

ФУНКЦИИ НЕКОДИРУЮЩИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНОМА МЛЕКОПИТАЮЩИХ

©2014 г. Л. И. ПАТРУШЕВ, Т. Ф. КОВАЛЕНКО

*Институт биоорганической химии им. академиков М.М.Шемякина
и Ю.А.Овчинникова РАН, Москва.*

I. Введение. II. Функции некодирующих последовательностей
гена человека и животных. III. Заключение.

I. ВВЕДЕНИЕ

Геном живого организма хранит и реализует в онтогенезе большую часть генетической (наследуемой) информации. Функционирование генома, проявляющееся, прежде всего, в эффективной экспрессии заключенных в нем генов, лежит в основе жизни на Земле. Для корректной работы генов, которые у эукариот представляют собой последовательности (ПП) нуклеотидов геномной ДНК, оказывающие

Принятые сокращения: 5mC – 5-метилцитозин; CAGE (Cap Analysis Gene Expression) – кэп-опосредованный анализ экспрессии генов; ceРНК – эндогенная конкурирующая РНК; CGI (CpG islands) – CpG-островки; ChIP – иммунопреципитация хроматина; CNV – вариант по числу копий; eРНК – энхансерная РНК; GWAS (Genome Wide Association Studies) – полногеномный поиск ассоциаций; HOR (higher order repeat) – повтор высокого порядка; HS-сайт – сайт, гиперчувствительный к ДНКазе; ICE (imprint control element) – элемент, контролирующий импринтинг; ICR (imprinting control region) – участок, контролирующий импринтинг; IGS (intergap segments) – межделеционные сегменты; IME (Intron-Mediated Enhancement) – опосредованное интронами усиление экспрессии генов; LCR – локус-контролирующая область; lincРНК – длинная межгенная некодирующая РНК; LINE (long interspersed elements) – длинные диспергированные элементы; lncРНК – длинная некодирующая РНК; miРНК – микроРНК; ncРНК – некодирующая РНК; NGS (Next generation sequencing) – секвенирование нового поколения; NIM (neutral indel model) – модель нейтральных делеций-вставок; NUMT (nuclear mitochondrial) – ядерно-митохондриальные (псевдогены); S/MAR (scaffold/matrix attachment regions) – матрикс-ассоциированный участок хроматина; shРНК – короткая РНК;

Окончание принятых сокращений: см. сл. стр.

Адрес для корреспонденции: patrush@mx.ibch.ru

Работа выполнена при финансовой поддержке гранта РФФИ 14-08-00801-А «Поиск и разработка новых генетических маркеров распространенных многофакторных заболеваний».

влияние на какой-либо фенотипический признак [1, 2], требуется их способность безошибочно отвечать на многочисленные эндогенные и экзогенные регуляторные сигналы. Выполнение этой функции генома обеспечивают регуляторные ПП, распознаваемые соответствующими регуляторными молекулами организма, также синтезируемыми под контролем генов. В этой связи полногеномный поиск регуляторных элементов генома необходим для понимания его организации и является одной из важнейших задач современной функциональной геномики [3].

Данные, полученные к настоящему времени при секвенировании генома человека, обнаружили несколько характерных особенностей его строения. Неожиданно оказалось, что ПП нуклеотидов ДНК, кодирующие экзонные области генов, составляют лишь небольшую его часть – 2.94%, а экзоны, кодирующие только аминокислотные ПП (без нетранслируемых экзонов и их частей), еще меньше – 1,2% [4]. Остальные ПП, составляющие 98% генома, в данном обзоре мы будем называть некодирующими ПП. Общее число генов оказалось также ниже ожидаемого (~25 000), однако механизмы, регулирующие их экспрессию, и регуляторные ПП демонстрировали исключительное разнообразие [5]. К тому же ПП индивидуальных геномов человека оказались высокополиморфными. Исчерпывающий анализ индивидуальных геномов, общее число которых в настоящее время приближается к 1.5 тысячам, выявил десятки миллионов однонуклеотидных замен (SNP и SNV), миллионы коротких вставок и делеций (indels), а также многочисленные протяженные ПП, варьирующие по числу копий (CNV) [6, 7]. Полученные результаты позволили сделать вывод о том, что первичная структура генома каждого чело-

Окончание принятых сокращений

SINE (Short interspersed elements) – короткие рассеянные элементы; siРНК – малые интерферирующие РНК; SNP – однонуклеотидный полиморфизм; SNV – однонуклеотидный вариант; snРНК – малые ядерные РНК; TAD (topologically associating domain) – топологически ассоциированный домен (хромосом); TERRA (Telomeric Repeat containing RNA) – РНК, содержащая теломерные повторы; TERT – каталитическая субъединица теломеразы; TR – теломеразная РНК; TSS (transcription start site) – точка инициации транскрипции; uORF (upstream open reading frame) – выше расположенная открытая рамка считывания; UTR – нетранслируемая область (РНК); vlincRNA – очень длинные межгенные некодирующие РНК; гяРНК – гетерогенные ядерные РНК; интРНК – интронные РНК; м.п.н. – миллион пар нуклеотидов; МГЭ – мобильные генетические элементы; мтДНК – митохондриальная ДНК; п.н. – пара нуклеотидов; ПП – последовательности; т.н. – тысяча нуклеотидов; т.п.н. – тысяча пар нуклеотидов; satДНК – α-сателлитная ДНК; тДНК – теломерная ДНК.

