюционного предшественника. То есть эти два семейства составляют суперсемейство, а степень сходства между отдельными представителями этого суперсемейства отражает их эволюционную близость.

1.1.2. Геномы органелл эукариот

Первые указания на присутствие экстрахромосомных генов были получены еще в 50-х годах прошлого века при исследовании необычного характера наследования некоторых признаков у *Neurospora crassa*, *S. cerevisiae* и фотосинтетической водоросли *Chlamidomonas reinhardtii*. Электронная микроскопия в сочетании с биохимическими методами позволили сделать заключение о существовании митохондриального генома и генома хлоропластов, которые независимы и отличаются от ядерного генома.

Почти все эукариоты имеют митохондриальные геномы. Все фотосинтезирующие эукариоты имеют геномы хлоропластов. Изначально считалось, что геномы всех органелл представляют собой кольцевые молекулы ДНК. Электронная микроскопия показала, что в некоторых органеллах присутствуют как кольцевые, так и линейные молекулы ДНК. Было предположено, что линейные молекулы просто являются фрагментами кольцевых геномов и получены в результате их частичного разрушения. В настоящее время известно, что кольцевые геномы нередко сосуществуют с линейными геномами, например в хлоропластах. Кроме того, обнаружено, что геном органелл может быть фрагментирован. Так, геном хлоропластов одного из представителей морских динофлагеллят расщеплен на множество кольцевых геномов, каждый из которых несет один ген. Кроме того, известно, что у целого ряда низших эукариот геном митохондрий разбит на множество линейных фрагментов (*Paramecium*, *Chlamidomonas* и некоторые дрожжи).

Множественность копий геномов органелл не всегда понятна. Каждая митохондрия человека содержит около 10 копий идентичных молекул. Это означает, что в клетке присутствует почти 8000 копий таких молекул. В клетках *S. cerevisiae* общее количество митохондриальных геномов вероятно меньше, даже несмотря на то, что у них присутствует более 100 геномов на митохондрию. Фотосинтезирующие организмы, такие как *Chlamidomonas*, имеют приблизительно 1000 хлоропластных геномов на клетку, которые включают всего лишь одну пятую часть от количества хлоропластных генов у высших растений.

Размеры митохондриальных геномов очень переменчивы у разных видов эукариот и не соотносятся со сложностью организма. Большинство многоклеточных животных имеют небольшие митохондриальные геномы с компактной генетической организацией. Гены тесно прилегают друг к другу, расстояния между ними невелики. Человеческий митохондриальный геном имеет размер, равный 16569 пар оснований. Он типичен для митохондриального генома позвоночных. Низшие эукариоты, такие как *S. cerevisiae*, так

же как и цветковые растения, имеют намного большие и менее компактные геномы митохондрий с множеством генов, содержащих интроны.

Геномы хлоропластов менее вариабельны по размерам и большинство имеют структуру, подобную структуре генома хлоропластов риса.

Размеры геномов органелл намного меньше размеров ядерных геномов. Репертуар составляющих их генов также менее разнообразен. В митохондриальных геномах, имеющих высокую вариабельность от вида к виду, количество генов варьирует от пяти у малярийного плазмодия до 92 у простейшего *Reclinomonas americana*. Однако имеются и общие для всех митохондриальных геномов характеристики. Все митохондриальные геномы содержат гены некодирующих рибосомальных РНК и гены некоторых компонентов дыхательной цепи, то есть гены, необходимые для выполнения основной функции митохондрий. Более крупные митохондриальные геномы кодируют также транспортные РНК, рибосомальные белки, белки, обеспечивающие транскрипцию, трансляцию и транспорт других белков в митохондрию из цитоплазмы.

Большинство геномов хлоропластов несут набор примерно из 200 генов или сходное число генов. Они кодируют рибосомальные РНК, транспортные РНК, рибосомальные белки, и белки, вовлеченные в фотосинтез. Общей чертой геномов органелл является наличие генов, кодирующих белки этих органелл. Однако это не все белки, которые им необходимы. Недостающие белки кодируются ядерным геномом, синтезируются в цитоплазме и транспортируются из цитоплазмы в органеллы.

Почему ядерные геномы не кодируют все белки, необходимые органеллам? Твердого ответа на этот вопрос пока нет. Предполагается, что, по крайней мере, сверхгидрофобные белки, кодируются органеллами, поскольку не могут транспортироваться через мембрану, которая окружает митохондрию или хлоропласт. В этой ситуации остается только один путь – синтезировать белок внутри органеллы.

Открытие геномов у митохондрий и хлоропластов вызвало множество спекуляций об их происхождении. В настоящее время большинство биологов согласны с эндосимбиотической теорией, предложенной еще в 60-х годах прошлого века. Эндосимбиотическая теория основывается на том, что экспрессия генов в органеллах напоминает экспрессию генов в бактериях. Кроме этого, анализ нуклеотидных последовательностей генов органелл обнаружил, что они более сходны с бактериальными, чем с эукариотическими. Эндосимбиотическая теория предполагает, что митохондрии и хлоропласты являются реликтами свободно-живущих бактерий, которые образовали симбиотические ассоциации с предшественниками эукариотических клеток.

Подтверждением эндосимбиотической теории является открытие организмов, для которых обнаружены менее продвинутые стадии симбиоза, чем симбиоз между митохондриями (хлоропластами) и клеткой. Демонстрацией ранних стадий симбиоза могут быть простейшие, такие, например, как *Ceanoïphora paradoxa*. Их фотосинтезирующие структуры, названные цианеллами, отличаются от хлоропластов и напоминают непереваренные цианобактерии. Сходным образом, риккетсий, живущих внутри эукариотических клеток, можно рассматривать в качестве современной версии бактерий, которые дали начало митохондриям. Также предполагается, что гидрогеносомы трихомонад, некоторые из которых имеют геном, являются одним из вариантов эндосимбиотических митохондрий.

Если митохондрии и хлоропласты когда-то были свободно живущими бактериями, то эндосимбиоз должен был привести к переносу генов из органелл в ядро. Каким образом происходил такой перенос, остается неясным. Неизвестно, осуществлялся ли быстрый и массовый перенос генов из бактерии (органеллы) в ядро или шел постепенный процесс миграции генов. Однако в недавних исследованиях продемонстрирована возможность переноса наследственной информации между органеллами и ядерным геномом, а также между органеллами клетки. Так, обнаружено, что у некоторых растений геномы хлоропластов содержат сегменты ДНК митохондрий. Геном митохондрий *Arabidopsis* содержит различные сегменты ядерной ДНК. Ядерный геном этого растения включает несколько коротких сегментов геномов хлоропластов и митохондрий, в частности, митохондриальную последовательность размером 270 килобаз, локализованную внутри центромерного региона хромосомы 2. Также документально продемонстрирован перенос митохондриальной ДНК в ядерный геном позвоночных.

1.2. Анатомия прокариотического генома

Геном прокариот существенно отличается от генома эукариот. Он обычно значительно меньше как по размеру, так и по количеству генов. Лишь в отдельных случаях между ними имеются некоторое сходство в размерах геномов и количестве генов. Оно может наблюдаться при сравнении геномов прокариот с геномами низших прокариот. Например, геном кишечной палочки *E. coli K12* имеет размеры, равные 4639 килобаз и содержит только 4405 генов. Напомним, что геном *S. cerevisiae* содержит 5800 генов, однако его размеры значительно больше (12,6 мегабаз). В то же время *V. megaterium* имеет геном, протяженностью 30 мегабаз.

Физическая организация геномов также различается у эукариот и прокариот. Традиционный взгляд на прокариотический геном заключается в том, что он представляет собой единственную кольцевую молекулу ДНК, то есть единичную хромосому. Однако, в дополнение к ней имеются еще кольцевые молекулы ДНК, называемые плазмидами. Гены плазмид полезны микроорганизмам, поскольку кодируют такие свойства, как устойчивость к антибиотикам, способность утилизировать сложные соединения из внешней среды и так далее. Плазмиды не обязательны для жизни микроорганизмов, бактерии могут существовать и без них.

Традиционные взгляды на строение генома прокариот основаны на представлениях, полученных при изучении генома кишечной палочки как типичного прокариота. На самом деле прокариоты демонстрируют значительные различия в организации геномов. Некоторые имеют унипартитные геномы, как например кишечная палочка. У других строение генома является более сложным. Например, *Borrelia burgdorferi* В31 имеет линейную хромосому размером 911 килобаз, содержащую 853 гена. Она сопровождается 17 или 18 линейными и кольцевыми молекулами, размер которых суммарно составляет 533 килобазы. Они кодируют 430 генов. Мультипартитный геном известен и для многих других бактерий и архей.

Несмотря на разнообразие в строении геномов прокариот, кишечную палочку все же рассматривают в качестве типичного представителя прокариот. Типичным является наблюдаемая у кишечной палочки компактность прокариотического генома. На рисунке 1, где представлен фрагмент генома кишечной палочки размером 50 килобаз, сразу бросается в глаза, что в составе этого фрагмента имеется намного больше генов, чем у эукариот, а межгенные пространства очень невелики.

Представленный фрагмент генома содержит 43 гена, занимающих 85,9% пространства этого сегмента. Некоторые гены не имеют выраженных границ между собой. Так, гены *thrA* и *thrB* разделены всего одним нуклеотидом. Ген *thrC* начинается с нуклеотида, немедленно следующим за последним нуклеотидом гена *thrB*. Эти три гена являются примером оперона, то есть группы генов, вовлеченных в один сигнальный биохимический путь (в данном случае – в синтез треонина). Они экспрессируются координировано по отношению друг к другу. В настоящее время опероны используются как модельные системы для изучения принципов регуляции генов.

В целом, гены прокариот короче, чем гены эукариот. Средний размер бактериального гена составляет примерно две трети от эукариотического гена, даже если не учитывать интроны. В то же время бактериальные гены длиннее, чем гены архей.

Две другие особенности прокариотического генома становятся понятны при дополнительном анализе рис. 1. Во-первых, в геноме кишечной палочки отсутствуют интроны. Фактически в ее геноме нет прерывистых генов и это является типичной чертой для прокариот, но не архей. Второй отличительной чертой является низкая частота повторяющихся последовательностей. Большинство прокариотических генов не имеет каких-либо эквивалентов высококопийных повторяющихся последовательностей, найденных в эукариотических геномах. Однако, они несут низкокопийные последовательности, например инсерционные последовательности IS1 и IS186, представленные во фрагменте размером 50 килобаз. Имеются и другие примеры транспозабельных элементов, последовательности которых имеют способность перемещаться по геному, и, как в случае инсерционных элементов, переноситься от одного организма другому даже иногда между организмами разных видов.

Позиции IS1 и IS186 элементов, показанные на рис. 1, характерны именно для данного изолята кишечной палочки. Если изучать различные изоляты, то IS последовательности будут занимать разные позиции или могут полностью отсутствовать в геноме. Большинство других прокариотических геномов имеют очень мало повторяющихся последовательностей. Так, они практически отсутствуют в геноме *Campylobacter jejuni* NCNC11168. Однако, имеется исключение – *Neisseria meningitidis* Z2491, геном которой включает свыше 3700 копий различных типов повторяющихся последовательностей, составляющих в общей сложности 11% от генома и имеющих объем 2,18 мегабаз.

Кольцевая хромосома кишечной палочки имеет длину окружности, равную 1,6 мм. В то же время размеры клеток *E. coli* составляют в среднем 1,0 x 2,0 мкм. Как и эукариот, упаковка достигается с помощью ДНК-связывающих белков, которые упаковывают геном в определенном порядке. Итоговая структура внешне не похожа на эукариотическую хромосому, но, тем не менее, исследователи используют термин «бактериальная хромосома» для обозначения наследственного материала бактериальных клеток. Используется также термин «нуклеоид»

Значительная часть данных об организации ДНК в составе нуклеоида получена при изучении кишечной палочки. Первоначально было обнаружено, что кольцевой геном кишечной палочки является суперскрученным (supercoiled). Суперскрученность возникает, когда в двойную спираль ДНК вводятся дополнительные повороты (положительная суперскрученность) или когда они, наоборот, исчезают (отрицательная суперскрученность). В линейной молекуле возникающие торсионные силы тотчас раскрутили бы эту супер- или наоборот гипозакрученность и привели бы спираль в

нормальное состояние, но в кольцевой молекуле, не имеющей концов, такого не происходит. Вместо этого кольцевая молекула, вращаясь вокруг себя, образует более компактную структуру.

Суперскрученность рассматривают как идеальный способ упаковки круговой молекулы в компактную структуру в замкнутом пространстве. Данные об участии суперскрученности в упаковке круговой молекулы ДНК у кишечной палочки были получены в 70-х годах при изучении изолированного нуклеоида и затем были подтверждены как свойство ДНК, характерное для живой клетки. У кишечной палочки суперскрученность генерируется и контролируется двумя ферментами – ДНК-гиразой и ДНК-топоизомеразой I. Они играют важную роль в репликации ДНК. Изучение изолированного нуклеоида в живых клетках показало, что если в нить ввести разрывы, то молекула ДНК получает ограниченную свободу вращения. Наиболее вероятное объяснение заключается в том, что бактериальная ДНК связана с белками, которые ограничивают свободу ее вращения. Расщепление молекулы ферментами приводит к расслаблению структуры ДНК лишь в ограниченных участках. Согласно существующей модели ДНК присоединяется к белковому ядру (кору). Из точки присоединения выходит 40-50 суперскрученных петель. Каждая петля содержит приблизительно 100 килобаз суперскрученной ДНК. Это то количество ДНК, которое становится свободным после единичного расщепления молекулы.

Белковый компонент нуклеоида включает гиразу и топоизомеразу I. Эти два фермента изначально ответственны за поддержание суперскрученного состояния ДНК. Имеется также набор из минимум 4-х белков, играющих более специфическую роль в упаковке бактериальной ДНК. Одним из таких белков является упаковочный белок HU, который структурно сильно отличается от эукариотических гистонов, но выполняет сходную с ними функцию, образуя тетрамеры, вокруг которых наматывается приблизительно 60 пар снований ДНК. На одну клетку *E. coli* приходится 60 тысяч молекул таких белков, что достаточно для покрытия одной пятой части молекулы ДНК. Остается неизвестным, размещаются ли такие тетрамеры по всей молекуле ДНК или только вдоль ограниченного участка молекулы.

В последние годы стало ясно, что представленные выше взгляды на анатомию прокариотического генома слишком упрощены. Хотя большинство хромосом бактерий и архей на самом деле кольцевые, тем не менее, все больше обнаруживается линейных геномов. Первым из обнаруженных линейных геномов является описанный в 1989 году геном *Borrelia burgdorferi*, возбудителя болезни Лайма (боррелиоз). В течение

последующих лет было описано еще несколько линейных геномов, например у стрептомицетов.

Традиционные представления о прокариотическом геноме осложняют также данные, полученные при изучении плазмид. Как известно, плазида – это небольшая молекула ДНК, часто, но не всегда кольцевая. В бактериальной клетке она сосуществует с основной хромосомой. Некоторые типы плазмид способны объединяться с основным геномом, другие постоянно сосуществуют независимо от него. Плазмиды несут гены, которые обычно не присутствуют в основной хромосоме. Многие плазмиды способны перемещаться из одной клетки в другую, некоторые плазмиды обнаруживаются в бактериях, принадлежащих разным видам. Имеющиеся в настоящее время данные позволяют сделать заключение о том, что плазмиды являются независимыми генетическими элементами. Принято считать, что плазмидный состав прокариотической клетки не зависит от основного генома и не влияет на работу генома.

У *E.coli* K12, имеющей хромосому длиной 4,6 мегабаз и различные комбинации небольших плазмид, размером несколько килобаз и не играющих принципиальной роли для жизни клетки, основную хромосому можно обозначить как геном. Для других прокариот эта ситуация не так проста. *Vibrio cholerae* имеет две кольцевые молекулы ДНК. Одна из них имеет размер 2,96 мегабаз а другая – 0,7 мегабазы. На большей хромосоме присутствует 71 % из 3885 генов, составляющих геном. То есть геном составляют обе молекулы ДНК. Но более детальный взгляд показывает, что большинство генов, обеспечивающих основную клеточную активность (экспрессия генома, генерация энергии), а также гены патогенности локализуются на большей молекуле ДНК. Вторая молекула содержит менее существенные гены. Они характерны для плазмид, что позволяет назвать меньший геном мегаплазмидой, приобретенной предшественником *Vibrio* в какой-то период эволюции.

Геном *Deinococcus radiodurans* R1, содержащий множество генов, обеспечивающих устойчивость к радиационному воздействию, построен сходным образом. Но важные для жизнеобеспечения гены распределены между двумя кольцевыми плазмидами и двумя кольцевыми хромосомами. То есть, геном прокариот может быть мультипартитным, включающим множество отдельных молекул ДНК.

В свою очередь, такое заключение касается далеко не каждого генома прокариот. В связи с этим, интересны данные, полученные при изучении генома бледной трепонемы, вызывающей сифилис (*Treponema pallidum*). Геном трепонемы представляет собой одноцепочечную кольцевую молекулу ДНК длиной 1138 килобаз, содержащий 1041 ген.

Дополнительные различия в структуре прокариотических геномов обнаруживаются при сравнении способов упаковки бактериальных геномов и геномов архей. В отличие от бактерий, археи не имеют таких упаковочных белков, как HU. Взамен у них присутствуют белки, которые сходны с гистонами. В настоящее время нет информации о структуре нуклеоида архей, но предполагается, что гистоноподобные белки играют важную роль в упаковке ДНК.

Ранее уже упоминалось, что бактериальные геномы имеют очень компактную организацию с очень небольшими расстояниями между генами. В геноме кишечной палочки имеется некодирующая ДНК, но она составляет только 11-15 % от общего количества ДНК и распределена по геному небольшими сегментами. В этом отношении кишечная палочка типична для всех прокариот, чей геном к настоящему времени секвенирован. Имеется теория, что такая компактная организация является преимуществом прокариот, поскольку позволяет геному относительно быстро реплицироваться.

Характерной чертой геномов прокариот является наличие в их составе оперонов. Оперон – это группа генов, локализованных в геноме последовательно, между такими генами может находиться один или два разделяющих их нуклеотида. Все гены в опероне экспрессируются совместно, как одна единица. Такой тип построения и экспрессии генома свойственен только прокариотам. Типичным примером оперона является широко известный лактозный оперон – первый из открытых оперонов. Он содержит три гена участвующие в превращении дисахарида лактозы в моносахариды – глюкозу и галактозу. Моносахариды являются субстратом гликолиза, поставляющего клетке энергию.

Лактоза не является обычным компонентом природного окружения кишечной палочки. Поэтому большую часть времени оперон не экспрессируется и ферменты для утилизации лактозы не синтезируются. Когда во внешней среде появляется лактоза, оперон включается, три гена экспрессируются совместно, приводя к координированному синтезу лактозо-утилизирующих ферментов. Этот механизм является классическим примером регуляции активности генов у бактерий.

В геноме *E.coli* имеется почти 600 оперонов. Каждый содержит два и более генов. Сходная картина обнаружена в геноме *Bacillus subtilis* и других микроорганизмов. В большинстве случаев гены в составе оперона функционально связаны, кодируют группу белков, которые участвуют в реализации одного биохимического пути, например утилизация сахаров или синтез аминокислот. Привлекает относительная простота работы этих систем. С их помощью бактериальная клетка способна быстро и эффективно регулировать работу тех или иных групп генов, в зависимости от возникшей ситуации.

Однако и здесь обнаружены исключения из правила. Представитель архей *Methanococcus jannaschii* и бактерия *Aquifex aeolicus* имеют опероны, в которых гены редко связаны между собой по биохимической функции. Например, один из оперонов *Aquifex aeolicus* содержит шесть генов. Эти гены кодируют два белка, участвующие в рекомбинации ДНК, фермент, используемый при синтезе белков, фермент, вовлеченный в синтез нуклеотидов, фермент липидного синтеза и белок, необходимый для подвижности клеток. Такой оперон является типичным для двух последних микроорганизмов. Заметим, что в традиционном понимании экспрессия оперона приводит к координированному синтезу ферментов, необходимых для реализации одного биохимического пути. У *Methanococcus jannaschii* и *Aquifex aeolicus* это правило не выдерживается. Таким образом, расшифровка геномов сотен микроорганизмов привело к изменениям взгляда на оперон в его классическом понимании.

Реализация многочисленных проектов, направленных на расшифровку геномов различных организмов, привели к изменениям в понимании того, что такое вид у прокариот. В микробиологии определение вида всегда было проблемой, поскольку стандартные биологические определения вида трудно применить к микроорганизмам. Таксономия Линнея описывала вид в морфологических терминах, где все члены вида имеют ту же или очень схожую морфологию. В микробиологии важный вклад в классификацию внес Роберт Кох, который применил окрашивание бактерий и биохимические тесты для их классификации. Такой подход имеет много минусов, поскольку бактерии одного и того же вида могут иметь разную морфологию и разные метаболические черты.

В настоящее время при определении вида исследователи основываются на генетическом сходстве организмов. Однако и здесь возник ряд проблем. Для прокариот межвидовой барьер, связанный с переносом генов, не является решающим. Разные штаммы одного вида могут иметь очень существенные различия в геномных последовательностях. Обнаружены довольно большие штамм-специфические наборы генов у разных штаммов одного вида. Впервые это было показано при сравнении двух штаммов *Helicobacter pylori* - микроорганизма, вызывающего желудочно-кишечные заболевания. От больных было изолировано два штамма, имеющих протяженность геномов 1,67 и 1,64 мегабаз, соответственно. Большой геном содержал 1552 гена, меньший 1495 генов. Из этого числа генов у данных штаммов одинаковы были лишь 1406. Остальные 6-7 % генома были уникальны для каждого из охарактеризованных штаммов.

Более серьезные различия были выявлены при сравнении геномов стандартного лабораторного штамма *E.coli K12* и одного из наиболее патогенных штаммов кишечной

палочки O157:H7. Длина геномов у них очень сильно различалась – 4,64 мегабазы для K12 и 5,53 мегабазы для O157:H7.

Геном штамма O157:H7 имеет более 200 отдельных вставок или, как их обозначили, O-островков, содержащих в общей сложности 1387 генов, отсутствующих у *E. coli* K12. Многие из этих генов кодируют токсины и другие белки, участвующие в реализации патогенных свойств штамма O157:H7. Но этим не ограничиваются отличия данных штаммов друг от друга. Штамм K12 имеет 234 сегмента уникальной ДНК, называемых теперь K-островками. В среднем они меньше O-островков, но в сумме содержат 528 генов, отсутствующих в клетках штамма O157:H7. Таким образом, штамм-специфические гены у двух сравниваемых штаммов *E. coli* составляют 26 и 12 % от общего количества генов.

Особенности прокариотических геномов оказались еще более выраженными, когда были изучены геномы других бактерий и архей. Поскольку гены могут сравнительно легко перемещаться из генома одного прокариота в другой, то не является удивительным, что один и тот же ген может быть найден в геномах разных видов. Латеральный перенос генов неоднократно документирован при анализе секвенирования геномов прокариот. Большинство геномов содержат несколько сотен килобаз ДНК, полученной от других видов. Например, 12,8 % генома *E. coli* K12 (0,59 мегабаз) получены клетками путем латерального переноса генетической информации. Более того, перенос генетической информации осуществляется не только между разными видами бактерий, но и между бактериями и археями. Так, термофильная бактерия *Thermatoga maritima* имеет 877 генов. Как предполагают, 451 ген получен от архей. С такой же вероятностью и в таких же масштабах происходит перенос генов в обратном направлении (от бактерий к археям). Вероятно прокариоты, живущие в сходных экологических нишах, обмениваются генами с целью повышения жизненных возможностей. Многие гены *Thermatoga maritima*, полученные от архей, обеспечивают бактериям устойчивость к высоким температурам.

Горизонтальный (латеральный) перенос генов играет важную роль в эволюции прокариот. В отличие от высших организмов эволюционная история бактерий и архей не может быть полноценно описана в виде дерева, поскольку включает горизонтальный перенос генов между видами, что необходимо учитывать при анализе эволюционных взаимоотношений между прокариотами.

Несмотря на то, что в настоящее время опубликована нуклеотидная последовательность многих прокариотических геномов, тем не менее составить полный каталог генов хотя бы для одного организма пока невозможно. Причина заключается в

том, что функция многих генов остается неизвестной. Например, неизвестна функция более 1000 генов *E. coli K12*.

1.3. Повторяющиеся последовательности геномной ДНК

Повторяющиеся последовательности ДНК изучены намного лучше, чем многие другие компоненты геномов. Повторяющиеся последовательности найдены во всех организмах, включая организм человека. Существуют различные типы повторяющихся последовательностей, создано несколько систем их классификации. Схема, которой мы будем пользоваться, делит повторяющиеся последовательности по типу их кластеризации и по тому, как они распределены по геному.

1. Тамдемно повторяющаяся ДНК.

Ее наличие является общей чертой эукариотических геномов. В геномах прокариот тамдемно повторяющиеся последовательности встречаются редко. Такой тип повторов называют также сателлитной ДНК, поскольку фрагменты ДНК, содержащие тамдемно повторяющиеся последовательности, образуют «сателлитные» полосы при фракционировании ДНК центрифугированием в градиенте плотности.

Когда расщепленную на фрагменты длиной 50-100 килобаз человеческую ДНК фракционируют, то она дает основную и три сателлитных полосы. Основная полоса содержит фрагменты ДНК, представляющие собой в основном однокопийные последовательности с содержанием ГЦ нуклеотидов, близким к 40,3 %. Это значение в целом характерно для человеческого генома. Сателлитные полосы содержат фрагменты повторяющейся ДНК, в котором содержание ГЦ пар атипично.

В качестве сателлитной ДНК классифицировано два типа тамдемно повторяющейся ДНК. Это минисателлиты и микросателлиты. Минисателлиты образуют кластеры длиной до 20 тысяч оснований. Повторяющиеся единицы в их составе имеют протяженность до 25 пар оснований. Микросателлитные кластеры короче. Они обычно меньше 150 пар оснований. Повторяющаяся единица составляет 13 пар оснований или меньше.

Сателлитная ДНК найдена в центромерах и является минисателлитной. Участки построены из серий тамдемных повторов в сотни килобаз длиной. Отдельный геном может содержать несколько различных типов сателлитной ДНК, каждая содержит разные повторяющиеся единицы длиной от 5 до 200 пар оснований. Три сателлитные полосы человеческой ДНК включают, по крайней мере, четыре разных типа повторов.

Один из типов тамдемно повторяющейся ДНК уже упоминался – это альфоидная ДНК, найденная в центромерных участках хромосом. Хотя некоторые сателлитные ДНК

разбросаны по всему геному, большинство локализовано в центромерах, где они играют структурную роль, как место связывания специальных центромерных белков. Кроме того, предполагается, что наличие повторяющейся ДНК в центромерах отражает тот факт, что это последний по времени реплицирующийся регион хромосом. Вследствие отсутствия тандемно повторяющихся последовательностей здесь отсутствуют точки инициации репликации (так называемые ориджины репликации). В результате репликация откладывается на самый конец клеточного цикла.

Теломерная ДНК также является примером минисателлитной ДНК. Кроме теломерных микросателлитов некоторые эукариотические геномы содержат различные другие кластеры сателлитной ДНК. Многие, но не все, расположены около конца хромосом.

Функции микросателлитов во многих случаях загадочны. Типичный микросателлит состоит из 1-, 2-, 3- или 4-х пар оснований. Каждая единица повторяется 10-20 раз, как это проиллюстрировано для микросателлитов в локусе гена бета-цепи Т-клеточного рецептора. Хотя каждый микросателлит относительно короток, их в геноме очень много.

У человека, например, микросателлиты, состоящие из повторов цитозин-аденин, составляют до 0,25 % генома. В целом их протяженность имеет длину порядка 8 мегабаз. Простые повторяющиеся последовательности, построенные из одинаковых нуклеотидов, такие например как поли(А), составляют 0,15% генома. Хотя их функция неизвестна, микросателлиты широко используются в практических целях генетиками. Многие микросателлиты вариабельны. Это означает, что количество повторяющихся последовательностей варьирует у разных представителей вида. Вариабельность возникает, когда микросателлиты копируются в ходе репликации ДНК. Это приводит к изменению количества повторяющихся единиц. Нет ни одного человека, у которого комбинация микросателлитов была бы идентичной комбинации микросателлитов другого человека. Если знать, как построены микросателлиты у каждого отдельного человека, то можно построить генетический паспорт каждой персоны. Исключение здесь может быть только одно – однояйцевые близнецы. Генетические профили широко используют в криминальной медицине, при определении отцовства. Анализ микросателлитов может быть использован также для определения взаимоотношений между организмами в популяциях животных и растений.

Тандемно повторяющиеся последовательности появляются при экспансии последовательности-предшественника или при проскальзывании репликации (replication

slippage), как это описано для микросателлитов, или путем рекомбинации ДНК. Каждое из таких событий приводит к появлению серии взаимно связанных повторов.

2. Интерсперсные геномные повторы.

Интерсперсные (разбросанные по геному) повторяющиеся последовательности появляются в геноме с помощью разных механизмов, каждый из которых может приводить к копированию повторяющихся единиц. Наиболее частый путь, с помощью которого это происходит – транспозиция. Большинство интерсперсных повторов являются результатом транспозиционной активности.

Существует два альтернативных механизма транспозиции. Один из них постулирует участие РНК-интермедиатов, а второй – нет. Механизм, который реализуется с участием РНК-интермедиатов, называется ретротранспозицией. Он включает несколько этапов:

1. РНК-копия транспозона синтезируется с помощью обычного процесса транскрипции.

2. РНК-транскрипт копируется в ДНК, которая изначально существует как независимая молекула, вне генома. Превращение РНК в ДНК, то есть обратная транскрипция, требует присутствия обратной транскриптазы или ревертазы. Часто обратная транскриптаза кодируется геном, находящимся в самом транспозоне, и транслируется с помощью РНК-копии, синтезированной на первом этапе.

3. ДНК-копия транспозона интегрируется обратно в геном, в ту же самую хромосому или в другую хромосому. В конечном итоге появляется две копии транспозона в разных участках генома.

РНК-транспозоны или ретроэлементы являются характерной чертой генома эукариотических клеток, но для прокариот такие элементы не известны. РНК-транспозоны привлекают особое внимание, поскольку они имеют явное сходство с другими типами ретроэлементов и свободно живущими вирусами, называемыми ретровирусами. Заметим, что к ретровирусам относят вирус иммунодефицита человека. Структурные взаимоотношения между ними могут быть суммированы следующим образом:

1. Ретровирусы - это РНК-содержащие вирусы. Они инфицируют многие типы позвоночных. Попав в клетку, РНК-геном копируется в ДНК с помощью обратной транскриптазы, кодируемой вирусным геном *pol*. ДНК копия затем интегрируется в геном клетки-хозяина. Новые вирусы могут продуцироваться путем копирования интегрированной ДНК в РНК и упаковки РНК в белки, кодируемые геномом вируса с образованием вирусной частицы.

2.Эндогенные ретровирусы. Являются ретровирусными геномами, интегрированными в хромосомы позвоночных. Некоторые из них активны и на определенных стадиях жизни клетки могут направлять синтез экзогенных вирусов, но в большинстве случаев – они не имеют возможности образовывать вирусные частицы. Неактивные последовательности образуют геномные повторы, но не способны к дополнительной пролиферации.

3. Ретротранспозоны. Имеют последовательности, сходные с эндогенными ретровирусами, но чаще встречаются в геномах не позвоночных эукариот. Это растения, грибы, беспозвоночные, простейшие. В некоторых геномах ретротранспозоны имеют очень большое количество копий, существует несколько различных типов ретротранспозонов. У кукурузы большинство повторяющихся последовательностей представлено ретротранспозонами. Они составляют почти половину генома кукурузы.

Имеется два семейства ретротранспозонов. Первое – это семейство *Ty3/gypsy*. Члены этого семейства несут тот же набор генов, что и эндогенные ретровирусы. *Ty3* и *gypsy* являются представителями этого семейства ретротранспозонов у дрожжей и дрозофилы, соответственно. Второе семейство ретротранспозонов – это семейство *Ty1/copia0020*. Члены этого семейства не имеют гена *env*. Оба семейства способны к транспозиции через механизм, представленный выше, но отсутствие гена *env* означает, что семейство *Ty1/copia* не может формировать инфекционные вирусные частицы. В отличие от них ретротранспозоны семейства *Ty3/gypsy* могут формировать вирусы, которые выглядят как ретровирусы беспозвоночных. Хотя ретротранспозоны по своей природе являются интерсперсными элементами, тем не менее, иногда их находят кластеризованными, как результат присутствия сайтов интеграции подвижных элементов в геном.

Три типа описанных ретроэлементов называют LTR элементами. Все они имеют длинные концевые повторы, которые принимают участие в процессе транспозиции, то есть перемещении этих элементов по геному. Наряду с LTR-элементами существуют ретроэлементы, не содержащие LTR. Их называют ретропозонами. У млекопитающих к ним относятся:

- LINE –длинные интесперсные ядерные элементы (long interspersed nuclear elements). Они содержат ген, подобный обратной транскриптазе. Предполагается, что он участвует в транспозиции этих элементов. Примером является человеческий элемент LINE-1, который имеет длину 6,1 килобаз и содержится в геноме человека в количестве, равном 615 тысяч копий.

SINE – короткие интерсперсные ядерные элементы (short interspersed nuclear elements). Они не имеют гена обратной транскриптазы, но, тем не менее, сохраняют свойство перемещаться по геному. Вероятно, они используют обратную транскриптазу других ретроэлементов. В геноме человека чаще всего встречается элемент, называемый Alu, количество копий которого свыше 1 миллиона. Некоторые Alu элементы активно копируются в РНК, создавая возможности для увеличения их копийности.

Не все транспозоны нуждаются в РНК-интермедиате. Многие способны перемещаться прямым способом из одного участка ДНК в другой. Их называют ДНК-транспозонами. Эти элементы перемещаются с помощью двух различных механизмов транспозиции. Один заключается в прямом взаимодействии между донорским транспозоном и целевым сайтом, приводя к копированию донорского элемента. Такой механизм называют репликативной транспозицией. При втором механизме происходит вырезание донорского элемента и реинтеграция в новый сайт (консервативная транспозиция). Оба механизма осуществляются с участием ферментов, которые обычно кодируются генами, расположенными внутри элемента.

У эукариот ДНК-транспозонов значительно меньше, чем ретротранспозонов. Однако первый транспозон был обнаружен именно у эукариот. В 1950 году Barbara McClintock обнаружила первый транспозон, названный Ac/Ds, в геноме кукурузы. Ее заключения о том, что некоторые гены подвижны и могут перемещаться из одного участка хромосомы в другой, были построены на генетических экспериментах. Молекулярные основы этого явления были расшифрованы только в 70-х годах.

ДНК-транспозонов намного больше в прокариотических геномах, чем РНК-транспозонов. Инсерционные последовательности IS1 и IS186, представленные при описании фрагмента генома размером 50 килобаз, являются примерами ДНК-транспозонов кишечной палочки. В ее геноме может содержаться более 20 транспозонов разных типов. IS элементы могут перемещаться как репликативно, так и консервативно. У кишечной палочки известны также и другие типы ДНК-транспозонов. Они типичны для прокариот:

- Композитные (сложные) транспозоны. В своей основе имеют IS элементы, которые фланкируются сегментами ДНК, содержащими обычно один или большее количество генов. Гены часто кодируют антибиотикорезистентность. Перемещение композитного транспозона катализируется транспозазой, кодируемой одним или большим количеством IS элементов. Композитные транспозоны используют консервативный механизм транспозиции.

- Транспозоны Tn3 типа. Имеют собственный ген транспозазы и поэтому не нуждаются в фланкирующих элементах для перемещения по геному. Tn элементы перемещаются консервативно.

- Транспозабельные фаги. Являются бактериальными вирусами, которые перемещаются по геному репликативно, что является частью их нормального инфекционного цикла.

1.4. Геном человека

Основной характеристикой нуклеиновой кислоты является ее нуклеотидная последовательность. Если известна нуклеотидная последовательность, то можно идентифицировать гены, которые содержатся в последовательности и детально исследовать функциональные свойства генов. Начиная с середины 70-х годов прошлого века, исследователи имели возможность получать нарастающую информацию о нуклеотидной последовательности все более длинных участков генома. Развитие методологии секвенирования ДНК привело к тому, что в 90-х годах были получены полные последовательности геномов нескольких организмов. Подобные исследования проводились в рамках больших геномных проектов. Самым крупным проектом в этой области исследований стал проект «Геном человека». Он был начат в 1984 году. Намного меньший по размеру митохондриальный геном человека был секвенирован еще в начале 80-х годов.

Проект «Геном человека» финансировался правительствами ряда стран, в том числе в этом проекте участвовал и Советский Союз. Второй геномный проект осуществлялся частной компанией Celera Genomics штата Maryland (США). Он был начат в 1998 году. Оба проекта завершили предварительный анализ человеческого генома в 2001 году и первые результаты были опубликованы в ведущих научных журналах Nature и Science в феврале того же года (HGS, 2001; Venter *et al.*, 2001). Было отсеквенировано примерно 83-84 % целого генома. При этом секвенированными оказались участки, наиболее важные для человеческого генома. Большинство нерасшифрованных последовательностей, составляющих 16-17 % генома, в основном представляли собой концевые последовательности хромосом (теломеры) и центромерные последовательности, где локализовано очень мало генов (Bork, Coley, 2001).

Несмотря на всю неполноту анализа генома, к 2001 году была определена последовательность 2,6 миллиарда пар оснований. Если печатать последовательность таким образом, чтобы 60 нуклеотидов умещались в отрезок, равный 10 сантиметрам, то

последовательность займет строчку длиной 5 тысяч километров – расстояние от Монреаля до Лондона. Последовательность займет примерно 3000 книг стандартного размера. Окончательное завершение проекта было осуществлено в 2004 году. Геном, размером 2,85 миллиарда оснований был расшифрован на 99 %. Аккуратность секвенирования в конечной версии составила 1 ошибку на 100 тысяч позиций.

Было определено, что геном содержит 20-25 тысяч генов. Функция известна лишь для половины из них. Средний размер генов, включая экзоны и интроны, составляет 28 тысяч пар нуклеотидов. Соотношение интронов к экзонам равно 24:1. Межгенные участки с неизвестной функцией составляют 75 % генома. То есть у человека кодирующие участки составляют лишь 1 % от всего генома. Если не считать псевдогены и фрагментированные гены, то остальные участки генома – это интроны, повторяющиеся последовательности и бессмысленные последовательности.

Предполагают, что бессмысленные участки ДНК служат пассивной защитой от опасных вирусов, поскольку вероятность попадания разрушающей вирусной информации в смысловую область резко уменьшается. Огромные участки ДНК остаются "нераспаханной целиной" в течение всей жизни клеток. Пустые концевые участки хромосом, как и область центромеров (первичные места спаривания родительских парных хромосом), важны для сохранения вида: они определяют строгое распознавание макрорельефа хромосом как органелл клетки (а не микрорельефа молекулы ДНК) одного вида по принципу "ключ-замок".

Другими словами, спермии человека не оплодотворяют яйцеклетку обезьяны и наоборот, потому что хромосомы клеток двух видов не распознают друг друга. Поэтому предполагается, что "бессмысленные участки" ДНК осмысленно работают в хромосоме, защищая вид от вторжения чужеродной ДНК.

Несмотря на завершение проекта в 2004 году, детальная информация о геномной характеристике каждой из хромосом печаталась в *Nature* и *Science* вплоть до середины 2006 года. Обнаружено, что более всего генов содержит первая хромосома, а наибольшая плотность генов наблюдается в 19 хромосоме, в связи с чем ее назвали «генной столицей». Наименьшее количество генов содержит У хромосома. В таблице 1 представлена сжатая информация о части хромосом и ссылки на публикации, где можно найти более детальные данные.