века уникальна. Действительно, секвенирование каждого нового генома обнаруживает в нем до 3 500 000 SNP и ~1000 больших (>500 п.н.) SNV, которые отсутствуют в референсной ПП. При этом до 500 000 SNP в каждом геноме являются неизвестными ранее. В среднем любой секвенированный геном внешне здорового человека содержит до 25 000 SNP в его кодирующих частях. Из них до 10 000 представляют собой несинонимические замены нуклеотидов, среди которых до 100 SNP в гетерозиготном состоянии уже отнесены к вредным мутациям, приводящим к развитию патологий. Основная же часть полиморфных нуклеотидов локализована в некодирующих частях генома, функции большей части которых неизвестны. Предполагается, что в совокупности индивидуальные особенности строения генома, в том числе и его некодирующих ПП, могут определять многие характерные черты жизнедеятельности организма человека, хотя и далеко не однозначно [8]. В настоящее время эта концепция развивается в рамках большой нерешенной генетической проблемы взаимодействия генотипа и фенотипа [9].

Представление о важной роли генетических факторов в этиологии сложных многофакторных заболеваний находит подтверждение в многочисленных современных исследованиях ассоциированности обнаруженных SNP с развитием патологического процесса. В последнее время активно развивается и приносит плоды система полногеномного поиска таких ассоциаций (Genome Wide Association Studies – GWAS) [10]. При таком подходе в группах «контроль–клинический случай», которые насчитывают сотни тысяч участников, с использованием биочиповых технологий одновременно исследуют ассоциированность с конкретными заболеваниями миллионами SNP, клиническая значимость для большинства из которых неизвестна. В результате выявлено более 12 000 новых генетических локусов, ассоциированных с исследованными патологическими состояниями, что в ряде случаев привело к изменению представлений о молекулярных механизмах развития болезней (www.genome.gov/gwastudies/). Однако интерпретация результатов, полученных методом GWAS, сталкивается со значительными затруднениями. Связь SNP с болезнями оказывается слабой (отношение шансов, как правило, не превышает 1.2), а большинство самих SNP также локализовано в некодирующих ПП генома [10]. Это позволяет предположить, что изменения генома, обусловленные данными SNP, на самом деле связаны с функциональными ПП, участвующими в регуляции экспрессии генов, и могут нарушать межгенные взаимодействия.

На более важную, чем ранее представлялось, функциональную роль некодирующих ПП генома эукариот может указывать и открытое недавно явление «всепроникающей транскрипции» [11]. Анализ транскриптома человека современными методами показал, что большая часть ПП в исследованных эукариотических геномах транскрибируется. Кроме того, высокая филогенетическая консервативность многих некодирующих участков генома у видов, в эволюционном плане далеко отстоящих друг от друга, также позволяет предположить их функциональную значимость [12].

В обзоре всесторонне проанализирована проблема функциональности некодирующих ПП в геноме млекопитающих. Приведены данные в пользу функциональной значимости большей части некодирующих ПП, кратко рассмотрены особенности функционирования основных регуляторных элементов генома, включая 5'- и 3'-UTR, интронов, энхансеров транскрипции, инсуляторов, участков метилирования ДНК, S/MAR-ПП, генов некодирующих РНК, а также повторяющихся ПП. Приведены сведения о мутациях в некодирующих частях генома, ассоциированных с заболеваниями человека, подтверждающие функциональную значимость некодирующих ПП.

II. ФУНКЦИИ НЕКОДИРУЮЩИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНОМА ЧЕЛОВЕКА И ЖИВОТНЫХ

Проблема «избыточности» генома человека в отношении ПП нуклеотидов ДНК, некодирующих белки, остается нерешенной. До сих пор представляется загадочной эволюционная консервативность гигантского размера геномов человека (размер его гаплоидного генома составляет 3.3×10^9 п.н.) и многих других эукариот, насыщенных некодирующими ПП. Попытки объяснить это явление привели к созданию многочисленных теорий, которые распадаются на две большие группы. Приверженцы одной из них считают, что большинство некодирующих ПП нефункционально и отражает эволюционное накопление в геноме обломков теперь уже ненужных генов или «эгоистически» размножающейся ДНК. Другая группа исследователей, к которой относят себя и авторы данного обзора, напротив предполагает функциональность большей части некодирующих ПП [13, 14].

С появлением референсной ПП генома человека выявление всех функционально значимых участков генома представляется реальной задачей. Для ее решения в 2003 году был создан международный консорциум по созданию энциклопедии функциональных элементов

ДНК (ENCyclopedia Of DNA Elements – ENCODE, <http://genome.ucsc.edu/ENCODE/>), задачей которого является поиск и составление исчерпывающего каталога всех функциональных элементов генома человека, контролирующих экспрессию генов. Для обнаружения функционально-значимых ПП в рамках консорциума и независимыми исследователями используют три группы методов: эволюционные, биохимические и генетические [15]. При эволюционном подходе методами сравнительной структурной геномики выявляют эволюционно консервативные участки генома млекопитающих, находящиеся под давлением естественного отбора. Биохимики получают доказательства их молекулярной активности, а генетики фиксируют фенотипические изменения организма в ответ на мутационные перестройки первичной структуры геномной ДНК. Все три подхода являются высокоинформативными, взаимно дополняют друг друга, хотя их вклад в аннотирование генома различен. Анализ мутационных изменений ДНК, приводящих к развитию заболеваний, также вносит вклад в понимание функциональной роли некодирующих ПП.

ЭВОЛЮЦИОННАЯ КОНСЕРВАТИВНОСТЬ НЕКОДИРУЮЩИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНОМА

При эволюционном подходе к определению функционально значимых ПП ДНК в геноме человека используют параметр α_{sel} , которым обозначают долю нуклеотидов генома, подверженных действию очищающего (отрицательного) отбора, среди всех нуклеотидов генома [16]. Для определения этого параметра в настоящее время применяют две группы методов. Во-первых, в сравнительных исследованиях выявляют консервативные геномные ПП у далеко дивергировавших видов, предполагая, что ПП, которые выполняют важные функции, остаются неизменными на протяжении длительного эволюционного периода. Данный прием мало применим для анализа быстро эволюционирующих регуляторных участков генома, для чего используют модель нейтральных делеций-вставок (neutral indel model – NIM) [17]. В этом случае осуществляют полногеномную оценку размеров фрагментов ДНК, располагающихся между двумя соседними indel-сайтами, получивших название межделеционных сегментов (intergap segments – IGS). При анализе целого генома распределение по длинам IGS сильно смещено в сторону протяженных ПП, что резко отличается от картины распределения длин IGS, расположенных в нейтрально эволюционирующих участках генома. Это интерпретируется как указание на наличие отбора, направленного на сохранение в геноме соответствующих длинных ПП, и, как следствие, на их возможную функциональность.

При первом сравнении ПП геномов человека и мыши путем их выравнивания друг относительно друга было установлено, что среди коротких сегментов генома длиной в 50–100 п.н. доля ПП, подверженных давлению отрицательного отбора, составляет ~5% [18].

В недавнем исследовании эволюционной консервативности ПП геномной ДНК геном человека сравнивали с 28 ПП геномов максимально дивергировавших плацентарных млекопитающих [19]. Большая часть геномов с неизвестной к началу исследования первичной структурой была секвенирована в этой же работе. В результате компьютерного выравнивания сравниваемых ПП друг относительно друга было установлено, что приблизительно 5.5% генома человека несет следы очищающего (отрицательного) отбора, а 4.2% включает консервативные, предположительно функциональные, ПП. Эти консервативные элементы ДНК, общим числом ~3.6 млн, были представлены в наибольшей степени экзонами, входящими в состав мРНК (~30%), интронами генов (~30%), а также межгенными участками (~40%). Более чем в четверти всех известных генов человека в экзонах были обнаружены запреты на синонимические замены нуклеотидов. В этом случае консерватизм ПП может поддерживаться, например, необходимостью сохранения определенной структуры мРНК для прохождения сплайсинга, взаимодействия с регуляторными микроРНК (miРНК) или ее редактирования на посттранскрипционном уровне. Было обнаружено около 4000 новых экзонов-кандидатов, локализованных в транскриптах генов, кодирующих белки, их интронах и нетранслируемых ПП (UTR) мРНК. Найдено более 1000 консервативных ПП, предположительно ассоциированных с новыми классами некодирующих РНК, а также 2.7 млн коротких, предположительно регуляторных, ПП. Кроме того, до 40% нуклеотидов выявленных консервативных ПП остались неохарактеризованными с точки зрения их возможной функциональности.

При использовании усовершенствованного вышеупомянутого подхода NIM, который позволяет выявлять быстро эволюционирующие участки генома, было установлено, что до 8–9% общего числа нуклеотидов генома человека находятся под давлением отрицательного отбора и могут быть функциональны [20]. При этом только около 2% ПП оказались консервативными при сравнении геномов человека и мыши. На основании такого рода данных был сделан вывод о том, что в исследованных геномах имеется множество короткоживущих видоспецифических ПП, подверженных давлению негативного отбора. Большая часть этих ПП также оказалась локализованной в некодирующих частях геномов, и была отнесена

к известным функциональным элементам, в том числе к участкам, гиперчувствительным к ДНКазе I, сайтам связывания факторов транскрипции, промоторам, энхасерам и генам длинных некодирующих РНК (long non-coding RNA – lncРНК). Последние оказались эволюционно наиболее нестабильными.

Таким образом, современные методы секвенирования ДНК и биоинформатики позволили по-новому рассмотреть проблему «некодирующих ПП», которые теперь представляются, напротив насыщенными потенциально функциональными элементами генома. Дальнейшее развитие методической базы молекулярной биологии подтвердило и расширило это новое представление.