Таблица 1. Характеристика хромосом человека

Хро- мо- сома	Кол-во белок- кодирующих генов	Кол-во псевдо генов	Примечания	Ссылка
1	3141	991	Наибольшее кол-во генов. Много перекрывающихся генов	Nature. 2006 May 18;441(7091):315-21.
2	1346	1,239	Хромосома 2 – продукт слияния двух небольших анцестральных хромосом	Nature. 2005 Apr 7;434(7034):724-31.
5	923		Низкая плотность генов и множество дупликаций генов	Nature. 2004 Sep 16;431(7006):268-74.
6	1557	633	6% генома, 166880988 пар оснований, Ген HLA-B наиболее полиморфен в геноме.	Nature. 2003 Oct 23;425(6960):805-11
7	1917		Большое кол-во (440) хромосомных перестроек, ассоциированных с болезнями	Science. 2003 May 2;300(5620):767-72.
9	1149	426	109044351 пар оснований, много дупликаций генов	Nature. 2004 May 27;429(6990):369-74.
10	816	430	131666441 пар оснований, перекрывание кодирующих генов и 67 антисенс- транскриптов. В сравнении с шимпанзе – нонсенс-мутация в 21 гене	Nature. 2004 May 27;429(6990):375-81.
11	1524 гена	765	Высокая плотность генов-11,6 генов на мегабазу. Четверть генов перекрывается с другими генами. Гены высококкластеризованы. 40% генов – в 28 кластерах. С	Nature. 2006 Mar 23;440(7083):497-500.

			хромосомой ассоциировано 171 заболевание	
12	1400 генов		487 локусов наследственных заболеваний	Human Genome Sequencing Center, USA.
13	633 гена	296	Идентифицировано 95.4% белок-кодирующих генов. Низкая плотность генов – 6,5 генов на 1 мегабазу, 105 РНК- кодирующих генов	Nature. 2004 Apr 1;428(6982):522-8.
14	1050 генов и генных фрагментов	393	Короткое плечо содержит в основном гены, кодирующие РНК. Длинное плечо кодирует гены белков.	Nature. 2003 Feb 6;421(6923):601-7.
16	880 генов, 19 генов, кодирующих РНК	341, + 3 РНК псевдоге на		Nature. 2004 Dec 23;432(7020):988-94.
19	1461 генов	321	Наибольшая плотность генов (более 12 на 1 мб) «Генная столица»	Nature. 2004 Apr 1;428(6982):529-35.
21	127 извест-ных, 98 предсказанных генов	59	Наименьшая аутосома	Nature. 2000 May 18;405(6784):311-9.
У	78 генов		Наименьшее кол-во генов. Содержит 8 массивных палиндромов, в которых локализовано по крайней мере 6 тестис-генов.	Nature. 2003 Jun 19;423(6942):825-37.
Х	1098 генов		Одну треть Х-хромосомы занимают LINE-1 повторы. 113 Х-связанных гена вызывают более 168 наследственных заболеваний	Nature. 2005 Mar 17;434(7031):325-37.

В ходе реализации проекта были расшифрованы последовательности фрагментов многих геномов. Геномы имеют между собою отличия. Как известно, идентичные геномы могут быть только у однойцевых близнецов. Различия между геномами формируются во многом благодаря так называемым однонуклеотидным полиморфизмам (SNP, single nucleotide polymorphism). В русской транскрипции они звучат как СНИПы, то есть, позиции, в которых обнаруживаются однонуклеотидные различия между геномами. В настоящее время идентифицировано более 1,4 млн СНИПов, в среднем один на 2000 пар нуклеотидов (SNP Group, 2001). Кроме того, в среднем на каждые две тысячи пар нуклеотидов имеется по одному микросателлиту, называемому также коротким tandemным повтором (STR). В составе микросателлитов присутствует серия повторяющихся нуклеотидов, например САСАСАСА. Количество повторов вариабельно и отличается у разных лиц. Существуют СНИПы и микросателлиты, для которых отсутствует информация об их влиянии на функции генома. Однако, во многих других случаях такая информация получена. По сути дела СНИПы – это закрепившиеся в популяции данного вида точечные мутации, возникшие на каком-то этапе эволюции. Внутри генов лежит 60 тысяч СНИПов. Часть из них оказывают прямое влияние на активность генов, приводя к вариациям между отдельными людьми на уровне биологических признаков. Кроме того, обнаруживают все большее количество СНИПов, вызывающих заболевания или опосредованно ассоциированных со склонностью к различным заболеваниям.

В соответствии с данными проекта «Геном человека» различия между геномами отдельных лиц могут составлять 0,1 %. Различия между геномами человека и шимпанзе составляют примерно 2 %. Однако, в настоящее время эти данные пересматриваются. Обнаружено, что некоторые гены и даже отдельные участки генома могут быть представлены несколькими копиями. Это отражается на физиологических и биохимических признаках организма, на поведенческих реакциях. Приводятся данные о возможном различии геномов человека, достигающем 12 %. Ведется активное изучение этого вопроса. Одновременно проводятся работы, направленные на сравнительное изучение геномов человека и шимпанзе (<http://www.vz.ru/society/2006/12/9/59727.html>).

Для того, чтобы получить общие представления о строении человеческого генома рассмотрим один из сегментов хромосомы 7 длиной 50 килобаз. Этот сегмент формирует часть локуса бета-цепи человеческого Т-клеточного рецептора, который имеет намного большую длину, равную 685 килобаз. Он кодирует бета-цепь Т-клеточного рецептора, участвующего в реализации иммунного ответа. В составе указанного фрагмента присутствуют следующие составные части:

- Один ген. Этот ген называется TR γ 4 и содержит информацию об аминокислотной последовательности трипсиногена – неактивного предшественника пищеварительного фермента трипсина. Ген TR γ 4 является представителем семейства генов трипсиногена, представленного двумя кластерами в конце локуса бета-цепи Т-клеточного рецептора. Эти гены не имеют никакого отношения к иммунному ответу. Они просто перекрываются с локусом бета цепи Т-клеточного рецептора. TR γ 4 является примером прерывистого гена. Заложённая в гене информация об аминокислотной последовательности трипсиногена, расщеплена на пять экзонов, разделённых четырьмя некодирующими интронами.

- Два генных сегмента. В составе фрагмента имеется два генных сегмента V28 и V29-1, каждый из которых кодирует часть бета-цепи Т-клеточного рецептора. Сегменты V28 и V29-1 не являются полными генами. Они представляют собой только фрагменты гена и перед тем как экспрессироваться должны связаться с другими сегментами гена этого же локуса. Такие перестройки генетического материала происходят только в Т-лимфоцитах и являются примером того, как могут происходить изменения в активности генов в процессе дифференцировки клеток. Как и TR γ 4, сегменты V28 и V29-1 являются прерывистыми генами и состоят из экзонов и интронов.

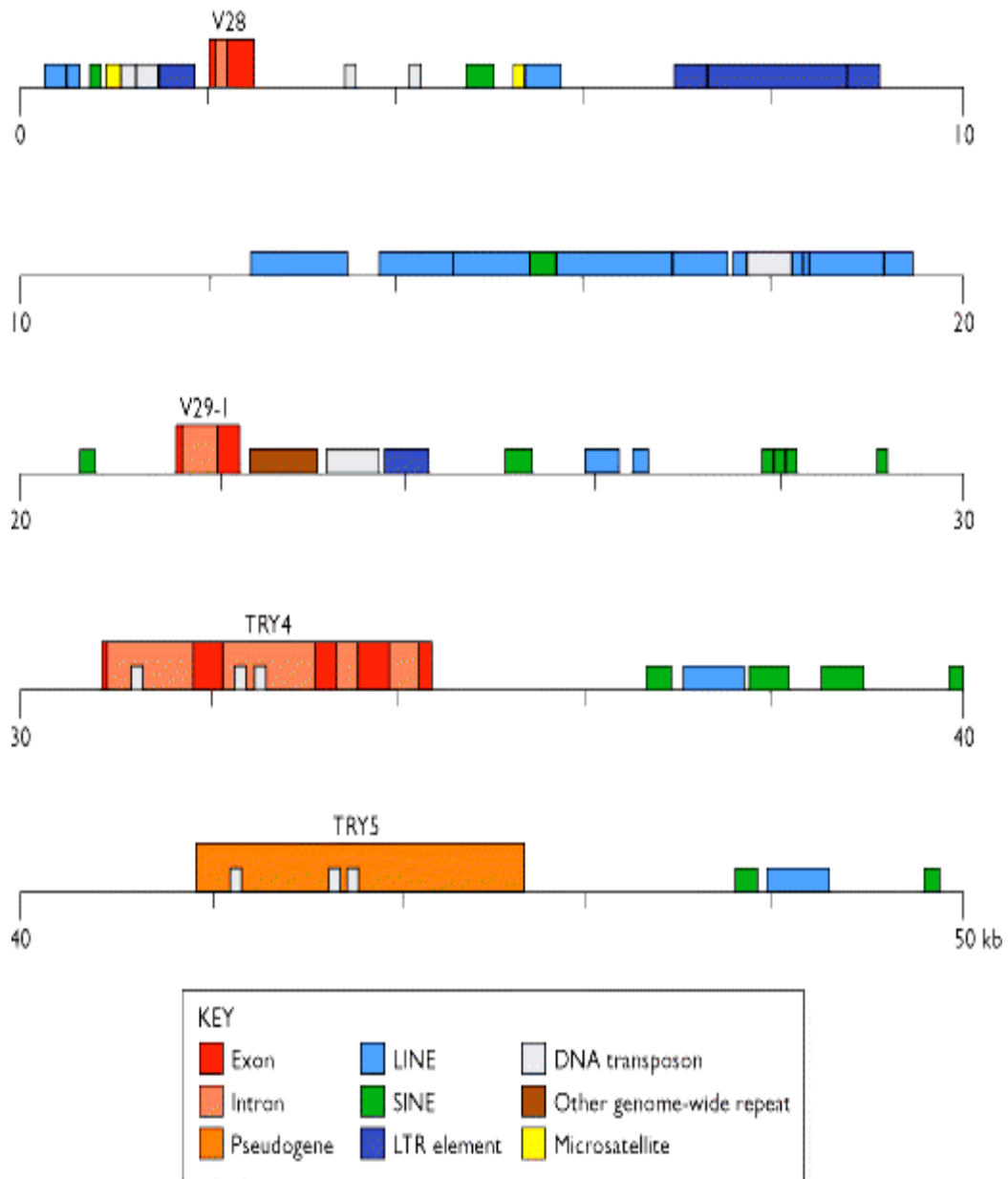


Рис. 2. Структура фрагмента 7 хромосомы генома человека длиной 50 килобаз (Т.А.Браун, 2002).

- Один псевдоген. Псевдоген является нефункционирующей копией гена. Обычно нуклеотидная последовательность в составе такого гена меняется так, что ген теряет способность к считыванию. Этот псевдоген назван TRY5. Он также относится к семейству генов трипсиногена.

- 52 повторяющиеся последовательности. В составе короткого фрагмента присутствуют длинные интерсперсные ядерные повторы LINEs, короткие интерсперсные ядерные повторы (SINEs), LTR элементы и ДНК-транспозоны.

- Два микросателлита, которые, как замечено выше, являются последовательностями, в которых тандемно повторяются короткие мотивы. Один из микросателлитов имеет мотив GA, повторяющийся 16 раз. Второй микросателлит имеет последовательность, построенную из шести повторяющихся повторов TATT.

- Наконец, примерно 50 % последовательности сегмента построено из неповторяющихся, некодирующих, однокопийных участков ДНК с неизвестной функцией и значением.

Теперь рассмотрим указанные компоненты генома более подробно.

Мы рассматриваем гены как наиболее важные части человеческого генома, поскольку они несут биологическую информацию. Большинство генов определяют строение одного или большего количества белковых молекул. Экспрессия этих генов осуществляется через промежуточную молекулу – матричную РНК – которая транспортируется из ядра в цитоплазму, где направляет синтез белков, кодируемых геном. Часть генов не кодирует белок. Их продуктом являются некодирующие молекулы РНК, играющие различную роль в клетке.

Каждый из генов и генных сегментов, представленных на фрагменте размером 50 килобаз, является прерывистым. Биологическая информация, заложенная в них, разделена на серию экзонов с помощью некодирующих интронов. Большинство человеческих генов являются прерывистыми, в среднем содержат девять экзонов на ген, хотя имеются некоторые гены, где экзонов намного больше. Ген, имеющий наибольшее количество экзонов у человека - это ген мышечного белка титина. Он содержит 178 экзонов и длину, равную 80780 парам оснований. Это самый длинный известный в настоящее время ген.

В процессе экспрессии гена первоначально на матрице ДНК синтезируется матричная РНК, которая является копией гена, включая экзоны и интроны. Процесс, называемый сплайсингом, удаляет интроны из пре-матричной РНК и соединяет экзоны вместе, формируя матричную РНК, направляющую синтез белка. Во многих случаях пре-матричная РНК подвергается альтернативному или дифференциальному сплайсингу, который дает серию мРНК, состоящих из различных комбинаций экзонов и дающих разные белки. Так же как и гены сами по себе, пре-матричная РНК, транскрибируемая с последовательности гена, содержит на концах участки, фланкирующие первый и последний экзоны. Они называются 5'-нетранслируемый регион и 3'-нетранслируемый регион, соответственно.

Одновременно с созреванием матричной РНК, может происходить образование микроРНК, кодируемых интронами. Один интрон может кодировать несколько микроРНК длиной примерно 22 нуклеотида. В образовании микроРНК участвует РНК-полимераза II. Сначала образуется кэпированный и полиаденилированный предшественник про-микроРНК, затем с помощью ферментной системы, названной «микропроцессором» образуется пре-микроРНК, которая переносится экспортином-5 из ядра в цитоплазму. Здесь происходит окончательное дозревание микроРНК с помощью цитоплазматической РНК-азы III Dicer и включение однонитевой микроРНК в состав комплекса, индуцирующего сайленсинг РНК (RISC). МикроРНК интронов обеспечивают пост-транскрипционный сайленсинг генов. RISC комплементарно связывается с 3'-нетранслируемым регионом матричной РНК, индуцируя ингибирование трансляции или деградацию мРНК (Valencia-Sanchez et al., 2006).

В составе генома человека существуют гены, представленные генными сегментами, расположенными на значительном расстоянии друг от друга. Так, в незрелых Т-лимфоцитах имеется более 100 генных сегментов, кодирующих вариабельные участки альфа и бета-цепей Т-клеточного рецептора. При созревании Т-лимфоцитов они связываются вместе в различных комбинациях, давая разные по строению и функциональным свойствам активные центры Т-клеточных рецепторов. Заметим, что такой механизм не является сплайсингом, поскольку перестройке подвергаются не РНК, а гены сами по себе. Его называют соматической рекомбинацией. Перестроенный ген, получаемый после соматической рекомбинации, напоминает усредненный ген, но имеет одну очень важную особенность. Он образуется только в Т-лимфоцитах. Соматической рекомбинации подвергаются также генные сегменты легкой и тяжелой цепей иммуноглобулинов. Но это происходит только в В-лимфоцитах.

Большинство генов кодирует белки, но около 2500 генов определяет строение различных типов некодирующих РНК. Почти четверть белок-кодирующих генов участвуют в экспрессии, репликации и поддержании генома. Еще 20 % генов определяют строение компонентов сигнал-передающих путей, которые регулируют экспрессию генов и другие активности клетки, в том числе ответ на внутриклеточные сигналы. Ферменты, ответственные за реализацию основных биохимических функций, кодируют 17,5 % генов. Продукты остальных генов участвуют в транспорте компонентов в клетку и из нее, в фолдинге белков, иммунном ответе, синтезе структурных белков.

Важной чертой генома человека является наличие в нем псевдогенов, что характерно и для многих других организмов. Псевдогены являются эволюционными реликтами, показывающими, что геном человека постоянно подвергается изменениям. В геноме

человека присутствует более 3 тысяч псевдогенов (Venter et al., 2001). Различают два типа псевдогенов.

- Традиционные псевдогены – это гены, которые инактивировались вследствие мутаций, возникающих как результат изменения нуклеотидной последовательности. Многие мутации имеют минимальный эффект на активность генов, но некоторые достаточно важны для функциональной целостности генов. Даже одиночная мутация может приводить к полной потере функциональной активности гена. Когда псевдоген становится нефункциональным, он может подвергнуться дальнейшей деградации с накоплением мутаций. Примером такого гена является псевдоген TRY5.

- Процессированные псевдогены - возникают не вследствие мутаций, а за счет нарушений, возникающих при экспрессии гена. Процессированные псевдогены возникают из копий мРНК, с которых синтезируется ДНК, интегрирующаяся затем в геном. Поскольку процессированные псевдогены являются молекулами матричной РНК, они не содержат интронов, присутствовавших в родительском гене. В их составе также отсутствуют 5'-НТР - нуклеотидные последовательности, лежащие в родительском гене перед кодирующей последовательностью. Поскольку в этих участках расположены сигналы для включения экспрессии генов, псевдогены становятся неактивными.

Кроме псевдогенов геномы содержат также другие эволюционные реликты – усеченные гены (truncated), которые потеряли нуклеотидную последовательность с одного из концов полного гена, и фрагменты генов - короткие участки того или иного гена.

Секвенирование генома показало, что приблизительно 62 % человеческого генома составляют межгенные участки. Это часть генома, которая размещена между генами. Функции этих участков не известны. Для этих последовательностей используют термин мусорная, бессмысленная (junk) ДНК. Но данный термин теряет свое значение, поскольку в последние годы появились данные о том, что функция у этих участков все же может существовать.

Геномные повторы и сателлиты. Повторяющаяся ДНК делится на две категории: разбросанные по геному (genome-wide) или интерсперсные повторы, чьи индивидуальные повторяющиеся единицы распределены по всему геному беспорядочным образом, и тандемно повторяющаяся ДНК, чьи повторяющиеся единицы расположены друг рядом с другом. Интересной чертой всех этих типов повторов является то, что все они являются производными транспозабельных элементов – мобильных сегментов ДНК, которая способна перемещаться по геному из одного места в другое. Многие из этих элементов оставляют копии самих себя, когда перемещаются в другое место. Этим объясняется большая распространенность этих элементов по геному.

Как указано выше, в геноме человека имеется два основных класса транспозабельных элементов: те, которые перемещаются через промежуточную РНК, и те, которые не используют такой механизм. Элементы LINE, SINE и LTR принадлежат к первому классу, ДНК-транспозоны – ко второму. Каждый класс элементов подразделяется на субкатегории. Например, SINE элементы, которые наиболее многочисленны в геноме среди интерсперсных повторов, подразделяются на три субтипа: Alu элементы (1090000 копий в геноме), MIR (393000 копий), MIR3 (75 тысяч копий). Все вместе интерсперсные повторы занимают до 44 % всей последовательности генома (1400 мегабаз ДНК).

Микросателлиты и минисателлиты являются примерами тандемно повторяющейся ДНК. Известно, что микросателлиты произошли благодаря ошибкам в процессах, отвечающих за копирование генома в ходе клеточного деления. Они могут быть побочными продуктами репликации генома.

2. ЭКСПРЕССИЯ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ

2.1. Этапы процессинга генетической информации

Геном является хранителем биологической информации. Однако он сам по себе не способен реализовать эту информацию. Использование биологической информации требует координированной работы ферментов и других белков, которые реализуют сложный комплекс биохимических реакций, называемых экспрессией генома. Начальным продуктом экспрессии генома является транскриптом – коллекция молекул мРНК, считываемых с генов кодирующих белок, чья биологическая информация нужна клетке в тот или иной момент времени. Молекулы мРНК направляют синтез конечных продуктов экспрессии генома, составляющих протеом, то есть репертуар клеточных белков, которые осуществляют биохимические реакции, необходимые клетке для ее существования.

Транскриптом формируется в процессе транскрипции. В ходе транскрипции отдельные гены копируются с образованием молекул РНК. В построении протеома важную роль играют процессы трансляции молекул мРНК в белки. Транскрипция и трансляция являются принципиальными событиями, однако их недостаточно для эффективной работы генома. Предложенная ранее схема экспрессии генома является чрезвычайно упрощенной. В настоящее время схема экспрессии генома выглядит намного сложнее и включает следующие этапы:

- Обеспечение доступности генома к экспрессии (Accessing the genome). Он включает в себя различные процессы, которые воздействуют на структуру хроматина,

расположение нуклеосом (позиционирование) в разных частях генома, содержащих активные гены.