ВСЕПРОНИКАЮЩАЯ ТРАНСКРИПЦИЯ

Первое указание на то, что большая часть генома человека транскрибируется, было получено при исследовании транскрипции небольших хромосом 21 и 22 и подтверждено консорциумом ENCODE при попытке проведения исчерпывающего анализа функциональной активности 1% генома человека [5, 21–24]. Оказалось, что совокупно в клетках разных тканей человека транскрибируется с разной эффективностью по меньшей мере 93% нуклеотидов генома. Неаннотированные транскрипты с неизвестной функцией были названы «темной материей» [25]. Получение этих данных оказалось возможным, в основном, благодаря применению двух современных методов исследования транскриптома: использования гибридизации кДНК с упорядоченными олигонуклеотидными зондами, организованными в виде биочипов высокой плотности (tiling arrays), и RNA-seq-модификации методов секвенирования ДНК нового поколения (Next Generation Sequencing – NGS) [26–28].

Современные биочипы могут содержать миллионы олигонуклеотидных зондов, ПП которых перекрывают весь геном человека. При анализе транскриптома суммарную РНК, выделенную из небольшого числа клеток исследуемых тканей, освобождают от продуктов активно транскрибирующихся генов (особенно рРНК) и используют для синтеза кДНК, которую далее амплифицируют в ПЦР, фрагментируют, метят флуоресцентными красителями и гибридизуют с биочипом. Флуоресцентные сигналы считывают и анализируют на компьютере в автоматическом режиме. В принципе, биочиповые технологии позволяют проводить до 10^9 индивидуальных гибридизаций на одном чипе и получать уникальную информацию об особенностях транскрипции больших геномов, в том числе, осуществлять картирование транскриптов с разрешением в несколько

п.н. [23, 24]. Несмотря на выдающиеся возможности биочиповых технологий, они не лишены ряда недостатков, среди которых необходимо отметить значительную стоимость индивидуальных экспериментов, необходимость знания ПП анализируемого генома, высокий фон от неспецифических перекрестных гибридизаций, а также ограниченный диапазон количественных измерений уровней транскрипции генов, связанный с вышеупомянутым фоном и быстрым насыщением гибридизационного сигнала [29, 30].

Методология RNA-Seq (RNA sequencing) успешно конкурирует с биочиповыми технологиями в исследованиях транскриптома. Основное преимущество этого подхода заключается в непосредственном определении ПП нуклеотидов исследуемых кДНК высокопроизводительными методами NGS и прямом подсчете индивидуальных молекул кДНК в исследуемых образцах. RNA-Seq позволяет как картировать транскриптом, так и количественно оценивать уровни транскрипции индивидуальных участков генома. При осуществлении анализа транскриптома этими методами суммарную или фракционированную РНК используют для получения библиотеки фрагментов кДНК, к которым с одного или двух концов лигированием присоединяют олигонуклеотидные адаптеры, что позволяет одновременно, параллельно амплифицировать на твердом носителе и/или непосредственно секвенировать иммобилизованные на подложке миллионы фрагментов исследуемой кДНК с одного или двух концов с использованием только одного или двух олигонуклеотидных праймеров [31, 32]. Применение RNA-Seq-методов в функциональной транскриптомике, в принципе, не требует знания первичной структуры исследуемых геномов и позволяет выявлять структурные варианты (например, SNP) транскриптов [33]. Кроме того, для этой группы методов, в отличие от биочиповых технологий, характерен очень низкий уровень фоновых сигналов. Основными же их недостатками являются артефакты, связанные с приготовлением библиотек кДНК и остающаяся желать лучшего эффективность методов биоинформатики, разработанных для анализа больших массивов данных [34]. О других подходах к исследованию транскриптома см. в работах [35, 36].

Феномен всепроникающей транскрипции в настоящее время надежно подтвержден экспериментально [11, 37]. Во всех исследованных клетках и тканях человека, здоровых и опухолевых, молекулы РНК неизвестных функций составляют большую часть суммарных препаратов РНК, освобожденных от рРНК и митохондриальной РНК [38]. Обнаружены сотни геномных областей, в которых транскрибируются очень длинные межгенные некодирующие РНК

(vlincRNA), размеры которых превышают 100 т.н., и такие области распространяются на ~4% межгенных ПП генома. При этом наборы vlincRNA оказываются тканеспецифичными, и большинство геномных ПП представлено в суммарной РНК, выделенной из многих тканей. При секвенировании отдельных молекул РНК неожиданно оказалось, что доля интронных (инт) РНК составляет 30–50% в суммарной РНК человека, освобожденной от рРНК и мтРНК. При этом разные гены и даже отдельные участки в интронах различались по способности генерировать интРНК [38]. Исчерпывающее функциональное значение интРНК еще предстоит выяснить, однако, уже сейчас известно, что в них заключена приблизительно половина ПП известных miРНК, о которых речь пойдет ниже, в соответствующем разделе нашего обзора.

Значительный вклад во всепроникающую транскрипцию вносят мобильные генетические элементы (МГЭ) генома человека, которые составляют ~45% всех его ПП. Исследование глобальной транскрипции повторяющихся ПП долгое время сдерживалось трудностями выявления конкретных транскриптов из-за их перекрестной гибридизации на биочипах. Разработка системы кэп-опосредованного анализа экспрессии генов (Cap Analysis Gene Expression – CAGE) совместно с методами NGS позволило решить эту методическую проблему [39, 40]. Оказалось, что до 30% кэпированных транскриптов берут свое начало в повторяющихся ПП генома человека. Всего же в геномных ретротранспозонах было обнаружено ~250 000 точек инициации транскрипции (Transcription Start Site – TSS), характер их распределения в геноме демонстрировал тканеспецифичность, и они были кластеризованы в участках генома, насыщенных генами.

В целом, несмотря на отсутствие знаний о функциональной роли большинства вновь обнаруженных РНК транскриптома млекопитающих, уже сегодня ясно, что многие некодирующие РНК выполняют важные функции в регуляции экспрессии генов и поддержании стабильности генома. Структурно-функциональные особенности основных классов некодирующих РНК будут рассмотрены нами ниже в соответствующих разделах обзора.

CIS-ДЕЙСТВУЮЩИЕ РЕГУЛЯТОРНЫЕ ЭЛЕМЕНТЫ ТРАНСКРИПЦИИ И ТРАНСЛЯЦИИ

Регуляция экспрессии генов, кодирующих белки, осуществляется на многих уровнях, которые включают, транскрипцию, ко- и постранипционный процессинг мРНК, трансляцию и посттрансляционный процессинг синтезированных полипептидов [41]. В контроле

транскрипции и трансляции важную роль играют некодирующие ПП, получившие название регуляторных элементов генома. К таким элементам относятся промоторы, ПП, кодирующие регуляторные сигналы 5'- и 3'-UTR мРНК, сайты и энхансеры/сайленсеры сплайсинга. Кроме того, в эту категорию попадают регуляторные элементы, осуществляющие глобальную регуляцию транскрипции: энхансеры, локус-контролирующие области, инсуляторы, S/MAR-ПП, специфические участки метилирования ДНК и т.п. Большинство этих элементов располагается в непосредственной близости от контролируемых генов, в составе одних с ними доменов хромосом, и, следовательно, осуществляют регуляторное действие *in cis*, то есть внутри контролируемой области генома. Ниже будут рассмотрены основные особенности структуры и функционирования некодирующих *cis*-действующих регуляторных элементов генома, которые составляют его существенную и неотъемлемую часть.

5'-Концевые регуляторные элементы генов

Промоторы. Генные промоторы представляют собой специализированные ПП геномной ДНК, которые включают TSS – первый нуклеотид (положение +1), с которого начинается синтез РНК [42]. Функциональное значение промоторов заключается в обеспечении сборки инициационного комплекса с последующей инициацией транскрипции, ключевым компонентом которого является ДНК-зависимая РНК-полимераза. У человека гены белков и многих некодирующих РНК транскрибируются РНК-полимеразой II.

Область в окрестностях TSS разделяют на выше расположенный протяженный проксимальный промотор и минимальный (или коровый) промотор, окружающий TSS. В проксимальном промоторе локализованы сайты связывания некоторых «проксимальных» факторов транскрипции, которые, в свою очередь, могут объединяться в кластеры *cis*-регуляторных модулей.

Коровым промотором называют минимальную ПП ДНК, способную обеспечивать точную инициацию транскрипции РНК-полимеразой II [43]. Очищенная РНК-полимераза II неспособна распознавать коровые промоторы и для реализации ее функционального потенциала требуется участие основных (базальных) факторов транскрипции [44]. К ним относятся TFIIA (transcription factor, RNA polymerase II, A), TFIIВ, TFIIС, TFIIЕ, TFIIF и TFIIH. Распознавание коровой ПП обеспечивает фактор TFIIС, который включает в себя ТАТА-боксовывающий белок (ТВР) и ТВР-ассоциированный фактор (ТАФ). В присутствии этих факторов РНК-полимераза II взаимодействует с

кóровым промотором, входя в состав большого прединициационного комплекса (PIC).

Функциональные свойства кóрового промотора определяют составляющие его специфические ПП, получившие название элементов (или мотивов) корового промотора. В настоящее время не обнаружено элементов, которые были бы универсальны для всех промоторов. Самым распространенным элементом коровых промоторов является инициатор (Inr), включающий в себя TSS (консенсусная ПП: YYANWYY (где A+1)), с которым прежде всего взаимодействует TFIID. Консенсусная ПП ТАТА-бокса – ТАТАWAAR, 5'-концевой Т которой располагается в положении –30 или –31 от A+1 Inr. С этой ПП взаимодействует TBP-субъединица фактора TFIID. Два BRE-элемента (TFIIB recognition element) могут располагаться как перед ТАТА-боксом (BREu), так и после него (BREd), с которым они функционально связаны и могут оказывать на базальный уровень транскрипции как позитивное, так и негативное влияние. Близко расположенные друг к другу и иногда даже перекрывающиеся элементы DPE (downstream core promoter element) и MTE (motif ten element) контактируют с субъединицами TAF6 и TAF9 фактора TF2D. Эти ПП содержат четыре области в положениях 18–22 (CGANC) и 27–29 (CGG) у MTE, а также 27–29 и 30–33 у DPE (относительно A+1 Inr), особенно важные для их функционирования. При этом, наличие первой и третьей областей достаточно для получения функционального корового промотора, хотя такое сочетание редко обнаруживается в природных промоторах.