- Сборка комплексов, инициирующих транскрипцию. Любой такой инициирующий комплекс включает набор белков, работающих совместно при копировании генов в РНК. Сборка инициирующего комплекса - высоко упорядоченный, целенаправленный процесс. Иницирующие транскрипцию комплексы собираются в конкретных участках генома для активации конкретных генов.

- Синтез РНК, в ходе которого гены транскрибируются в РНК.

- Процессинг РНК. Включает серию изменений в последовательности молекулы РНК и ее химической структуре. Процессинг происходит перед тем, как молекула РНК транслируется в белок. В случае процессинга некодирующих РНК – перед тем, как они начинают осуществлять свои функции в клетке.

- Деградация РНК. Контролируемый процесс, который играет активную роль в формировании транскриптома и, таким образом, является неотъемлемой составляющей общей картины экспрессии генома.

- Сборка комплекса, инициирующего трансляцию. Происходит на 5'-конце кодирующей молекулы РНК. Сборка обязательна для осуществления трансляции.

- Синтез белка, в ходе которого осуществляется трансляция молекул РНК

- Фолдинг и процессинг белка. Оба процесса могут осуществляться одновременно. Фолдинг приводит к формированию правильной трехмерной структуры белка. При процессинге осуществляется модификация белка путем добавления химических групп, а для некоторых белков – путем удаления одного или нескольких сегментов аминокислотной последовательности.

- Деградация белка. Влияет на формирование протеома и, как и деградация РНК, является важной составляющей процесса экспрессии генома.

Экспрессия генома намного более сложна, чем схема «ДНК делает РНК делает Белок». Недостатком двухэтапной схемы экспрессии генома является то, что в ней не учитываются точки регуляции потока биологической информации, передаваемой от генома к протеому. Для регуляции каждого из этапов, представленных выше, существуют контрольные механизмы, обуславливающие быстрое и точное построение транскриптома и протеома, что позволяет клетке приводить свои биохимические процессы в соответствие с изменениями внеклеточного окружения и сигналами, приходящими от других клеток. Регуляция важна не только для функционирования отдельной клетки, но и для правильной реализации процессов дифференцировки и развития.

2.2. РНК, содержащаяся в клетке

Типичная бактериальная клетка содержит 0,05 – 0,10 пикограмм РНК, что составляет около 6 % от общего веса клетки. Клетка млекопитающего по размерам намного больше бактериальной клетки. Она содержит и значительно больше РНК - 20-30 пг. Но это составляет только 1% от веса клетки. Важно заметить, что не все молекулы РНК составляют транскриптом клетки. Последний состоит только из белок-кодирующих молекул РНК, то есть матричных РНК, транслирующихся в белок. Большинство молекул РНК не попадает в эту категорию, поскольку являются некодирующими. Понимание различий между кодирующими и некодирующими РНК важно для обсуждения вопросов экспрессии генома.

Матричная РНК редко превышает по содержанию 4% от общего количества внутриклеточной РНК. Она является короткоживущей молекулой, то есть быстро разрушается после синтеза. Бактериальная РНК имеет полупериод жизни не более чем несколько минут. Эукариотические РНК деградируют в течение нескольких часов после синтеза. Это означает, что транскриптом не является фиксированным образованием. Транскриптом может быстро меняться, как за счет синтеза разных матричных РНК, так и за счет изменений в скорости синтеза индивидуальных матричных РНК.

Некодирующие РНК более разнообразны, чем кодирующие и имеют множество различных функций, которые осуществляются непосредственно молекулами РНК. Как у эукариот, так и у прокариот имеется два основных типа некодирующих РНК:

-рибосомальные РНК. Составляют более 80 % от общего количества РНК в активно делящейся бактериальной клетке. В соответствии с названием рибосомальные РНК являются составной частью рибосом.

-транспортные РНК – небольшие молекулы, функцией которых является перенос аминокислот на рибосомы.

Рибосомальные и транспортные РНК присутствуют в клетках всех типов. Другие РНК имеют ограниченное распространение. Например, у эукариот имеется множество коротких некодирующих РНК. Их разделяют на несколько категорий:

1. Малая ядерная РНК (snRNA, называемая также U-RNA). Богата уридином. Участвует в процессинге матричной РНК.

2. Малая ядрышковая (нуклеолярная) РНК (snoRNA). Играет центральную роль в процессинге рибосомальной РНК.

3. Малая цитоплазматическая РНК (scRNA). Многочисленная группа, включающая множество разных типов РНК с разными, часто не до конца изученными функциями.

4. МикроРНК интронов (miRNA). Небольшие молекулы РНК размером чуть более 20 нуклеотидов (22 нуклеотида). Кодированы внутри интронов. Выщепляются из состава интронов одновременно с созреванием матричной РНК. Обладают регуляторными свойствами (Young-Kook Kim, N Narry Kim, 2007).

5. Интерферирующая РНК (siRNA). Двунитевая РНК длиной 21-25 пар нуклеотидов. Участвует в механизмах деградации мРНК.

Бактерии и археи также содержат некодирующие РНК, отличающиеся от рибосомальной и транспортной РНК. Но эти молекулы не составляют существенной фракции в общем пуле РНК. У бактерий к ним относится один очень интересный тип некодирующей РНК, присутствующий у многих, если не у всех видов. Это так называемая транспортно-матричная РНК (tmRNA). Она выглядит как транспортная РНК, присоединенная к матричной РНК. Функция ее заключается в том, что она добавляет короткий пептид к белкам, которые некорректно синтезированы, участвуя, таким образом, в процессах деградации белков в бактериальных клетках.

Ферменты, ответственные за транскрипцию ДНК в РНК, называются ДНК-зависимыми РНК-полимеразами. Название означает, что ферментативная реакция, которую они катализируют, приводит к полимеризации РНК из рибонуклеотидов. ДНК при этом выступает в качестве матрицы, с которой копируется нуклеотидная последовательность РНК. Заметим, что имеются и другие РНК-полимеразы. Например, у вирусов присутствуют гены, кодирующие РНК-зависимые РНК-полимеразы. Кроме того, существуют не зависящие от матрицы РНК-полимеразы. Так, для экспрессии генома важное значение имеет поли(А)-полимераза, обеспечивающая нематричный синтез поли(А)-хвостов мРНК в ходе ее созревания (процессинга).

Рибонуклеотиды добавляются по одному к растущему 3'-концу РНК-транскрипта. При добавлении каждого последующего нуклеотида от внедряющегося нуклеотида удаляются бета- и гамма- фосфаты от внедряющегося нуклеотида, а от 3'-нуклеотида растущей цепи удаляется гидроксильная группа. Процесс полимеризации РНК подобен процессу полимеризации ДНК. В основе лежит комплементарность оснований, образующих пары А – Т или U, G – C.

РНК-полимераза является главным компонентом комплекса, инициирующего транскрипцию, который собирается каждый раз при транскрипции гена выше его на нуклеотидной последовательности ДНК собирается такой комплекс. Сборка идет в строго определенных местах, называемых промоторами. У бактерий промоторы прямо узнаются РНК-полимеразой. У эукариот и архей для присоединения РНК-полимеразы необходимы

специальные промежуточные ДНК-связывающие белки, которые образуют платформу для посадки полимеразы.

Наряду со зрелыми молекулами РНК в клетках содержатся молекулы-предшественники. Многие РНК, особенно у эукариот, сначала синтезируются в виде предшественника или пре-РНК, которая должна дозреть, процессироваться перед выполнением своей функции. Существует множество вариантов процессинга:

- концевая модификация. Происходит при синтезе мРНК эукариот и архей. Матричная РНК в большинстве случаев имеет на 5'-конце необычный нуклеотид 7-метилгуанозин, называемый кэпом. На 3'-конце в составе мРНК присутствует поли(А) конец. Кэп и поли(А)-хвост участвуют в сборке комплекса, инициирующего трансляцию мРНК.

- сплайсинг. Представляет собой удаление интронов из РНК-предшественника. Многие белок-кодирующие гены эукариот содержат интроны и они копируются при транскрипции гена. Интроны удаляются из пре-мРНК путем вырезания и сшивания нуклеотидной последовательности. Не прошедшая сплайсинг пре-мРНК образует фракцию внутриклеточной РНК, называемую гетерогенной ядерной РНК (гяРНК, hnRNA). Некоторые эукариотические пре-рибосомальные РНК (пре-рРНК) и пре-транспортные РНК (пре-тРНК) также имеют интроны, как и некоторые транскрипты архей. У бактерий интроны очень редки.

- нарезание РНК. Происходит при созревании рибосомальных и транспортных РНК, многие из которых синтезируются из транскрипционных единиц, определяющих строение нескольких молекул. Пре-рРНК и пре-тРНК нарезаются на фрагменты, которыми являются зрелые молекулы рРНК и тРНК. Такой тип процессинга происходит как у прокариот, так и у эукариот.

- химическая модификация. Осуществляется при созревании рибосомальных, транспортных и матричных РНК. Рибосомальные и транспортные РНК всех организмов модифицируются путем добавления новых химических групп. Эти группы добавляются к специальным нуклеотидам внутри каждой РНК. Химическая модификация матричной РНК называется РНК-редактированием. Наблюдается у различных групп эукариот.

Процессинг мРНК очень сильно влияет на построение транскрипта. Например, редактирование РНК может приводить к превращению одной и той же молекулы пре-мРНК в две разные молекулы мРНК, кодирующие совершенно различные белки. Этот способ модификации не является общим механизмом формирования разнообразия мРНК. Наиболее распространен другой вариант процессинга - альтернативный сплайсинг. С помощью альтернативного сплайсинга пре-мРНК может образоваться несколько молекул

зрелых мРНК, возникающих вследствие комбинации экзонов. Матричные РНК, возникающие вследствие редактирования и альтернативного сплайсинга, часто имеют тканевую специфичность. Процессинг увеличивает кодирующую емкость генома, повышая количество образующихся белковых продуктов без увеличения числа генов.

Хотя транскриптом занимает менее 4 % от общего количества РНК, тем не менее, это наиболее важный компонент среди молекул РНК, поскольку содержит кодирующие РНК, которые определяют построение протеома и, тем самым, биохимические возможности клетки. Важно отметить, что транскриптом никогда не синтезируется *de novo*. Каждая клетка получает при делении часть родительского транскриптома, а затем поддерживает его на протяжении всей своей жизни, изменяя в соответствии с необходимостью. Даже споры бактерий и семена растений имеют транскриптом, хотя экспрессия транскриптома в протеом может быть полностью выключена. То есть, транскрипция поддерживает транскриптом, заменяя деградирующие мРНК на новые, осуществляя тем самым постоянный обмен компонентов транскриптома путем включения или выключения тех или иных генов.

Даже у наиболее простых организмов, таких как дрожжи или бактерии, разные гены активируются в разное время. Транскриптом содержит множество копий сотен и тысяч различных мРНК. Обычно каждая мРНК представляет очень небольшую часть транскриптома и редко составляет более 1 % от общей фракции молекул мРНК. Исключения известны лишь для клеток, у которых идут высоко специализированные биохимические процессы. Наличие таких процессов отражается на строении транскриптома, так как некоторые виды мРНК могут значительно преобладать над другими. Примером могут служить семена пшеницы. Они синтезируют большие количества белка глиаина, который аккумулируется в гранулах и является источником аминокислот для развивающегося зародыша. В прорастающих семенах мРНК глиаина составляет более 30 % от транскриптома клетки.

В настоящее время строение транскриптома можно изучать с помощью биочипов, позволяющих регистрировать существующий в клетке набор матричных РНК. С использованием биочипов проведено сравнение транскриптомов человека и дрожжей.

Дрожжи, имеющие менее 6 тысяч генов, являются идеальной моделью для изучения строения транскриптома. Многие пионерские проекты по изучению транскриптома были выполнены с использованием *Saccharomyces cerevisiae*. В-первых же наблюдениях было обнаружено, что хотя мРНК дрожжей постоянно деградируют и ресинтезируются, тем не менее, состав дрожжевого транскриптома подвергается очень небольшим изменениям, если условия культивирования дрожжей постоянны. Так, если

дрожжи растут на богатой среде, позволяющей клеткам делиться с максимальной скоростью, транскриптом достаточно стабилен. Только 19 мРНК изменяют свое содержание более чем вдвое за два часа. Значительные изменения в строении транскриптома происходят только тогда, когда среда истощается, заставляя клетки переключаться с анаэробного на анаэробное дыхание. В ходе такого переключения увеличивается содержание более чем 700 различных мРНК и уменьшается содержание более 1000 мРНК. То есть, изменение окружающих условий приводит к значительной перестройке транскриптома клетки.

Дрожжевой транскриптом подвергается серьезной реструктуризации также в ходе дифференцировки дрожжевых клеток. Это было обнаружено при изучении споруляции дрожжей, индуцированной стрессирующими факторами. Споруляция делится на 4 стадии – раннюю, среднюю, средне-позднюю и позднюю. Такое деление проведено на основании изучения морфологических и биохимических характеристик. Каждая стадия характеризуется экспрессией разных наборов генов. Изучение изменений в транскриптоме позволило разделить эти стадии на подстадии. Изменения, которые были выявлены в транскриптоме, показали, что ранняя стадия может быть разделена на три подстадии. На ранней стадии увеличивалось содержание более, чем 250 разных мРНК, содержание других 158 мРНК повышалось в ходе средней стадии, последующее повышение еще 61 различных мРНК обнаруживалось на средне-поздней стадии и, наконец, на поздней стадии повышалось содержание еще 66 генов. Всего, таким образом, менялось содержание почти 600 разных мРНК, кодирующих различные белки, которые вероятно необходимы для образования спор.

Описанная работа важна в двух отношениях. Во-первых, она открыла путь к изучению взаимоотношений между геномом и окружающими клетку сигналами. Во-вторых, некоторые мРНК, содержание которых менялось в ходе споруляции, являются транскриптами генов, чьи функции пока не известны. Изучение транскриптома поможет в аннотации геномных последовательностей, в идентификации генов, чья роль в геноме еще не определена.

Поскольку геном человека содержит в несколько раз больше генов, чем геном дрожжей, то и человеческий транскриптом должен быть намного сложнее. Он активно изучается. Получены данные об особенностях его строения в разных клетках, на разных стадиях их дифференцировки, при различных заболеваниях. Например, проведено картирование транскриптомов восьми различных типов клеток, что привело к получению глобальной картины экспрессии генов. Показано, что транскриптом раковых клеток значительно отличается от транскриптома нормальных клеток.

Данные о реструктуризации транскриптома раковых клеток впервые были получены на примере рака прямой кишки. Было обнаружено, что клетки эпителия нормальной прямой кишки отличаются от злокачественно трансформированных гомологов значительно повышенным содержанием 289 матричных РНК. Содержание более половины из этих РНК увеличивается при раке поджелудочной железы. Хочется надеяться, что понимание различий между транскриптомами нормальных и опухолевых клеток откроет новые перспективы лечения онкологических заболеваний.

Изучение транскриптома имеет значение и для совершенствования методов диагностики онкологических заболеваний. В 1999 году было показано, что транскриптом клеток острой лимфобластной лейкемии отличается от транскриптома клеток острой миелобластной лейкемии (Golub, 1999). Были изучены транскриптомы опухолевых клеток 27 больных лимфобластной и миелобластной лейкемией. Обнаружено, что, хотя транскриптомы имеют небольшие различия у больных с одним и тем же типом опухоли, тем не менее, существуют достаточно явные различия между транскриптомами двух типов опухолевых клеток. Практическое значение этой работы заключается в том, что с помощью данного подхода можно определять злокачественную трансформацию на ранних стадиях онкологического процесса до появления морфологических признаков заболевания.

2.3. Протеом клетки

Протеом является конечным продуктом экспрессии генома и включает все белки, представленные в клетке в настоящий момент времени и на данной стадии развития. Типичная клетка млекопитающего, например, гепатоцит, содержит 10-20 тысяч различных белков, около 8×10^9 индивидуальных молекул, составляя примерно 0,5 нг белка или 18-20 % от общего веса клетки. Количество копий белковых молекул в клетке может быть самым разным: от менее чем 20 тысяч белковых молекул до 100 миллионов копий.

Белки, которые присутствуют в клетке в количестве более чем 50 000 копий, считаются относительно многочисленными, и свыше 2000 белков попадают в эту категорию. Когда сравнили протеомы различных типов клеток млекопитающих, обнаружили, что различия в содержании этих белков сравнительно невелики. Это свидетельствует об их принадлежности к белкам, реализующим основные биохимические реакции, важные для жизни клетки (по аналогии с генами их можно отнести к белкам домашнего хозяйства). Белки, обеспечивающие специфические клеточные функции, могут

находиться в больших количествах, например, белки секреторных гранул, гемоглобин и другие.

Протеом может рассматриваться как центральный связующий элемент между геномом и клеткой. Он, с одной стороны, является кульминацией экспрессии генома, а, с другой стороны, является стартовой точкой для осуществления биохимических реакций, составляющих клеточную жизнь.

Поток информации от ДНК к РНК путем транскрипции не представляет собой концептуальной сложности. ДНК и РНК являются полинуклеотидами, имеющими очень сходную структуру. Поэтому, легко можно понять каким образом РНК копирует ген с помощью матричного синтеза, используя пары комплементарных оснований. Последующие стадии экспрессии генома, в ходе которых одна структура обеспечивает синтез другой структуры (направление синтеза белка с помощью РНК), понять сложнее.

В начале 50-х годов сразу после открытия строения ДНК несколько молекулярных биологов предприняли попытки расшифровать пути, с помощью которых четырехнуклеотидный триплетный код определяет синтез белков, состоящих из 20 аминокислот. В 1957 году Ф. Крик предположил, что матричная РНК – это адаптерная молекула, связывающаяся с синтезируемым полипептидом. Далее было установлено, что адаптерными молекулами являются не матричные РНК, а некодирующие транспортные РНК. Сразу после этого были быстро сформированы представления о матричном синтезе белка на рибосомах с участием разных типов РНК.

Вторая важная проблема, которая стояла перед молекулярными биологами – это информационная проблема, связанная с пониманием генетического кода, лежащего в основе механизма трансляции, а следовательно, играющего важную роль в понимании связи между транскриптомом и протеомом. Здесь мы не будем останавливаться на свойствах генетического кода, поскольку они описаны во многих учебниках и пособиях по биохимии и молекулярной биологии. Обратимся к следующему уровню информационного процессинга – некоторым вопросам, касающимся связи между протеомом и биохимическими процессами, происходящими в клетке.

Хорошо известно, что белки имеют четыре уровня организации: от первичной, то есть аминокислотной последовательности, до четвертичной, отвечающей за формирование крупных биологических структур. И здесь важна иерархическая природа этих четырех видов структуры, определяющая ключевое значение первичной структуры для формирования остальных уровней организации. Второй аспект заключается в множественности химических свойств белков, выполняющих в клетке самые разнообразные функции.

Связь между различными уровнями организации белковой молекулы особенно хорошо заметна при рассмотрении взаимоотношений между первичной и вторичной структурами. Из-за различий в строении радикалов некоторые аминокислоты чаще встречаются в составе альфа-цепей, другие в составе бета-структур. Это обусловлено различиями в химическом строении радикалов аминокислот. Образование вторичной структуры происходит в несколько этапов: нуклеация, рост и завершение роста. Нуклеация с последующими этапами происходит по всей полипептидной цепи, и если физико-химические свойства какой-либо аминокислоты не позволяют ей участвовать в формировании альфа-цепи или бета-структуры, то формирование структуры тормозится или она совсем не образуется.

Зная, какие именно аминокислоты способны участвовать в формировании вторичных структур, биохимики способны предсказать вторичную структуру белка с достаточно большой достоверностью. В настоящее время существует программное обеспечение, достаточно легко выполняющие подобные задачи. Сложнее предсказать третичную и четвертичную структуры, что связано с множеством проблем, поскольку на этом уровне значительное влияние оказывают условия микроокружения, внешние воздействия. В частности, в формировании структуры одних белков могут принимать участие другие белки (шапероны).