Помимо указанных базовых элементов, промоторы некоторых генов могут содержать специфические последовательности, необходимые для связывания определенных активирующих факторов. К таким элементам относится, например, E-бокс с канонической ПП SACGTG. E-бокс является сайтом связывания регуляторных белков семейства «спираль–петля–спираль» и присутствует во многих генах, экспрессирующихся в мышечной, нервной ткани и в поджелудочной железе [45]. Недавние исследования показали, что E-бокс или E-боксоподобные элементы обнаруживаются в промоторах так называемых clock-генов, экспрессия которых подчиняется циркадным ритмам (то есть, циклически изменяется в течение суток) [46]. Другой пример – элементы теплового шока (heat shock elements – HSEs), которые содержатся в промоторах генов белков теплового шока. Данные элементы являются сайтами связывания транскрипционных факторов, известных как факторы теплового шока. HSEs присутствуют в виде нескольких копий пентануклеотидной последовательности NGAAN-3', которые

расположены в различной ориентации относительно друг друга. Количество пятинуклеотидных мономерных единиц может быть различным в промоторах разных генов, однако для успешной активации транскрипции таких мономеров должно быть не менее трех [47].

По нуклеотидному составу и функциям у млекопитающих различают три типа промоторов. К типу I относят промоторы с низким содержанием GC, включающие ТАТА-бокс. У большинства промоторов этого типа инициация транскрипции ограничивается одним или несколькими близко расположенными нуклеотидами (сфокусированная транскрипция), а ТАТА-бокс расположен на определенном расстоянии от TSS, что в сочетании с консенсусной ПП инициатора определяет выбор TSS на ДНК. Промоторы типа I направляют тканеспецифическую транскрипцию во взрослом организме и метилированы в активном состоянии.

Для промоторов типа II и III характерно высокое содержание GC, а также наличие множественных TSS, расположенных на участке 50–100 нт (дисперсная транскрипция). У промоторов типа II имеются короткие CpG-островки (CGI), отсутствует ТАТА-бокс, и они направляют транскрипцию большинства генов «домашнего хозяйства». Промоторы типа III содержат протяженные CGI, распространяющиеся на ПП генов, маркированы триметилированными гистонами H3K27me₃, и регулируются системой репрессии Polycomb. Промоторы этого типа управляют генами, участвующими в регуляции развития и дифференцировки клеток.

Отдельную нишу занимают, так называемые, TCT-промоторы, характерные для высокоэкспрессирующихся генов рибосомных белков. Они получили свое название от характерной структуры инициатора, который содержит полипиримидиновую ПП YUC+1TUTTYYU (TSS соответствует +1T) [48].

5'-Концевые ПП некоторых генов могут содержать более одного промотора. Такие альтернативные промоторы одного и того же гена могут функционировать в разных тканях или на разных этапах развития организма [49]. Несмотря на то, что в результате синтеза РНК с альтернативных промоторов образуются транскрипты разного размера, кодируемые белки могут быть идентичны как следствие использования одного и того же сайта инициации трансляции [49]. Другой тип альтернативных промоторов – промоторы, в результате активности которых образуются различные изоформы белков с разными свойствами. Примером могут служить промоторы гена метилтрансферазы DNMT1. 5'-концевая область данного гена включает три промотора. В соматических клетках активен промотор 1s. В

ооцитах функционирует промотор 1 α . При этом трансляция происходит с ATG-кодона, расположенного в экзоне 4. Образующийся белок короче фермента, синтезирующегося в соматических клетках на 118 N-концевых аминокислот. В сперматозоидах активен промотор 1 β . Однако образующаяся при этом мРНК не транслируется. В результате, сперматозоиды не содержат фермента DNMT1 [50].

Геном эукариот, и в частности человека, содержит множество генов, расположенных в непосредственной близости друг от друга и транскрибируемых с разных цепей одной и той же молекулы ДНК в противоположных направлениях. Промоторные области таких генов частично перекрываются, а расстояние между их TSS не превышает 1000 п.н. Участок молекулы ДНК, расположенный между TSS двух генов, транскрибируемых с разных цепей, получил название двунаправленного промотора [51, 52]. Подобным образом, в частности, организована значительная часть генов системы репарации ДНК и гены шаперонов. В большинстве случаев двунаправленные промоторы обеспечивают одновременную экспрессию двух генов, однако иногда инициация транскрипции одного гена из генной пары вызывает подавление транскрипции другого [51]. Первичная структура двунаправленных промоторов имеет некоторые особенности. Например, только 9% таких промоторов содержат ТАТА-бокс. Усредненное содержание GC-пар в двунаправленных промоторах несколько выше, чем в обычных (66% и 55% соответственно).