Как указано выше, белки выполняют множество функций. Биологическая информация, заложенная в геноме, находит свое конечное выражение в том, какую функцию и каким образом выполняет тот или иной белок. В зависимости от множества факторов протеом может быть самым разным, поддерживая самые разные проявления жизни.

Давно известно, что белки выполняют следующие основные функции:

- Биохимический катализ. Его выполняют белки, называемые ферментами или энзимами. Центральные энергетические пути, процессы биосинтеза и распада низкомолекулярных соединений, самих белков, нуклеиновых кислот, жирных кислот катализируются множеством ферментов, которые способны нередко проявлять несколько каталитических функций, в том числе за счет того, что построены из доменов, обладающих разными функциональными активностями. Биохимический катализ лежит в основе экспрессии самого генома, например, в экспрессии генома на стадии транскрипции участвуют РНК-полимеразы.

- Структурная роль. Хорошо известно, что структура клетки, то есть ее морфологические свойства, определяются в первую очередь белками. Из белков построен

цитоскелет, из белков построен межклеточный матрикс, клеточные органеллы и так далее.

- Двигательная функция. Осуществляется двигательными белками, среди которых наиболее известны миозин, актин, динеин, флагеллин.

- Транспортная функция. Весьма существенна для перемещения молекул как внутри клетки, так и во внеклеточном пространстве, в том числе и в плазме крови. Приведем лишь два общеизвестных примера. Всем известно, что гемоглобин переносит кислород. Наряду с гемоглобином, являющимся внутриклеточным транспортным белком, в плазме крови имеется внеклеточный транспортный белок. Это альбумин, переносящий жирные кислоты и целый ряд других биологически активных соединений.

- Регуляторные функции. Внутриклеточные процессы регулируются каскадами сигнальных белков и белками, воздействующими на экспрессию генома, участвующими во включении или выключении и тех или иных генов Широко известны внеклеточные регуляторные белки – это например гормоны, цитокины и так далее.

- Защитная функция. К этой группе белков в первую очередь относят разнообразные белки, секретируемые клетками иммунной системы - антитела, являющиеся по своей химической природе иммуноглобулинами. Важная роль принадлежит белкам свертывания крови, целому ряду других протеинов, обеспечивающих сохранение живого организма.

- Запасающая функция. Осуществляется такими белками как ферритин, который запасает ионы железа в печени, глиадины, которые являются хранилищем аминокислот в семенах растений.

Множественность функций, выполняемых белками, позволяет протеому успешно конвертировать информацию, содержащуюся в геноме, в протекающие сложным образом жизненные процессы.

2.4. Механизмы экспрессии генома

2.4.1. Обеспечение доступности генома

Для того, чтобы клетка использовала биологическую информацию, содержащуюся в геноме, группы генов, а также отдельные гены, несущие единицы информации, должны функционировать координировано. Высокая скоординированность генной экспрессии достигается при участии транскриптома, который определяет природу протеома, который, в свою очередь, определяет жизненные проявления клетки и организма в целом.

Когда рассматривают геномные последовательности, записанные в виде серии знаков (A, T, G, C), произвольно подразумевают, что вся последовательность доступна

для считывания. На самом деле такой взгляд на ситуацию очень далек от действительности. ДНК в ядре эукариотических клеток или в нуклеоидах прокариот присоединена к множеству различных белков, которые непрямым образом участвуют в экспрессии генома. Чтобы РНК-полимераза и другие, необходимые для считывания генов белки получили доступ к генам, нужно чтобы ДНК была свободна в этих участках от других белков, закрывающих ген, когда он находится в молчащем состоянии. То есть последние белки должны отсоединиться от ДНК или переместиться в другие участки.

О механизмах обеспечения доступности генов у прокариот в настоящее время известно немного. Однако, для эукариот получено достаточно данных для формирования начальных представлений об упаковке генов в хроматине и влиянии этой упаковки на экспрессию генов. Эта актуальная в настоящее время область молекулярно-биологических исследований показывает, что гистоны и другие упаковочные хромосомные белки являются не просто инертными структурами, окружающими ДНК. Они выступают активными участниками процессов, определяющих экспрессию того или иного участка генома в конкретной клетке. Многие прорывные направления в этой области связаны с появлением новых данных о субструктуре ядра.

Предшествующие исследования сформировали представление о ядре, как гомогенной по архитектуре структуре, как о черном ящике. В настоящее время ясно, что ядро – это сложная внутриклеточная структура, в которой осуществляются множество биохимических реакций. При этом ядро не разделено мембранами как цитоплазма, что еще более усложняет происходящие в нем события.

Сложность ядерного матрикса послужила основанием для того, чтобы составить представление о нем как о структуре, в которой движение белковых молекул ограничено. Тем не менее, последние исследования с использованием флуоресцентного анализа показали, что белки внутри ядра способны перемещаться достаточно активно. Обнаружено, что белки могут проходить через поры в ядро или обратно в цитоплазму за несколько минут. Белки, участвующие в экспрессии генома имеют свободу передвижения, активно перемещаясь из одного участка в другой. Этот факт особенно важен с позиции образования ДНК-белковых комплексов, которые обеспечивают экспрессию генов.

Ранее мы рассматривали структуру хроматина, его многоуровневую иерархию строения, начиная от упаковки в нуклеосомы и 30-нанометровые хроматиновые нити, до метафазных хромосом, возникающих при делении клеток. После деления хромосомы опять расплетаются и возникают участки, названные гетерохроматином и эухроматином. Гетерохроматин локализован в основном по периферии ядра и содержит ДНК, которая находится в относительно более компактной организации, чем эухроматин.

Различают два типа гетерохроматина:

- конститутивный гетерохроматин. Он постоянно присутствует во всех клетках и представляет собой ДНК, которая не содержит генов и поэтому все время находится в компактном состоянии. Эта фракция ДНК содержит центромерную и теломерную ДНК, а также некоторые другие участки хромосом. Например, большая часть человеческой Y хромосомы построена из конститутивного гетерохроматина.

- факультативный гетерохроматин. Это не постоянная структура, появляется в разных частях ядра в разное время. Факультативный гетерохроматин содержит гены, которые неактивны в определенные периоды жизни клетки или совсем не активны в данном типе клеток. ДНК таких неактивных генов и окружающие участки компактизируются в факультативный гетерохроматин.

Полученные данные предполагают, что организация гетерохроматина компакта настолько, что белки, участвующие в экспрессии генов, просто не могут достигнуть этого участка генома. И наоборот, остальные регионы хромосомной ДНК, часть которых содержат активные гены, являются менее компактными, что обеспечивает доступность этих участков для белков, принимающих участие в активации генов. Эти участки называются эухроматином. Они разбросаны по всему ядру. Детальная организация ДНК внутри эухроматина не известна, но в электронный микроскоп внутри регионов эухроматина можно видеть петли ДНК. Каждая петля имеет длину от 40 до 100 килобаз и преимущественно находится в форме 30-нанометровых нитей. Петли присоединены к ядерному матриксу через АТ-богатые сегменты ДНК, называемые матрикс-ассоциированными регионами (matrix-associated regions (MARs) or scaffold attachment regions (SARs)).

Петли ДНК между точками прикрепления к ядерному матриксу называются структурными доменами. Нерешенным остается вопрос о соответствии структурных доменов хроматина его функциональным доменам. Функциональные домены могут быть идентифицированы путем обработки очищенного хроматина ДНК-азой I, которая расщепляет экспрессирующиеся гены, поскольку они имеют более открытую, доступную для белков, и соответственно, нуклеаз, структурную организацию. Показано, что некоторые MARs, у которых маркированы границы структурных доменов, локализованы в тех же участках, что и функциональные домены. Но, соответствие не всегда является полным, поскольку некоторые структурные домены содержат функционально не связанные между собой гены.

Границы функциональных доменов маркируются последовательностями длиной 1-2 килобазы, названными инсуляторами. Инсуляторные последовательности впервые

были открыты у дрозофил. Сейчас они идентифицированы у широкого круга видов эукариот. Лучше всего изучены две последовательности дрозофилл, названные *scs* и *scs'* (*scs* - specialized chromatin structure) и локализованные на краях двух генов *Hsp70*.

Инсуляторы обладают двумя свойствами, имеющими отношение к их роли разделителей функциональных доменов. Первое свойство – это способность преодолевать так называемый позиционный эффект. Он обнаружен при проведении экспериментов по клонированию генов у эукариот и связан с вариабельностью генной экспрессии, возникающей после того, как новый ген внедряется в структуру эукариотической хромосомы. Позиционный эффект - это результат неупорядоченного внедрения генов. При попадании в регион с высокоупакованным хроматином ген становится неактивным, если ген попадает в регион открытого хроматина, то высока вероятность его экспрессии. Способность гасить позиционный эффект продемонстрирована путем помещения инсуляторов *scs* на края генов окраски глаз у дрозофил. Когда такие гены фланкированы инсуляторами, то при помещении их в геном дрозофилы они экспрессируются с высокой долей вероятности в отличие от вариабельной экспрессии генов, клонированных без использования инсуляторов. Вывод из этих и других подобных экспериментов состоит в том, что искусственно введенные в геном инсуляторы вызывают модификации в структуре хроматина и, внедряясь в новые участки генома, формируют новые функциональные домены.

Поддерживая независимость каждого функционального домена, инсуляторы предотвращают перекрытия между соседними доменами. Если вышеприведенные инсуляторы *scs* или *scs'*, находящиеся в нормальном положении, перенести в другой участок, разделяющий тело гена и вышележащий регуляторный участок, то ген становится изолированным от его регуляторного участка. Это свидетельствует о том, что в нормальной позиции инсулятор предохраняет ген, находящийся внутри домена, от действия регуляторных участков, находящихся в соседних доменах.

Как инсуляторы осуществляют эту роль пока остается неизвестным, но предполагается, что функциональными компонентами являются не изолирующие последовательности сами по себе, а ДНК-связывающие белки, такие как *SU (Hw)* у дрозофил, специфически присоединяющийся к инсуляторам. Связываясь с инсуляторами, такие белки, кроме того, образуют ассоциаты с ядерным матриксом, показывая, что функциональные домены могут определять и структурные домены внутри хроматина. Однако последнее утверждение нельзя считать строго доказанным.

Некоторые функциональные домены содержат участки нуклеотидной последовательности, названные регионами, контролирующими локус (*LCR*). Они

контролируют структуру хроматина внутри функциональных доменов. Образование и поддержание открытого функционального домена, по крайней мере в ряде случаев связано с такими последовательностями ДНК. То есть, LCR стимулируют экспрессию генов, находящихся внутри функциональных доменов. Как и инсуляторы, LCR могут вызывать позиционные эффекты при внедрении какого-либо гена в эукариотическую хромосому.

LCR были открыты при изучении генов человеческих бета-глобинов, и, как стало известно в настоящее время участвуют в регуляции экспрессии многих генов, по разному активных в различных тканях и на разных стадиях развития. LCR бета-глобинов имеет длину 12 килобаз, расположен выше генов в функциональном домене бета-глобинов, длиной 60 килобаз. Многие случаи заболевания талассемией являются результатом мутаций в кодирующих участках генов глобинов, но некоторые картированы в регионе, соответствующем LCR. Способность мутаций в последовательности, соответствующей LCR, вызывать талассемию является индикатором того, что нарушения в этом регионе приводят к потере экспрессии генов бета-глобинов.

Более детальное изучение LCR бета-глобинов показало, что он содержит пять отдельных гиперчувствительных к ДНК-азе I сайтов – коротких участков ДНК, которые расщепляются ДНК-азой I намного легче, чем другие участки функционального домена. Эти сайты совпадают с позициями, где нуклеосомы модифицированы или отсутствуют и с позициями, доступными для связывания с белками. Эти белки, а не сама по себе последовательность ДНК в LCR, контролируют структуру хроматина внутри функциональных доменов. Однако, каким образом это происходит, остается неизвестным.

Сайты, гиперчувствительные к действию ДНК-азы I, находятся также непосредственно перед каждым из генов бета-глобинов. Эти сайты соответствуют тем позициям, в которых собирается иницирующий транскрипцию комплекс. Места сборки иницирующего комплекса иллюстрируют интересные особенности сайтов гиперчувствительности. Они не являются инвариантными компонентами функциональных доменов. Заметим, что различные гены бета-глобинов экспрессируются на разных стадиях развития. Глобин ϵ активен на ранних эмбриональных стадиях развития, $G\gamma$ и $A\gamma$ – это фетальные белки, δ и β глобины появляются у взрослых. Гиперчувствительные сайты обнаруживаются, когда активируется какой-либо ген из семейства бета-глобинов. С этим связывают эффект дифференциальной экспрессии данных генов. Другими словами отсутствие генной активности связывается с тем, что нуклеосомы закрывают сайт сборки иницирующего комплекса. То есть присутствие или

отсутствие нуклеосом в данном участке является причиной различий в экспрессии генов. Гены выключаются, когда нуклеосомы закрывают сайт сборки комплекса, инициирующего транскрипцию, и включаются при открытии этого сайта.

Этот пример хорошо иллюстрирует, каким образом доступность биологической информации для считывания может зависеть от упаковки ДНК в хроматине. Структура хроматина влияет на экспрессию генома двумя способами. Во-первых, степень упаковки хроматина в разных участках хромосом определяет активность генов. Во-вторых, когда ген находится в доступном состоянии для активации, тогда его активность зависит от тонкой структуры и позиций нуклеосом. Если нуклеосомы находятся в участках, соответствующих сайтам, где собираются комплексы, инициирующие транскрипцию, то ген не активен.

Нуклеосомы являются наиболее важными структурами, определяющими активность генома у эукариот не только потому, что способны занимать различные позиции по отношению к молекуле ДНК, но и потому, что тонкая химическая структура гистоновых белков, составляющих нуклеосому, также влияет на степень упаковки разных сегментов хроматина.

Так, структура хроматина зависит от модификаций гистонов. Гистоны могут подвергаться разным типам модификаций. Лучше всего изучено ацетилирование гистонов, то есть присоединение ацетильных групп к остаткам лизина на N-конце каждого корового гистона. N-концы образуют хвосты, которые выступают за пределы нуклеосомного корового октамера. Ацетилирование снижает аффинность гистонов к ДНК и понижает уровень межнуклеосомного взаимодействия, что приводит к образованию 30-нанометровых хроматиновых нитей. Гистоны в гетерохроматине обычно неацетилированы. В то же время, в функциональных доменах они ацетилированы, показывая, что этот тип модификации связан с упаковкой ДНК.

Значение ацетилирования гистонов для экспрессии генома было выявлено в 1996 году, когда после нескольких лет исследований были идентифицированы гистон-ацетилтрансферазы (НАТ) – ферменты, которые катализируют добавление ацетильных групп к гистонам (Pennisi, 1997). Открытие этих ферментов, а также исследование разных паттернов ацетилирования гистонов в разных типах клеток привели к заключению, что ацетилирование гистонов играет существенную роль в регуляции экспрессии генов.

Отдельные гистон ацетилтрансферазы ацетилируют гистоны в пробирке, но не проявляют активности на интактных нуклеосомах, показывая, что в ядре НАТ почти всегда не работают в одиночку. Они образуют мультибелковые комплексы, такие как комплексы ADA и SAGA у дрожжей и комплекс TFTC у человека. Различные комплексы

ацетируют разные гистоны, а некоторые могут также ацетилировать другие белки, участвующие в экспрессии генома, такие как факторы транскрипции TFIIIE, TFIIIF. Имеются также данные о том, что в дополнение к локальным модификациям гистонов в участках, окружающих экспрессирующиеся гены, НАТ может также осуществлять более общие модификации глобального масштаба в объемах сего генома (Berger, 2000).

Ацетилирование – не единственный тип модификации гистонов. Концы коровых гистонов также имеют сайты присоединения метильных и фосфатных групп и сайты присоединения убиквитина (белка, участвующего в инициации деградации белков и в регуляции клеточного протеома). Хотя информация об этих модификациях довольно ограничена, тем не менее ясно, что они также влияют на структуру хроматина и имеют значительное влияние на клеточную активность. Например, фосфорилирование гистона H3 и линкерных гистонов ассоциировано с образованием метафазных хромосом, убиквитилирование гистона H2B является звеном в механизме, с помощью которого убиквитин контролирует клеточный цикл (Robzyk et al., 2000).

Эффект метилирования двух остатков лизина в 4 и 9 позициях от N-конца гистона H3 особенно интересен. Метилирование лизина в девятой позиции приводит к образованию сайта связывания белка HP1, который индуцирует упаковку хроматина и снижение генной экспрессии (Lachner et al., 2001, Vannister et al., 2001). Метилирование лизина в четвертой позиции вызывает противоположный эффект и вызывает формирование открытой структуры хроматина. В функциональном домене бета-глобинов метилирование лизина в четвертой позиции тесно связано с ацетилированием гистона H3 (Litt et al., 2001). То есть, оба типа модификации очень тесно работают при активации регионов хроматина. В настоящее время предложено понятие «гистоновый код», в соответствии с которым существует паттерн химических модификаций гистоновых белков. Эти модификации определяют порядок экспрессии разных участков генома в разных клетках в разное время в соответствии с различными внеклеточными воздействиями.

Вторым типом модификации хроматина, который может влиять на экспрессию генома, является так называемое ремоделирование нуклеосом (nucleosome remodeling). Этот термин отражает модификацию или изменение положения нуклеосом (репозиционирование) в пределах короткого участка генома. В результате ремоделирования ДНК-связывающие белки могут получать доступ к местам их потенциального расположения на молекуле ДНК. В отличие от ацетилирования и других химических модификаций, описанных выше, ремоделирование нуклеосом не сопровождается присоединением каких-либо остатков к гистоновым белкам.

Ремоделирование индуцируется энерго-зависимыми процессами, которые приводят к ослаблению контактов между отдельными нуклеосомами и между нуклеосомами и ДНК, с которой они связаны. Известно три типа изменений, называемых ремоделированием:

- ремоделирование в прямом смысле слова, представляющее собой изменение структуры нуклеосом без изменения их позиции по отношению к ДНК. Природа структурных изменений неизвестна, однако в условиях *in vitro* показано, что нуклеосомы удваиваются в размерах, ДНК в этих участках становится более доступной к действию ДНК-аз;

- слайдинг или цис-перемещение – физическое движение нуклеосом вдоль ДНК;

- трансфер или транс-перемещение – перенос нуклеосом на вторую молекулу ДНК или на отдаленную часть той же молекулы.

Как и гистонацетилтрансферазы, белки, ответственные за ремоделирование нуклеосом, выполняют свои функции совместно с другими белками в составе больших комплексов. Один из них Swi/Snf построен по крайней мере из 11 белков, представленных у многих эукариот (Sudarsanam, Winston, 2000). В настоящее время исследуются механизмы действия ремодулирующих комплексов. Известно, что ни один из компонентов комплекса Swi/Snf не обладает способностью связываться с ДНК. То есть, этой функцией должны обладать дополнительные белки, взаимодействующие с комплексом.

Обнаружено взаимодействие между комплексом Swi/Snf и НАТ (Synichaki et al., 2000). Это предполагает, что ремоделирование нуклеосом может происходить параллельно с ацетилированием гистонов. Это весьма привлекательная гипотеза, поскольку она связывает воедино два процесса, играющие центральную роль в активации генома. Однако в этой гипотезе существуют несоответствия, поскольку комплекс Swi/Snf не обладает глобальным воздействием на целый геном, влияя только на ограниченное число генов. Например, у дрожжей он влияет на экспрессию только 6 % генома. Более важным представляется взаимодействие этого комплекса с другими белками, которые соединяются с ограниченным числом генов, в отличие от НАТ, функционирующим на протяжении всего генома. Наиболее вероятными кандидатами являются активаторы транскрипции, специфические для конкретных групп генов и одновременно взаимодействующие с комплексом Swi/Snf.

Поскольку существуют механизмы активации генома, значит должны существовать и комплементарные им механизмы торможения активности генома, то есть механизмы, гасящие активность генов, переводящие их в неактивное, молчащее состояние. Перевод генов в молчащее состояние называют сайленсингом (умолканием).

Одним из механизмов сайленсинга является удаление ацетильных групп с гистоновых хвостов, что приводит к торможению активационного эффекта, осуществляемого с помощью НАТ. Торможение осуществляется ферментами, называемыми гистондеацетилазами (НАД). Связь между активностью этих ферментов и сайленсингом генов была установлена в 1996 году. Тогда был открыт первый из таких ферментов, получивший сокращенное название HDAC1.