Мутации легко нарушают функциональную целостность промоторов из-за их перегруженности регуляторными элементами, взаимодействующими с факторами транскрипции. Например, ген *TERT* кодирует каталитическую обратнотранскриптазную субъединицу фермента теломеразы, поддерживающей длину теломер (см. ниже раздел о теломерах). Повышенная теломеразная активность является одной из характерных особенностей опухолевых тканей. Промотор гена *TERT* содержит множество сайтов связывания различных активаторов и репрессоров. Мутации A-57C, C-124T и C-146T создают новый перекрывающийся сайт связывания для факторов транскрипции Ets и TCF, что повышает уровень экспрессии гена *TERT*. Указанные соматические мутации являются взаимоисключающими и встречаются с разной частотой при различных онкологических заболеваниях. При раке мочевого пузыря наиболее распространенной является мутация C-124T (встречается у 53,5% больных). В целом, указанные три мутации встречаются при раке мочевого пузыря у 65,4% пациентов, а при глиомах – у 80% больных. Данные генетически изменения ассоциированы с более агрессивным течением заболевания и неблагоприятным прогнозом [53].

В еще одном примере при Лейденской гемофилии В, которая характеризуется низким содержанием фактора IX свертывания крови (максимум – 60% от нормы), SNP G26C и T20A в промоторе гена *F9* нарушают перекрывающийся сайт связывания TATA/HNF-4. HNF-4 – ядерный фактор гепатоцитов, контролирующий экспрессию гена фактора IX. Неспособность HNF-4 связываться с измененным сайтом приводит к гемофилии [54].

5'-Концевые нетранслируемые области. 5'-Концевая нетранслируемая область (5'-UTR), как правило, кодируемая первым экзоном эукариотических генов, включает в себе ПП, обеспечивающие регулируемую трансляцию мРНК. Средняя длина 5'-UTR у человека составляет 210 нт, а минимальная – 18 нт. Максимальная длина этой ПП у человека отмечена для мРНК онкогена *Tre* – 2858 нт [55]. Длина 5'-UTR влияет на эффективность трансляции мРНК, поскольку инициаторный AUG-кодон удален от кэп-группы m7G, на которой происходит сборка рибосомы, и для его достижения рибосоме приходится преодолевать этот часто высокоструктурированный участок матрицы. Для 5'-UTR характерен высокий GC-состав. От 10% до 15% генов млекопитающих используют альтернативные 5'-UTR как следствие инициации транскрипции на разных промоторах. Тот же эффект в 13% генов достигается за счет альтернативного сплайсинга [56]. Использование альтернативных 5'-UTR имеет важные регуляторные последствия для трансляции и экспрессии генов.

Вторичная структура 5'-UTR играет большую роль в регуляции экспрессии генов, в частности, факторов транскрипции и факторов роста, а также протоонкогенов и их рецепторов. Более 90% их мРНК имеют в своих 5'-UTR ПП, образующие стабильные элементы вторичной структуры, как правило, расположенные недалеко от кэп-группы. С ними взаимодействуют регуляторные белки, что может сопровождаться ингибированием трансляции на этих мРНК. РНК-связывающие белки (RNA binding protein – RBP), которых у человека известно более 1000, разделяются на две большие группы: 1) RBP, необходимые для трансляции всех мРНК и 2) специальные RBP, регулирующие трансляцию некоторых мРНК. RBP, взаимодействующие с одними и теми же ПП, могут выполнять антагонистические функции ингибиторов или активаторов трансляции, например, CUGBP1 и кальретикулин, регулирующие экспрессию p21, который участвует в регуляции клеточного цикла [57].

К основным регуляторным элементам в 5'-UTR относятся и так называемые выше расположенные рамки считывания (upstream open reading frame – uORF), которые представляют собой короткие ORF,

расположенные перед основной рамкой считывания, со своим кодонным инициации, но без кодона терминации трансляции. До 50% РНК транскриптома человека и мышей содержат uORF, которые весьма гетерогенны по размерам и расположению [58]. Функциональная активность uORF определяется расстоянием от кэп-группы, ее первичной структурой, контекстом, в котором рамка находится, ее размером, эффективностью ее сайта инициации трансляции и числом расположенных в ней AUG-кодонов. uORF модулируют сканирование 5'-конца мРНК рибосомами, при котором uAUG-кодоны выступают в качестве ловушек. Кроме того рибосомы на разное время задерживаются в конце uORF по завершении их трансляции с образованием аттенуаторных пептидов, взаимодействующих с рибосомами [59]. Многие патологии человека связаны с мутациями, влияющими на uORF, включая предрасположенность к меланомам и раку груди, наследственные тромбоцитемии, болезнь Альцгеймера, гипотрихоз и многие другие (см. ссылки в обзорах [58, 60]).