Подобно НАТ и ремоделирующим ферментам, гистон-деацетилирующие ферменты также функционируют не как отдельные молекулы, а в составе мультибелковых комплексов. Одним из таких комплексов является комплекс Sin3 млекопитающих, состоящий из семи белков, включая как гистондеацетилирующие ферменты, так и вспомогательные белки, способные связываться с гистонами и с метилированной ДНК.

Не только модификация хроматина может приводить к сайленсингу генома. Метилирование ДНК также угнетает экспрессию генов. У эукариот хромосомная ДНК метилируется по остаткам цитозина. При этом образуется метилцитозин. Метилирование катализируется ДНК метилтрансферазами. Метилирование цитозина довольно редко наблюдается у низших эукариот. В то же время у позвоночных до 10 % остатков цитозина в составе хромосомной ДНК метилируется. В геноме растений эта цифра достигает 30 %.

Метилирование происходит не беспорядочно. Оно ограничено специфическими последовательностями. Чаще всего метилирование наблюдается в последовательности 5'-CG-3' (CpG-островки). У растений чаще всего метилируется последовательность 5'-CNG-3', где N – любой нуклеотид. Различают два типа метилирования. Первый называют поддерживающим метилированием. Оно осуществляется при репликации генома, отвечает за добавление метильных групп к остаткам цитозина на вновь синтезирующуюся молекулу ДНК в позициях, противоположных сайтам метилирования материнской нити ДНК. Поддерживающая активность обеспечивает паттерн метилирования дочерних нитей ДНК, подобный родительским нитям.

Второй тип метилирования – это метилирование *de novo*, при котором метильные группы добавляются к остаткам цитозина так, что меняется паттерн метилирования в локальных регионах генома. Обнаружены метилтрансферазы, отвечающие в основном за поддерживающее метилирование (Dnmt1) и метилтрансферазы, отвечающие в основном за метилирование *de novo* (Dnmt3a, Dnmtb). Показано, что активные гены локализованы в неметилированных участках генома. У человека 40-50 % всех генов локализовано близко к CpG островкам, отражающим активационный статус близлежащих генов. Гены домашнего хозяйства, то есть те, которые экспрессируются во всех клетках и необходимы для поддержания жизни клетки, имеют неметилированные CpG островки, в то время как

тканеспецифические гены не метилированы только в тех тканях, в которых эти гены экспрессируются. ДНК активно пролиферирующих клеток мало метилирована. Покоящиеся клетки характеризуются более высоким уровнем метилирования ДНК.

Ингибирующее действие генов метилирования реализуется через метил-CpG-связывающие белки, которые являются компонентами гистондеацетилирующих комплексов (Sin3). В соответствии с существующей моделью CpG островки являются мишенью для присоединения таких комплексов, что модифицирует окружающий хроматин и приводит к выключению генов.

Обнаружена связь между метилированием ДНК и двумя интригующими феноменами, называемыми геномным импринтингом и X-инактивацией.

Геномный импринтинг является важной чертой генома млекопитающих, в которых экспрессируется только один из пары генов, присутствующих на гомологичных хромосомах в диплоидном ядре. Второй ген молчит, поскольку метилирован. Третий ген человека и мыши подвержено импринтингу. Функция импринтинга исследуется. Предполагается, что он играет важную роль в развитии организма и участвует в реализации механизмов, связанных с особенностями экспрессии генома у животных разного пола.

X-инактивация менее загадочна. Это специальная форма импринтинга, которая приводит к инактивации одной из X-хромосом у самок млекопитающих. Существование X-инактивации объясняют тем, что самки имеют две X-хромосомы, в то время как самцы только одну. Если обе X-хромосомы самок будут активны, то белки, кодируемые генами этой хромосомы, будут синтезироваться во вдвое большем количестве, чем у самцов. Чтобы устранить это несоответствие одна хромосома молчит. Сайленсинг включается на ранних стадиях эмбрионального развития и контролируется X-инактивирующим центром, представляющим собой конкретный участок, присутствующий на каждой X хромосоме.

В клетках, подвергнутых X-инактивации, инактивирующий центр на одной из X хромосом инициирует образование гетерохроматина. Процесс распространяется затем на всю хромосому за исключением очень коротких сегментов, содержащих небольшие кластеры генов, остающихся активными. Более детальный механизм в настоящее время изучается. Так, обнаружено, что ген, названный Xist и локализованный в инактивирующем центре, транскрибируется в некодирующуюся РНК длиной 25 килобаз. При формировании гетерохроматина копии РНК покрывают хромосому. При X-инактивации гистон H2A, один из составляющих корового октамера нуклеосомы, замещается специальным гистоном макроH2A1. В гетерохроматине обычно наблюдается

также деацетилирование гистона H4 и происходит гиперметилование некоторых последовательностей ДНК.

2.4.2. Инициация транскрипции.

Совсем недавно первую стадию экспрессии генома описывали как транскрипцию ДНК в РНК. В настоящее время считается признанным, что процесс, который приводит к транскрипции генома, намного более сложен, чем просто синтез РНК. В предыдущем разделе мы рассмотрели первую стадию экспрессии генома, которая определяет возможность активации того или иного гена. Если такая возможность реализуется, то наступает следующая стадия экспрессии генома – сборка комплекса, инициирующего транскрипцию. В состав такого комплекса входит РНК-полимераза и различные вспомогательные белки. Смысловое содержание этого этапа заключается в осуществлении событий, которые определяют, будет ли происходить активация гена.

Центральную роль в инициации транскрипции играют ДНК-связывающие белки. С одной из групп таких белков мы уже сталкивались. Это гистоны. Существует еще множество ДНК-связывающих белков, принимающих участие в процессах репликации, репарации, рекомбинации. Известна группа РНК-связывающих белков. На данном этапе экспрессии генома центральная роль принадлежит факторам инициации транскрипции. Это ДНК-связывающие белки, сложным образом комплексно взаимодействующие с геномом и приводящие к инициации транскрипции с помощью ДНК-зависимой РНК-полимеразы (РНК полимеразы II).

На свойствах разнообразия ДНК-связывающих белков мы не будем останавливаться, поскольку они детально описаны во многих современных учебниках по биохимии и молекулярной биологии.

Комплексы, инициирующие транскрипцию, формируются на конкретных участках молекулы ДНК, которые узнаются у прокариот непосредственно РНК-полимеразой, а у эукариот специфическими ДНК-связывающими белками, образующими платформу для посадки РНК-полимеразы.

У бактерий РНК-полимераза присоединяется к последовательности, называемой промотором и расположенной прямо перед группой генов в составе оперона. Так, промотор лактозного оперона кишечной палочки состоит из двух сегментов, каждый по 6 нуклеотидов. Промоторные регионы могут находиться на разном расстоянии от начала кодирующего участка гена. Расстояние измеряется десятками и сотнями пар нуклеотидов. Величина интервала между двумя участками (боксами) имеет важное значение, поскольку

определяет особенности взаимодействия ДНК-связывающего участка РНК-полимеразы с молекулой ДНК.

Эукариотические промоторы имеют более сложное строение. У эукариот термин промотор используется для описания всех последовательностей, которые важны для инициации транскрипции гена. Для некоторых генов этих последовательностей может быть много и они могут иметь разные функции. Среди промоторных последовательностей эукариот различают коровый промотор, иногда называемый также базальным промотором. Он является местом сборки иницирующего комплекса. Имеется также один или больше вышележащих промоторных элементов, то есть участков, расположенных выше корового промотора. Сборка иницирующего комплекса на коровом промоторе может происходить и без участия вышележащих элементов. Однако чаще всего она малоэффективна. Это указывает на то, что белки, связывающиеся с вышележащими элементами, являются активаторами транскрипции.

Каждая из трех имеющихся у эукариот ДНК-зависимых РНК-полимераз распознает свой тип промоторной последовательности. Промоторы РНК-полимеразы I состоят из корового промотора, перекрывающего точку инициации транскрипции. Вышележащий контрольный элемент расположен выше стартовой точки на расстоянии 100 пар оснований.

Промоторы РНК-полимеразы II могут быть самыми разными и располагаются на расстоянии нескольких килобаз от сайта инициации транскрипции. Коровый промотор состоит из двух сегментов – ТАТА бокса, расположенного на расстоянии 25 нуклеотидов от точки старта (5'-TATAWAW-3', где W – это А или Т), и расположенной перед стартовым нуклеотидом иницирующей (Inr) последовательности (5'-YYCARR-3', где Y – это С или Т, а R – это А или G). Некоторые гены, транскрибируемые РНК-полимеразой II, имеют только один из двух компонентов корового промотора, а иногда ни одного. Такие гены называют «нулевыми» генами. Они транскрибируются путем взаимодействия между РНК-полимеразой II и лежащей внутри гена последовательностью MED-1.

Промоторы РНК-полимеразы III необычны тем, что тоже локализованы внутри генов, чью транскрипцию они иницируют. Они очень вариабельны. Различают три категории таких промоторов. Обычно коровый промотор занимает 50-100 пар оснований и построен из двух боксов. Одна из категорий промоторов РНК-полимеразы III похожа на промоторы РНК-полимеразы II, имеет ТАТА бокс и вышележащие промоторные элементы.

Транскрипция иницируется путем присоединения к промоторным участкам РНК-полимеразы и вспомогательных белков. Результатом является образование закрытого

промоторного комплекса, который превращается в открытый промоторный комплекс при расщеплении ограниченного количества пар оснований вокруг точки инициации трансляции. Затем РНК-полимераза движется от промотора к точке старта с образованием стабильного транскрипционного комплекса, начинающего считывание информации, заложенной в кодирующей части гена. Представленная схема развития событий при инициации транскрипции в целом верна для всех четырех типов РНК-полимераз, включая прокариотическую, хотя для каждого типа полимеразы существуют свои особенности, описанные в учебниках по биохимии и молекулярной биологии.

Инициация транскрипции является стадией, для которой регуляторный контроль более всего важен. Этап, определяющий доступность генома, то есть описанный выше этап, является первым этапом, с помощью которого регулируется инициация транскрипции. Структура хроматина влияет на экспрессию генов путем контроля за доступностью промоторных последовательностей для РНК-полимеразы и ассоциированных с ней белков. Если на этой стадии определяется возможность экспрессии той или иной группы генов, то на стадии инициации транскрипции происходит регуляция, определяющая какие конкретно гены будут экспрессироваться в разных условиях.

У бактерий, таких как кишечная палочка, различают два пути, регулирующих инициацию транскрипции. Это конститутивный контроль, который зависит от структуры промотора, и регуляторный контроль, который зависит от регуляторных белков. Промоторная структура определяет базальный уровень инициации транскрипции. Консенсусная последовательность промотора *E. coli* весьма вариабельна. Вариабельность касается обоих промоторных боксов и дополняется различиями в строении точки старта транскрипции, что также влияет на эффективность работы промотора.

Эффективность промотора определяется количеством продуктивных инициаций в секунду. Продуктивная инициация приводит к тому, что РНК-полимераза покидает промотор и начинает синтез полноразмерного транскрипта. Наиболее сильные промоторы направляют продуктивную инициацию в 1000 раз эффективнее, чем слабые. Базальная скорость транскрипции предопределена последовательностью промотора и в нормальных условиях не меняется. Изменения могут происходить при появлении мутаций, изменяющих важные для промотора нуклеотиды. Эффективность работы промотора может меняться также при наличии мутаций в структуре σ -субъединицы РНК-полимеразы, которая обладает ДНК-связывающей способностью.

Регуляторный контроль осуществляется при помощи последовательностей ДНК, расположенных на разных расстояниях от промотора и называемых операторами.

Операторы регулируют инициацию транскрипции оперона, в состав которого обычно входит несколько генов, выполняющих единую биохимическую функцию. Присоединяющиеся к оператору регуляторные белки меняют уровень транскрипции гена, предотвращая связывание РНК-полимеразы с промотором или увеличивая эффективность инициации транскрипции. Кроме операторов у бактерий обнаружены расположенные на более отдаленных расстояниях энхансеры и сайленсеры. Первые увеличивают эффективность транскрипции, вторые – понижают. Они не так часто встречаются у бактерий и реализуют свои регуляторные функции, если ДНК формирует петли, благодаря которым энхансеры или сайленсеры взаимодействуют с РНК-полимеразой. Один энхансер или сайленсер может контролировать экспрессию нескольких генов.

У эукариот нет четких различий между конститутивной и регуляторной формами регуляции инициации транскрипции. РНК-полимераза прокариот прямо связывается с промотором, имея к нему сильный аффинитет. Вследствие этого базальная скорость транскрипции достаточно велика. У эукариот базальный уровень инициации транскрипции обычно очень мал. Чтобы была достигнута эффективная инициация необходимо участие дополнительных белков. Некоторые из них обозначают как конститутивные активаторы, которые влияют на работу многих генов. Другие могут быть обозначены как регуляторные активаторы, поскольку они регулируют отдельные гены.

Любой белок, который способен стимулировать инициацию транскрипции называют активатором. Изначально считалось, что все активаторы являются ДНК-связывающими белками, участвующими в инициации транскрипции. Одни узнают промоторные элементы, другие взаимодействуют с сайтами, влияющими на активность генов и называемыми энхансерами. У эукариот энхансеры могут находиться на больших расстояниях от кодирующей последовательности гена. Независимо от того, на каком расстоянии от тела гена активатор связывается с ДНК, он в соответствии с традиционной точкой зрения выполняет одну и ту же функцию – стабилизирует преинициаторный комплекс, взаимодействуя с ним.

Традиционная точка зрения верна в отношении большинства обнаруженных в настоящее время активаторов. В то же время обнаружено множество активаторов, которые не являются ДНК-связывающими белками. Тем не менее, они относятся к активаторам, осуществляя белок-белковые взаимодействия с преинициаторным комплексом.

Многие активаторы являются составной частью белковых комплексов, принимающих участие в регуляции структуры хроматина. Например, они входят в состав модифицирующих хроматин комплексов SAGA и Swi/Snf. Другие белки

классифицированы как активаторы, влияющие на экспрессию генов путем прямого воздействия на структуру ДНК и создания в ней дополнительных изгибов и искажений (Thomas, Travers, 2001). Вся совокупность активаторов, взаимодействующих с регуляторными участками ДНК и между собой, может образовывать подвижную, динамическую структуру, называемую энхансерсомой. В ее состав может входить множество различных активаторов. Энхансерсома способна взаимодействовать с преинициаторным комплексом. Эффективность взаимодействия и, соответственно, инициация транскрипции зависит от строения энхансерсомы.

Активаторы традиционного типа связываются с преинициаторным комплексом с помощью специализированного домена, называемого активаторным доменом. Активаторные домены переменны. Их делят на три категории:

- Кислые домены. Богаты кислыми аминокислотами (аспарагиновая и глутаминовая кислота). Это наиболее часто встречающаяся категория активаторных доменов.
- Глутамин-богатые домены. Часто обнаруживаются в составе активаторов, чей ДНК-связывающий мотив представляет собой так называемый гомеодомен.
- Пролин-богатые домены. Встречаются редко.

Детали взаимодействия между активаторами и пре-инициаторным комплексом активно изучаются. Многочисленные исследования показали, разные активаторы взаимодействуют с различными белками, составляющими пре-иницирующий комплекс. У дрожжей идентифицирован большой белковый комплекс, названный медиатором. Он образует контакты между активаторами и С-концевым доменом РНК-полимеразы II. Обнаружение медиаторов показало, что, скорее всего активирующий сигнал передается активатором на пре-иницирующий комплекс через медиатор.

Медиаторы могут обладать протеинкиназной активностью, что делает их способными фосфорилировать РНК-полимеразу, стимулируя ее к перемещению с промоторной последовательности. Значение медиаторов для инициации транскрипции у дрожжей подтверждено наблюдениями о том, что его компоненты, известные ранее как коактиваторы, являются белками, необходимым для полной активации пре-инициаторного комплекса. Структуры, эквивалентные медиаторам дрожжей, идентифицированы у млекопитающих (Malik, Roeder, 2000).

Показано, что имеется несколько различных типов медиаторов млекопитающих. Каждый отвечает за разные, но перекрывающиеся наборы активаторов. В настоящее время считается, что медиаторы являются обязательным компонентом пре-инициаторного комплекса РНК-полимеразы II. Стимулирующие эффекты от всех активаторов проходят

через медиаторы. При этом одни и те же активаторы могут действовать через разные медиаторы.

Таким образом, в настоящее время сформировано представление о существовании множества активаторов, действующих через медиаторы и образующих сложные динамичные быстро меняющиеся структуры, называемые энхансерсомами. Состав энхансерсом непостоянен, их структура пластична и представляется исследователям как образование, напоминающее размытый пазл, собираемый из множества составляющих, не имеющих очерченных границ. У более высоко организованных видов в геноме более высоко относительное содержание генов, кодирующих регуляторные белки, в том числе факторы, инициирующие транскрипцию. С их помощью обеспечивается тонкая регуляция сложных биохимических процессов, характерная для более продвинутых в эволюционном плане организмов.

Наряду с активаторами транскрипции обнаружено множество ДНК-связывающих белков, репрессирующих инициацию транскрипции. Эти белки связываются с промоторными элементами, так и с более отдаленными участками, называемыми сайленсерами. Определенное влияние на экспрессию генома осуществляется через деацетилирование гистонов и метилирование ДНК, то есть через общие пути регуляции экспрессии генома. Однако существуют и другие более специфические репрессирующие воздействия на индивидуальные промоторы. Например, у дрожжей репрессоры, называемые Mot1 и NC2, ингибируют сборку пре-инициирующего комплекса, связываясь с одним из его компонентов (TBP) и нарушая его активность (Lee, Young, 2000). Оба репрессора имеют широкий спектр активности, инактивируя большой набор генов. Подобные репрессоры обнаружены и у других эукариот (Smith, Johnson, 2000).

Сложность процессов регуляции инициации транскрипции демонстрируется тем, что в зависимости от условий некоторые регуляторные белки могут осуществлять как репрессирующее, так и активирующее действие. Примером может служить упомянутый выше репрессор NC2. Этот белок ингибирует транскрипцию при наличии ТАТА бокса в составе промотора, но активирует транскрипцию, если ТАТА бокс отсутствует (Willy et al., 2000).

Наряду с регуляцией инициации транскрипции с помощью активаторов и репрессоров регулируется и работа самих активаторов и репрессоров. Одним из вариантов регуляции является контроль за их синтезом. Однако такой контроль не приводит к быстрым изменениям в экспрессии генома, поскольку нужно время для накопления в клетке или разрушения активаторов и репрессоров. Этот тип контроля используется клетками для контроля активаторов и репрессоров, ответственных за поддержание

стабильности общей картины экспрессии генома. Например, такой способ контроля характерен для процессов клеточной дифференцировки. Альтернативным путем является химическая модификация активаторов и репрессоров, например их фосфорилирование или индукция конформационных изменений. Такие изменения происходят намного быстрее, чем синтез белков *de novo*, и позволяют клетке оперативно отвечать на внеклеточные сигналы, индуцирующие быстрые изменения в экспрессии генома.

2.4.3. Дегградация матричной РНК

Хорошо известно, что успешная инициация транскрипции приводит к последующему синтезу и процессингу РНК. Механизмы транскрипции и процессинга хорошо известны и детально описаны (Жимулев, 2005). В связи с этим мы не будем обсуждать эти этапы экспрессии генома, а перейдем к обсуждению аспектов, связанных со следующим этапом экспрессии генома – дегградацией РНК.

Дегградация РНК является таким же важным процессом, что и синтез. Особенно это важно в отношении матричных РНК, чье присутствие или отсутствие в клетке определяет, какие именно белки будут формировать основные структурно-функциональные черты протеома. Специфическая дегградация мРНК является одним из путей регуляции генной экспрессии.

Скорость дегградации мРНК оценивается как период полужизни этих молекул. Между организмами имеются значительные колебания в периодах полужизни мРНК. Бактериальные мРНК обычно имеют очень небольшой период полужизни. Он редко длится больше нескольких минут, что отражает быстрые изменения в спектре синтезируемых белков. У бактерий время генерации может составлять 20 минут. Такой быстрый рост клеток должен быть обеспечен быстрыми изменениями метаболизма и соответствующими быстрыми изменениями транскриптома.

Эукариотическая мРНК является более долгоживущей. У дрожжей период полужизни мРНК в среднем составляет 10-20 минут, у млекопитающих - десятки минут и часы. Внутри отдельной клетки наблюдается широкий разброс в продолжительности периодов существования мРНК. Некоторые дрожжевые мРНК имеют период полужизни, равный только одной минуте, в то время у других он составляет 35 минут. Каким же образом осуществляется процесс дегградации мРНК и как этот процесс контролируется?

Бактериальные мРНК деградируют в направлении 3'-5'. В их дегградации участвует ряд рибонуклеаз и РНК-деградирующих ферментов (Carpousis et al., 1999). Среди них:

-РНК-аза E и РНК-аза III, которые являются эндонуклеазами, разрезающими молекулы РНК.

-РНК-аза II, которая является экзонуклеазой, удаляющей нуклеотиды в направлении 3'-5'.

-Полинуклеотидфосфорилаза (PNP), которая также последовательно удаляет нуклеотиды с 3'-конца, хотя как истинная фосфорилаза она нуждается в присутствии неорганического фосфата в качестве косубстрата.

Ферментов, которые осуществляли бы деградацию молекул мРНК в направлении 5'-3', у бактерий не обнаружено.

Набор ферментов, осуществляющих деградацию мРНК, показывает, что их расщепление осуществляется с 3'-конца. Однако такое понимание механизма деградации сразу вызвало сомнение, поскольку большинство мРНК имеют шпилько-образную структуру около 3'-конца. Подобная шпилька индуцирует терминацию транскрипции. Она блокирует перемещение по молекуле РНК рибонуклеаз и PNP.

Детальное изучение процесса деградации бактериальных мРНК показало, что РНК-аза E и PNP локализованы внутри комплексов, названных деградосомами. Еще одним компонентом деградосом является РНК-хеликаза, которая расплетает двуспиральную структуру петель РНК. Деградосомы способны расщеплять также рибосомальную РНК. Однако, до настоящего времени реальность деградосом подвергается сомнению, поскольку в их составе обнаруживаются ферменты, не имеющие отношения к распаду РНК. В частности в составе деградосом обнаружен фермент гликолиза енолаза. Независимо от того, каким образом будет решен вопрос о деградосомах, в бактериальных клетках имеется набор ферментов, обеспечивающий упорядоченную и специфическую деградацию мРНК.

Эукариоты имеют более широкий спектр механизмов деградации мРНК. Лучше всего механизмы деградации мРНК изучены у дрожжей. Идентифицировано по крайней мере четыре пути, обеспечивающие деградацию мРНК. Один из них идет с участием мультибелкового комплекса, названного экзосомой. Он разрушает транскрипты в направлении 3'-5' и содержит нуклеазы, сходные с ферментами бактериальной деградосомы. Экзосомы присутствуют и в клетках млекопитающих. Их важность совершенно очевидна, однако они недостаточно изучены. Предполагается, что их роль заключается не столько в деградации мРНК как таковой, а скорее в мониторинге полиаденилирования, в обеспечении правильного полиаденилирования молекул мРНК, покидающих после синтеза ядро клетки (Hileren et al., 2001).

Значительно больше известно о двух других механизмах деградации мРНК эукариот. Первый из них обозначают как декэпинг, зависимый от деаденилирования (deadenylation-dependent decapping). Этот механизм включается при удалении поли(А)-хвоста у молекулы мРНК. Удаление происходит либо за счет экзонуклеазного расщепления, либо за счет потери полиаденилат-связывающего белка, который стабилизирует хвост молекулы. Вслед за удалением поли(А)-хвоста происходит расщепление 5'-кэпа с помощью соответствующего фермента, названного Dcp1p. Декэпирование приводит к невозможности трансляции молекулы и соответственно к потере функциональной активности молекулы. После этого мРНК подвергается быстрому экзонуклеазному расщеплению с 5'-конца. Будет ли данная конкретная молекула деградировать определяется доступностью Dcp1p к кэпу, что в свою очередь зависит от ассоциации между кэпом и белками, связывающимися с ним для инициации трансляции.

Деградация также зависит от последовательностей, называемых элементами нестабильности и локализованными внутри транскриптов. Значение этих последовательностей продемонстрировано экспериментами, в которых элементы искусственно удаляли, что приводило к повышению уровня трансляции и снижению скорости деградации мРНК (Tucker, Parker, 2000).

Вторым хорошо изученным механизмом деградации эукариотических мРНК является механизм, основанный на использовании системы, называемой nonsense-mediated RNA decay (NMD) или mRNA surveillance - РНК-выживание. Первое имя дано благодаря выполняемой функции, поскольку на жаргоне молекулярных биологов нонсенс-последовательность является терминирующим кодоном. NMD приводит к специфической деградации мРНК, имеющих терминирующий кодон в неправильной позиции. Он может появляться вследствие мутаций или как результат некорректного сплайсинга.

Некорректный кодон выявляется с помощью так называемой машины выживания, которая состоит из комплекса белков. Он сканирует мРНК и способен отличить корректный терминирующий кодон, расположенный в конце кодирующего региона транскрипта, от некорректного, расположенного в неправильном месте (Culbertson, 1999). Показано, что корректное расположение терминирующего кодона - это расположение сразу перед точкой соединения экзона и интрона. В связи с этим, предполагается, что в основе правильного распознавания терминирующего кодона лежит способность ферментов выживания распознавать терминирующий кодон, лежащий сразу перед точкой соединения экзона и интрона.

У эукариот присутствуют также другие механизмы деградации РНК, которые вовлечены в основном в защиту клетки от атак со стороны чужеродных РНК, например

РНК, являющихся вирусными геномами. Примером является механизм, названный РНК-интерференция. Результатом этого механизма является расщепление вирусной двунитовой РНК на короткие фрагменты длиной 21-25 нуклеотидов (siRNA) с помощью фермента, названного Dicer. Фрагментация инактивирует вирусный геном. Если вирусный геном уже успел транскрибироваться, то запускается второй этап РНК-интерференции. Расплетенные нити siRNA спариваются с вновь образованной вирусной мРНК, образуя двунитовые участки, которые являются мишенью для специфической нуклеазы, называемой RDE-1. Эта нуклеаза разрушает вновь образованную вирусную мРНК.

2.4.4. Инициация трансляции

Шестым этапом экспрессии генетического материала, заключенного в клетке, является инициация трансляции матричной РНК. Как известно, трансляция, то есть синтез белка, происходит на рибосомах. Их строения мы в настоящем пособии касаться не будем, так же как не будем затрагивать особенностей строения транспортных РНК. Подробную информацию о них можно найти в учебниках по биохимии и молекулярной биологии.

Инициация трансляции вынесена в отдельный этап экспрессии генома, поскольку она, как и инициация транскрипции, также является важным звеном, контролирующим формирование клеточного протеома.

Несмотря на сходство в строении бактериальных и эукариотических рибосом, имеются различия в осуществлении трансляции у представителей этих двух миров живых организмов. Ярче всего эти различия проявляются на стадии инициации трансляции, когда рибосома собирается на молекуле мРНК в участке, расположенном выше иницирующего кодона. Основное отличие инициации трансляции у бактерий от эукариот заключается в том, что у бактерий комплекс, иницирующий трансляцию, выстраивается прямо на иницирующем кодоне, то есть точке, с которой начинается синтез белка. Эукариоты используют более сложный механизм инициации, который мы постараемся представить коротко ниже.

Но сначала обсудим инициацию трансляции у бактерий. Хорошо известно, что рибосомы, на которых не синтезируется белок, диссоциируют на субъединицы, которые остаются в цитоплазме в ожидании нового цикла трансляции. У бактерий процесс иницируется, когда малая субъединица, связанная с фактором инициации трансляции IF3, присоединяется к рибосома-связывающему сайту, называемому также последовательностью Шайна-Далгарно. Это короткий сайт, имеющий у *E. coli* последовательность 5'-AGGAGGU-3' и локализованный в районе 3-10 нуклеотида выше иницирующего кодона. Это точка, с которой начинается трансляция. Рибосома-

связывающий сайт комплементарен региону на 3'-конце 16S рРНК, находящемуся на малой субъединице. Этот регион и обеспечивает комплементарное взаимодействие. Спаривание оснований в этом регионе приводит к взаимодействию малой субъединицы с мРНК.

Присоединение малой субъединицы к рибосома-связывающему сайту закрывает иницирующий кодон. Это обычно кодон 5'-AUG-3', который кодирует метионин, хотя иногда используются и другие кодоны (5'-GUG-3', 5'UUG-3'). Все три кодона могут распознаваться одной и той же инициаторной тРНК. Инициаторная тРНК способна аминоацилировать метионин и соответственно модифицировать его, превращая в N-формилметионин. В присоединении формилметионина к малой субъединице принимает участие фактор инициации трансляции IF-2. Завершение инициаторного этапа происходит, когда с инициаторным комплексом связывается фактор IF-1. Он индуцирует конформационные изменения в инициаторном комплексе, обеспечивая присоединение большой субъединицы. Затем иницирующие факторы освобождаются, что приводит к наступлению фазы элонгации трансляции, то есть к наступлению следующего этапа экспрессии генома - этапа синтеза белка.

У эукариот первоначально происходит сборка пре-иницирующего комплекса. Эта структура включает 40S субъединицу рибосомы, несколько иницирующих факторов (eIF-2, eIF-1, tIF-1A, eIF-3), молекулу ГТФ и инициаторную метиониновую тРНК. После сборки пре-инициаторный комплекс ассоциирует с 5'-концом мРНК. Для этого необходимо участие кэп-связывающего комплекса, называемого также eIF-4F, и состоящего из факторов инициации eIF-4A, eIF-4E, eIF-4G. Последний из перечисленных факторов выступает в роли мостика. Мостик образуется между фактором eIF-4E, связывающимся с кэпом, и фактором eIF-3, присоединенным к преиницирующему комплексу (Hentze, 1997). В результате пре-иницирующий комплекс становится связанным с 5'-регионом мРНК. На присоединение пре-инициаторного комплекса с мРНК также влияет поли-адениловый хвост (поли(А)) на 3'-конце мРНК. Это взаимодействие опосредуется полиаденилат-связывающим белком (PADP), который присоединяется к поли(А)-хвосту. Полиадениловый хвост играет регуляторную роль, длина хвоста коррелирует с выраженностью иницирующих процессов.

После того, как пре-иницирующий комплекс связался с 5'-концом мРНК, его называют уже иницирующим комплексом. Он начинает сканировать молекулу мРНК в поиске иницирующего кодона. Лидирующая последовательность в эукариотических мРНК может иметь длину в несколько сотен и тысяч нуклеотидов и часто содержит регионы, образующие шпильки и другие спаренные структуры. Они удаляются с

помощью белков кэп-связывающего комплекса, обладающих хеликазной активностью (eIF-4A, eIF-4B). Иницирующий кодон у эукариот (5'-AUG-3') распознается в составе так называемой последовательности Козака (Kozak consensus) – 5'-ACCAUGG-3'.

Когда иницирующий комплекс достигает иницирующего кодона, к малой субъединице рибосомы присоединяется большая субъединица. Как и у бактерий это требует гидролиза ГТФ и освобождения иницирующих факторов.

Инициация может происходить также и без сканирования мРНК. Это впервые было замечено при изучении пикорнавирусов, группы РНК-содержащих вирусов, включающих человеческий полиовирус и риновирусы. Транскрипты этих вирусов не копируются. Вместо этого они имеют последовательности, сходные с рибосома-связывающим сайтом у бактерий (internal ribosome entry site -IRES). Эукариотические клетки также несут белки, которые способны иницировать трансляцию с помощью IRES. Примером является мРНК белка, связывающего тяжелую цепь иммуноглобулина млекопитающих. IRES найдены также в некоторых мРНК, кодирующих белки, транслирующиеся при стрессовых воздействиях. В этих условиях кэп-зависимая инициация трансляции подавляется. Ее заменяет альтернативный вариант инициации трансляции.

После этапа инициации трансляции наступают фазы элонгации и терминации, то есть непосредственно синтез белка, составляющий следующий этап экспрессии генома. Механизм синтеза, включающий эти фазы, хорошо известен, в связи с этим здесь мы не будем его рассматривать. Взамен обсудим механизмы регуляции инициации трансляции.

Различают два типа регуляции. Первый заключается в глобальной регуляции, которая приводит к общим изменениям в количестве синтезируемого белка. У эукариот это достигается, например, фосфорилированием eIF-2, что приводит к подавлению инициации трансляции. Фосфорилирование этого фактора инициации происходит также при стрессовых воздействиях, таких как тепловой шок, когда общий уровень синтеза белка падает и начинает превалировать IRES-опосредованная инициация трансляции.

Второй тип регуляции – это транскрипт-специфическая регуляция. Она использует механизмы, которые воздействуют на отдельные транскрипты или малые группы транскриптов, кодирующие отдельные белки. В качестве примера обычно приводят транскрипт-специфическую регуляцию оперона, включающего гены, кодирующие рибосомальные белки. В каждом опероне лидирующий регион мРНК содержит последовательность, которая действует как сайт связывания для одного из белков, кодируемых этим же опероном. Когда такой белок синтезируется, он может либо присоединяться к рибосомальной РНК, участвуя в образовании рибосомы, либо

связываться с лидирующим регионом РНК. Если в клетке присутствует свободная рибосомальная РНК, то он присоединяется к последней, формируя рибосому. Когда все рибосомальные РНК собираются в рибосомы, то рибосомальные белки начинают связываться уже с собственной мРНК, блокируя инициацию трансляции. В результате, блокируется синтез рибосомальных белков, кодируемых соответствующими им мРНК. Таким образом, синтез каждого рибосомального белка регулируется концентрацией свободных рибосомальных РНК в клетке.

Второй пример транскрипт-специфической регуляции касается мРНК ферритина – железо-запасяющего белка млекопитающих. В отсутствие железа синтез ферритина ингибируется белками, которые связываются с последовательностями, называемыми iron-response elements (элементами, отвечающими за железо). Они локализованы в лидирующей последовательности мРНК ферритина. Связавшийся белок блокирует рибосому, когда она пытается сканировать мРНК, перемещаясь по ней в поисках иницирующего кодона. В присутствии железа связывающийся белок отсоединяется и мРНК транслируется.

Интересно, что мРНК, кодирующая рецептор трансферрина, также имеет элемент, отвечающий на железо. Но в этом случае отсоединение связывающегося белка в присутствии железа приводит не к трансляции мРНК, а наоборот к ее деградации. Это происходит потому, что при появлении железа в клетке функциональная активность рецептора трансферрина не нужна, поскольку нет необходимости транспортировать железо из внеклеточного пространства.

2.4.5. Процессинг, фолдинг и деградация белков

Синтез белка не является последней стадией экспрессии генома. Полипептид, выходящий из рибосомы неактивен. Прежде чем приобрести функциональную активность, он должен подвергнуться, по крайней мере, первому из приведенных ниже четырех типов пост-трансляционных изменений:

- Фолдинг белка. Полипептид не активен, пока не приобретает корректную конформацию.

- Протеолитическое расщепление. Некоторые белки процессируются путем разрезания исходной полипептидной цепи с помощью протеаз. При разрезании от полипептидной цепи могут отрезаться последовательности, находящиеся на ее концах, что приводит к появлению укороченного варианта белка. Полипептидная цепь может также нарезаться на фрагменты, которые функционально активны.

- Химическая модификация. Отдельные аминокислоты в составе полипептидной цепи могут быть модифицированы путем присоединения к их радикалам химических групп.

- Интеиновый сплайсинг. Некоторые белки содержат интеиновые последовательности, подобно тому, как гены и мРНК содержат интроны. Интеины могут быть затем удалены, а экстеины сшиты таким образом, что сшитый блок полипептидов приобретает функциональную активность.

Часто различные типы процессинга происходят одновременно. Однако обнаружены случаи последовательного процессинга, когда, например, разрезание или химическая модификация происходят после фолдинга. Такие ситуации возникают в тех случаях, когда процессинг выступает как регуляторный механизм: подвергнутый фолдингу белок становится активным после дополнительного процессинга в виде разрезания полипептидной цепи или химической модификации конкретных аминокислот.

Одним из центральных принципов молекулярной биологии заключается в том, что вся информация, нужная полипептидной цепи для принятия правильной трехмерной структуры, содержится в аминокислотной последовательности. Представления о том, что аминокислотная последовательность содержит всю информацию о третичной структуре, получена при изучении рибонуклеазы еще в 60-е годы. (Anfinsen, 1973). Рибонуклеаза является небольшим белком, состоящим из 124 аминокислот, содержит 4 дисульфидные связи. Вторичная структура представлена преимущественно бета-структурой, альфа-спиральных участков мало. Денатурирующие агенты, такие как мочевины, приводят к потере функциональной активности белка. Удаление денатурирующего агента приводит к рефолдингу белка и возвращает его функциональную активность.

Изучение особенностей фолдинга привело к пониманию двухэтапности этого процесса (Hartl, 1996).

1. После удаления денатурирующего агента элементы вторичной структуры образуются в течение нескольких миллисекунд. Этот этап сопровождается свертыванием белка в компактную, но не конечную структуру с гидрофобными группами, расположенными внутри молекулы.

2. В ходе следующих нескольких секунд или минут элементы вторичной структуры взаимодействуют друг с другом и постепенно принимают конечную третичную структуру. Формирование конечной конформации часто идет через серию промежуточных состояний. Возможны альтернативные пути, приводящие молекулу в конечное конформационное состояние. Отклонение от путей корректного фолдинга приводит к неправильной конформации, которая обычно нестабильна, что создает условия

для возвращения к правильной конформации. Грубые отклонения в осуществлении фолдинга, например при резких температурных воздействиях, приводят к невозможности возвращения к правильной конформации и, соответственно, к денатурации белка.

Такие представления о механизме фолдинга верны для небольших белков. Более крупные белки не имеют такой способности. Два фактора предотвращают спонтанный фолдинг более крупных белков. Во-первых, их стремление образовывать нерастворимые агрегаты при удалении денатурирующих агентов. Во-вторых, большие белки стремятся принять множество непродуктивных конформаций, которые более стабильны и не позволяют завершиться фолдингу.

В функционирующей клетке фолдинг осуществляется еще до окончания белкового синтеза. Более того, в клетках фолдинг идет при помощи молекулярных шаперонов – белков, помогающих принять правильную конформацию. Они обнаружены как у прокариот, так и у эукариот (Hartl, 1996, Slavotinek, Biesecker, 2001). Хорошо известны шапероны *E. coli*, обозначаемые как DnaK, DnaJ, GrpE, и шаперонины GroEL/GroES. Молекулярные шапероны не воздействуют на третичную структуру белка. Они помогают белку принять наиболее выгодную с термодинамических позиций конформацию.

Протеолитическое расщепление имеет две функции в пост-трансляционном процессинге. Во-первых, оно удаляет короткие участки N- или C-концевых регионов полипептида, создавая укороченную молекулу, которая свертывается в активный белок. Во-вторых, оно приводит к расщеплению полипротеина на сегменты, часть которых становится активными белками. Протеолитический процессинг характерен для секретируемых белков, чья активность может быть нежелательна внутри клетки. Примером может служить меллитин – основной белок пчелиного яда. Меллитин лизирует клетки, как у млекопитающих, так и у пчел. Поэтому он изначально синтезируется в форме неактивного предшественника, который имеет 22 добавочные аминокислоты на N-конце. Эта последовательность удаляется внеклеточной протеазой.

Подобный тип процессинга характерен для инсулина - гормона, синтезируемого островками Лангерганса поджелудочной железы и контролирующего концентрацию сахара в крови. Инсулин синтезируется как преинсулин. Имеет длину 105 аминокислот. В ходе процессинга отщепляется 24 аминокислоты с превращением преинсулина в проинсулин. Затем происходит еще два добавочных разрезания полипептидной цепи, которые дают две активные цепи инсулина, связанные между собой дисульфидным мостиком.

Некоторые белки изначально синтезируются как полипротеины. Их расщепление может приводить к появлению набора коротких пептидов, обладающих различными функциями. Полипротеины встречаются у вирусов. У позвоночных полипротеины являются предшественниками некоторых гормонов. Например, полипротеин, называемый проопиомеланокортином, является предшественником, по крайней мере, 10 различных пептидных гормонов. Они освобождаются при протеолитическом расщеплении полипротеина, но не продуцируются в полном составе, поскольку имеются перекрытия между последовательностями отдельных пептидов. В результате в разных клетках синтезируются разные наборы таких пептидных гормонов.

Геном имеет способность кодировать 21 аминокислоту. Двадцать определяются стандартным генетическим кодом, селеноцистеин появляется в полипептидной цепи в результате контекст-зависимого считывания кодона 5'-UGA-3'. Этот набор аминокислот значительно увеличивается в результате пост-трансляционной химической модификации аминокислот.

Самый простой тип модификации представляет собой добавление небольших химических группировок к аминокислотным радикалам. Например, добавляются ацетильные, метильные, фосфатные группы. К концевым аминокислотам могут добавляться карбоксильные или аминогруппы. Для различных белков документировано более 150 различных модификаций. Каждая из них осуществляется с помощью специфического процесса.

Химическая модификация часто играет важную функциональную роль. Например, уже известная нам модификация гистонов приводит к изменениям в структуре хроматина, что весьма существенно для механизмов реализации информации, заложенной в геноме. Важную роль играет химическая модификация белков в передаче внутриклеточных сигналов.

Более сложным типом модификации является гликозилирование белков, то есть присоединение больших углеводных цепочек к полипептидной цепи. Имеется два основных типа гликозилирования: O-связанное гликозилирование – присоединение углеводных остатков через гидроксильные группы серина и треонина, и N-связанное гликозилирование – присоединение углеводов через аминогруппы боковых цепей аспарагина.

Гликозилирование может быть результатом присоединения к белку сложной структуры, построенной из сети углеводных остатков. Углеводные компоненты помогают белку найти свое положение в клетке, определяют стабильность белков, циркулирующих в крови.

Еще один тип крупномасштабной модификации представляет собой присоединение к полипептидам длинных липидных цепей. Присоединение идет к остаткам серина или треонина. Такой процесс назван ацилированием и характерен для белков, ассоциированных с мембраной. В ряде случаев встречается биотинилирование. Молекула биотина присоединяется к ферментам, которые катализируют карбоксилирование органических кислот, таких как ацетат и пропионат (Chapman-Smith, Cronan, 1999).

Еще одним типом процессинга является интеиновый сплайсинг. Интеины – это внутренние сегменты белков, которые в отличие от экстеинов удаляются после трансляции. Первый интеин был открыт в 1990 году у дрожжей. Более всего интеины распространены у бактерий и архей. В составе белков встречается как по одному интеину, так и по несколько интеинов.

Большинство из них имеют длину примерно по 150 аминокислот. У разных белков аминокислотные последовательности в месте соединения интеинов имеют некоторое сходство друг с другом. В частности, первая аминокислота, расположенная сразу после экстеина – это цистеин, серин или треонин. Несколько других аминокислот внутри последовательности интеина также консервативны. Они участвуют в процессе сплайсинга, который представляет собой самокатализируемый процесс (Paulus, 2000). В соответствии с этим их принято рассматривать как аутоферменты, которые сами себя расщепляют и сшивают.

Завершающим этапом экспрессии генома является специфическая деградация белков, синтезированных и процессированных на предыдущих этапах. Каждый белок имеет срок своего существования, протеом клетки постоянно меняется в соответствии с внутренними потребностями клетки и внешними воздействиями. Изменения протеома происходят не только за счет синтеза белков *de novo*, но и за счет удаления белков, чья функция уже выполнена. Такое удаление имеет высоко селективный характер. Ненужные белки быстро уничтожаются. Это имеет важное значение, поскольку в клетке постоянно идут высоко динамичные процессы, требующие эффективного обновления белков, составляющих протеом (Hunt, 1997). Таким образом, деградация белка - это важный фактор формирования протеома, фактор, контролирующий экспрессию генома.

Многие годы исследование процессов деградации белков было не модным. Однако в 90-е годы был сделан серьезный прорыв в понимании того, как процессы специфического протеолиза принимают участие в реализации клеточного цикла, в дифференцировке клеток. Появилось понимание этого процесса, как высокоспецифического процесса, участвующего в регуляции внутриклеточных событий.

В настоящее время ясно, что в клетках существует упорядоченная система деградации белков. Бактерии используют в этих целях целый спектр совместно работающих протеаз. У эукариот такие события идут с участием специализированной и достаточно сложно устроенной системы, включающей убиквитин и протеасомы. Детальное описание убиквитиновой системы и структурно-функциональная характеристика протеасом представлены в электронном пособии для преподавателей и сотрудников ВУЗов, входящем в данную серию пособий (Бабаев и др., 2007).

Таким образом, описанные выше стадии экспрессии генома приводят к формированию протеома, который в свою очередь определяет биохимическую активность клетки. Ни в одном организме биохимические процессы не могут быть постоянными и неизменными. Даже наиболее просто устроенные одноклеточные организмы меняют свой протеом в соответствии с сигналами, поступающими из окружающего пространства. Биохимические процессы постоянно отвечают на изменения условий существования клетки. Аналогичным образом и многоклеточные организмы отвечают на внеклеточные стимулы. Однако, у многоклеточных организмов добавляется система межклеточной регуляции метаболизма, включающая гормоны и другие молекулярные факторы.

Все эти воздействия приводят к быстрым перестройкам в активности генов, приводящим в свою очередь к ремоделированию протеома и модуляции биохимических реакций. Другие изменения активности генома являются более продолжительными. Они приводят к клеточной дифференцировке, адаптируя клетки к выполнению специализированных функций. У человека обнаружено более 250 типов специализированных клеток, формирующих органы и ткани. Сборка таких многоклеточных структур нуждается в координации активности геномов различных клеток. Эта координация включает как быстрые, так и более постоянные изменения в работе генома и происходит в организме весь период его развития, то есть всю жизнь.

Литература

- И.Ф.Жимулев «Общая и молекулярная генетика», 2005
- Б.Льюин. «Гены», 1987
- М.Сигнер, П.Берг «Гены и геномы», 1998
- MA. Adams, SE. Celniker, and RA. Holt, *et al.* (2000). The genome sequence of *Drosophila melanogaster* *Science* 287: 2185-2195.
- AGI (The Arabidopsis Genome Initiative). (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* *Nature* 408: 796-815.
- J. Ahringer. (2000). NuRD and SIN3: histone deacetylase complexes in development *Trends Genet.* 16: 351-356.
- A. Akhmanova, F. Voncken, and T. van Allen, *et al.* (1998). A hydrogenosome with a genome *Nature* 396: 527-528.
- AA. Alizadeh, MB. Eisen, and RE. Davis, *et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling *Nature* 403: 503-511.
- RA. Alm, L-SL. Ling, and DT. Moir, *et al.* (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori* *Nature* 397: 176-180.
- S. Anderson, AT. Bankier, and AG. Barriell, *et al.* (1981). Sequence and organization of the human mitochondrial genome *Nature* 290: 457-474. ([PubMed](#))
- SGE. Andersson, A. Zomorodipour, and JO. Andersson, *et al.* (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria *Nature* 396: 133-140.
- GJ. Barton. (1995). Protein secondary structure prediction *Curr. Opin. Struct. Biol.* 5: 372-376.
- FR. Blattner, G. Plunkett, and CA. Bloch, *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12 *Science* 277: 1453-1462.
- AJ. Bannister, P. Zegerman, and JF. Partridge, *et al.* (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain *Nature* 410: 120-124.
- SB. Baylin and JG. Herman. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics *Trends Genet.* 16: 168-174.
- AC. Bell, AG. West, and G. Felsenfeld. (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome *Science* 291: 447-450.
- SL. Berger. (2000). Local or global? *Nature* 408: 412-415.
- A. Bird. (1999). DNA methylation de novo *Science* 286: 2287-2288.
- T.A.Brown “Genomes” BIOS Scientific Publishers, LTD, 2002
- CJ. Bult, O. White, and GJ. Olsen, *et al.* (1996). Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii* *Science* 273: 1058-1073.

- M. Bustin. (2001). Chromatin unfolding and activation by HMGN chromosomal proteins *Trends Biochem. Sci.* 26: 431-437.
- C. Costanzi and JR. Pehrson. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals *Nature* 393: 599-601.
- H. Caron, B. van Schail, and S. van der Mee, *et al.* (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains *Science* 291: 1289-1292.
- AJ. Carpousis, NF. Vanzo, and LC. Raynal. (1999). mRNA degradation: a tale of poly(A) and multiprotein machines *Trends Genet.* 15: 24-28.
- A. Chapman-Smith and JE. Cronan. (1999). The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity *Trends. Biochem. Sci.* 24: 359-363.
- DA. Clayton. (2001). A big development for a small RNA *Nature* 410: 29-31.
- CESC (The *C. elegans* Sequencing Consortium). (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology *Science* 282: 2012-2018.
- ST. Cole, R. Brosch, and J. Parkhill, *et al.* (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence *Nature* 393: 537-544.
- GP. Copenhaver, K. Nickel, and T. Kuromori, *et al.* (1999). Genetic definition and sequence analysis of *Arabidopsis* centromeres *Science* 286: 2468-2474.
- G. Deckert, PV. Warren, and T. Gaasterland, *et al.* (1998). The complete genome of the hyperthermophile bacterium *Aquifex aeolicus* *Nature* 392: 353-358.
- RF. Doolittle. (1997). Microbial genomes opened up *Nature* 392: 339-342.
- WF. Doolittle. (1999). Lateral genomics *Trends Cell Biol.* 9: M5-M8.
- CM. Fraser, S. Casjens, and WM. Huang, *et al.* (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi* *Nature* 390: 580-586.
- CM. Fraser, SJ. Norris, and GM. Weinstock, *et al.* (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete *Science* 281: 375-388.
- EF. Fritsch, RM. Lawn, and T. Maniatis. (1980). Molecular cloning and characterization of the human α -like globin gene cluster *Cell* 19: 959-972.
- R. Feil and S. Khosia. (1999). Genomic imprinting in mammals: an interplay between chromatin and DNA methylation? *Trends Genet.* 15: 431-434.
- TI. Gerasimova, K. Byrd, and VG. Corces. (2000). A chromatin insulator determines the nuclear localization of DNA *Mol. Cell* 6: 1025-1035.
- E. Heard, P. Clerc, and P. Avner. (1997). X-chromosome inactivation in mammals *Ann. Rev. Genet.* 31: 571-610.
- JF. Heidelberg, JA. Eisen, and WC. Nelson, *et al.* (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae* *Nature* 406: 477-483.

- M. Holcik, N. Sonenberg, and RG. Korneluk. (2000). Internal ribosome initiation of translation and the control of cell death *Trends Genet.* 16: 469-473.
- WA. Houry, D. Frishman, C. Eckerskorn, F. Lottspeich, and FU. Hartl. (1999). Identification of *in vivo* substrates of chaperonin GroEL *Nature* 402: 147-154.
- CA. Hutchison, SN. Peterson, and SR. Gill, *et al.* (1999). Global transposon mutagenesis and a minimal *Mycoplasma* genome *Science* 286: 2165-2169.
- IHGSC (International Human Genome Sequencing Consortium). (2001). Initial sequencing and analysis of the human genome *Nature* 409: 860-921.
- S. Imai, CM. Armstrong, M. Kaerberlein, and L. Guarente. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase *Nature* 403: 795-800.
- F. Jacob and J. Monod. (1961). Genetic regulatory mechanisms in the synthesis of proteins *J. Mol. Biol.* 3: 318-356.
- R. Jaenisch. (1997). DNA methylation and imprinting: why bother? *Trends Genet.* 13: 323-329.
- T. Jenuwein and CD. Allis. (2001). Translating the histone code *Science* 293: 1074-1080.
- P. Jeppesen and BM. Turner. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression *Cell* 74: 281-289.
- PA. Jones. (1999). The DNA methylation paradox *Trends Genet.* 15: 34-37.
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* 2004,431(7011):915-916.
- H-P. Klenk, RA. Clayton, and J-F. Tomb, *et al.* (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus* *Nature* 390: 364-370.
- BF. Lang, G. Burger, and CJ. O'Kelly, *et al.* (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature *Nature* 387: 493-497.
- ES. Landers, LM. Linton, and B. Birren, *et al.* 2001. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome *Nature* 409: 860-921. .
- M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins *Nature* 410: 116-120.
- MA. Lever, JPH. Th'ng, X. Sun, and MJ. Hendzel. (2000). Rapid exchange of histone H1 *in vivo* on chromatin in living cells. *Nature* 408: 873-876.
- JD. Lewis and D. Tollervey. (2000). Like attracts like: getting RNA processing together in the nucleus *Science* 288: 1385-1389.
- Q. Li, S. Harju, and KR. Peterson. (1999). Locus control regions: coming of age at a decade plus *Trends Genet.* 15: 403-408.

- MD. Litt, M. Simpson, M. Gaszner, CD. Allis, and G. Felsenfeld. (2001). Correlation between histone lysine methylation and developmental changes at the chicken β -globin locus *Science* 293: 2453-2455.
- W-H. Li, Z. Gu, H. Wang, and A. Nekrutenko. (2001). Evolutionary analyses of the human genome *Nature* 409: 847-849.
- SC. Low and MJ. Berry. (1996). Knowing when not to stop: selenocysteine incorporation in eukaryotes *Trends Biochem. Sci.* 21: 203-208.
- J. Lykke-Andersen, C. Aagaard, M. Semionenkova, and RA. Garrett. (1997). Archaeal introns: splicing, intercellular mobility and evolution *Trends Biochem. Sci.* 22: 326-331.
- T. Misteli. (2001). Protein dynamics: implications for nuclear architecture and gene expression *Science* 291: 843-847.
- T. Misteli, A. Gunjan, R. Hock, M. Bustin, and DT. Brown. (2000). Dynamic binding of histone H1 to chromatin in living cells *Nature* 408: 877-881. A. Muto, C. Ushida, and H. Himeno. (1998). A bacterial RNA that functions as both a tRNA and an mRNA *Trends Biochem. Sci.* 23: 25-29.
- KE. Nelson, RA. Clayton, and SR. Gill, *et al.* (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermatoga maritima* *Nature* 399: 323-329.
- MS. Neuberger and J. Scott. (2000). RNA editing AIDs antibody diversification *Science* 289: 1705-1706.
- HH. Ng and A. Bird. (2000). Histone deacetylases: silencers for hire *Trends Biochem. Sci.* 25: 121-126.
- J. Nosek, L. Tomáška, H. Fukuhara, Y. Suyama, and L. Kovác. (1998). Linear mitochondrial genomes: 30 years down the line *Trends Genet.* 14: 184-188.
- H. Ochman, JG. Lawrence, and EA. Groisman. (2000). Lateral gene transfer and the nature of bacterial innovation *Nature* 405: 299-304.
- SG. Oliver, QJM. van der Aart, and ML. Agostini-Carbone, *et al.* (1992). The complete DNA sequence of yeast chromosome III *Nature* 357: 38-46.
- JD. Palmer. (1997a). The mitochondrion that time forgot *Nature* 387: 454-455.
- JD. Palmer. (1997b). Organelle genomes: going, going, gone! *Science* 275: 790-791.
- J. Parkhill, M. Achtman, and KD. James. (2000a). Complete genome sequence of a serogroup A strain of *Neisseria meningitidis* Z2491 *Nature* 404: 502-506.
- J. Parkhill, BW. Wren, and K. Mungall, *et al.* (2000b). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences *Nature* 403: 665-668. ([PubMed](#))
- J. Parkhill, BW. Wren, and NR. Thomson, *et al.* (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague *Nature* 413: 523-527.

- C. Patience, DA. Wilkinson, and RA. Weiss. (1997). Our retroviral heritage *Trends Genet.* 13: 116-120.
- H. Paulus. (2000). Protein splicing and related forms of protein autoprocessing *Ann. Rev. Biochem.* 69: 447-496.
- E. Pennisi. (1997). Opening the way to gene activity *Science* 275: 155-157.
- E. Pennisi. (1999). The race to the ribosome structure *Science* 285: 2048-2051.
- R. Percudani. (2001). Restricted wobble rules for eukaryotic genomes *Trends Genet.* 17: 133-135.
- TV. Pestova and CUT. Hellen. (1999). Ribosome recruitment and scanning: what's new? *Trends Biochem. Sci.* 24: 85-87.
- S. Pietrokovski. (2001). Intein spread and extinction in evolution *Trends Genet.* 17: 465-472.
- T. Preiss and MW. Hentze. (1998). Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast *Nature* 392: 516-520.
- NT. Perna, G. Plunkett, and V. Burland, *et al.* (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7 *Nature* 409: 529-532.
- BD. Peterson-Burch, DA. Wright, HM. Laten, and DF. Voytas. (2000). Retroviruses in plants? *Trends Genet.* 16: 151-152.
- S. Radford. (2000). Protein folding: progress made and promises ahead *Trends Biochem. Sci.* 25: 611-618.
- RA. Reenan. (2001). The RNA world meets behavior: A→I pre-mRNA editing in animals *Trends Genet.* 17: 53-56.
- R. Sachidanandam, D. Weissman, and S. Schmidt, *et al.* (2001). The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms *Nature* 409: 928-933
- J. Scott. (1997). RNA editing: message change for a fat controller *Nature* 387: 242-243.
- M. Tucker and R. Packer. (2000). Mechanisms and control of mRNA decapping in *Saccharomyces cerevisiae* *Ann. Rev. Biochem.* 69: 571-595.
- M. Selmer, S. Al-Karadaghi, G. Hirokawa, A. Kaji, and A. Liljas. (1999). Crystal structure of *Thermatoga maritima* ribosome recycling factor: a tRNA mimic *Science* 286: 2349-2352.
- CK. Stover, XQ. Pham, and AL. Erwin, *et al.* (2000). Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen *Nature* 406: 959-964.
- P. Sudarsanam and F. Winston. (2000). The Swi/Snf family: nucleosome-remodelling complexes and transcriptional control *Trends Genet.* 16: 345-351.
- P. Syntichaki, I. Topalidou, and G. Thireos. (2000). The Gcn5 bromodomain co-ordinates nucleosome remodelling *Nature* 404: 414-417.

- H. Tettelin, KE. Nelson, and IT. Paulsen, *et al.* (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae* *Science* 293: 498-506.
- A. Travers. (1999). The location of the linker histone on the nucleosome *Trends Biochem. Sci.* 24: 4-7.
- J. Scott. (1997). RNA editing: message change for a fat controller *Nature* 387: 242-243.
- M. Tucker and R. Packer. (2000). Mechanisms and control of mRNA decapping in *Saccharomyces cerevisiae* *Ann. Rev. Biochem.* 69: 571-595.
- JC. Venter, MD. Adams, and EW. Myers, *et al.* (2001). The sequence of the human genome *Science* 291: 1304-1351.
- MA.Valencia-Sanchez *et al.*, 2006 Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* 17: 3011-3016.
- PE. Warburton. (2001). Epigenetic analysis of kinetochore assembly on variant human centromeres *Trends Genet.* 17: 243-247.
- O. White, JA. Eisen, and JF. Heidelberg, *et al.* (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1 *Science* 286: 1571-1577.
- Young-Kook Kim, N Narry Kim, (2007). Processing of intronic micro RNA. *EMBO Journal*, 26, 771-783
- J. Zhang. (2000). Protein-length distributions for the three domains of life *Trends Genet.* 16: 107-109.
- Z. Zhang, BR. Green, and T. Cavalier-Smith. (1999). Single gene circles in dinoflagellate genomes *Nature* 400: 155-159.
- L. Zhang, W. Zhou, and VE. Velculescu, *et al.* (1997). Gene expression in normal and cancer cells *Science* 276: 1268-1272.
- Y-B. Zhou, SE. Gerchman, V. Ramakrishnan, A. Travers, and S. Muyltermans. (1998). Position and orientation of the globular domain of linker histone H5 on the nucleosome *Nature* 395: 402-405.