3'-Концевые регуляторные элементы генов

3'-Концевые некодирующие элементы генов, входящие в состав 3'-концевых нетранслируемых областей (3'-UTR) мРНК, выполняют важные регуляторные функции на посттранскрипционном и трансляционном уровнях экспрессии эукариотических генов [61]. Начинаясь на кодоне терминации трансляции, эти ПП участвуют в процессинге мРНК, контролируют их стабильность и локализацию транскриптов в клетке, а также оказывают влияние на скорость трансляции. Анализ 3'-UTR генов человека выявил значительную гетерогенность по размерам. Их длина в среднем составляет 1.3 т.п.н. или ~36% длины зрелой мРНК (без поли(А)-тракта), хотя часто может превышать и 5 т.п.н. (наибольший зафиксированный размер – 8.5 т.п.н.) [55, 62]. Это почти вдвое превышает средние размеры 3'-UTR других млекопитающих, что по-видимому, связано с наличием в них большего числа регуляторных ПП и более тонкой регуляцией трансляции в организме человека.

По завершению синтеза пре-мРНК ее сигналы полиаденилирования (поли(А)-сайты) с консенсусной ПП AAUAAA направляют сборку рибонуклеопротеинового комплекса из ~85 белков, который укорачивает пре-мРНК с 3'-конца и обеспечивает присоединение поли(А)-ПП длиной ~250 нт. Приблизительно 1/3 мРНК содержит два и более поли(А)-сайта, которые могут использоваться во время процессинга пре-мРНК как альтернативные. Поли(А)-ПП является сигналом для объединения мРНК с регуляторными поли(А)-связы-

вающими белками (poly(A) binding proteins – PABP), что необходимо для экспорта мРНК из ядра, регуляции ее стабильности, надзора за ошибками транскрипции, а также позитивной и негативной регуляции трансляции рибосомами, в том числе и с участием миРНК (обзор [63]). Во время инициации трансляции имеет место синергизм в действии 5'-концевой кэп-группы и 3'-UTR, что становится возможным благодаря их сближению в результате циркуляризации мРНК. В случае мРНК гена *p53* человека в ее 3'-UTR имеется ПП, комплементарная 5'-UTR той же мРНК, взаимодействующая с фактором трансляции RPL26, который стимулирует трансляцию этой мРНК в ответ на повреждение ДНК в клетке [64].

3'-UTR являются одними из основных мишеней, на которые действуют регуляторные miРНК и которые могут не только ингибировать, но и стимулировать трансляцию соответствующих мРНК, облегчая их циркуляризацию (см. ниже соответствующий раздел обзора).

Необходимость 5'- и 3'-концевых взаимодействий в мРНК, а также взаимодействий с miРНК при их трансляции накладывает определенные требования к длине и вторичной структуре 3'UTR. В общем, более длинные 3'-UTR ассоциированы с меньшим уровнем трансляции их мРНК, что, в частности, связано с наличием в них большего числа сайтов связывания miРНК. Так, некоторые изоформы мРНК, например транскрипты гена *Hip2*, используют альтернативные 3'-UTR разной длины, которые различаются по числу сайтов связывания miРНК [65].

AU-богатый элемент (AU-rich element – ARE), локализованный в 3'-UTR, содержащий одну или несколько (до пяти) ПП типа AUUUA, контролирует стабильность мРНК [66]. Нономер UUAUUUAUU или его аналог UUAUUUAWW являются минимальными ПП, обеспечивающими дестабилизацию, а 13-звенная ПП WWWUAUUUAUWWW рассматривается в качестве минимального ARE у человека. Увеличение числа копий пентамерного мотива в ARE снижает стабильность мРНК. Многофункциональные белки, взаимодействующие с ARE, ускоряют уменьшение длины поли(A)-ПП до 30–60 нт, что является критическим для стабильности мРНК, после чего этот процесс вступает во вторую фазу, когда происходит быстрая деградация центральной части мРНК. Функциональность ARE-связывающих белков, в свою очередь, регулируется посттрансляционными модификациями (фосфорилированием) или в результате направленной внутриклеточной локализации. При этом miРНК также являются регуляторами ARE.

Кроме вышеупомянутых мотивов в 3'-UTR ряда генов присутствуют в виде нескольких копий похожие по механизму действия CU-богатый элемент (CU-rich element – CURE) и элемент контроля дифференцировки (differentiation control element – DICE), которые взаимодействуют с гетерогенными ядерными (гЯ) РНП, что сопровождается ингибированием инициации трансляции из-за специфического подавления сборки 80S рибосом. Фосфорилирование гЯРНП при передаче сигналов активирует трансляцию.

GU-богатые элементы (GU-rich element – GRE) 3'-UTR, составленные из 2–5 перекрывающихся GUUUG-пентамеров, присутствуют в 5% мРНК человека [67]. Находясь в связанном с белками состоянии, GRE принимает участие в деаденилировании мРНК, их деградации и сплайсинге.

CA-богатые элементы (CA-rich element – CARE) представляют собой одни из наиболее часто встречающихся динуклеотидных повторов в геноме человека, как в кодирующих, так и некодирующих ПП. Они оказывают стабилизирующее действие на мРНК после взаимодействия с гЯРНП L.

Элементы, чувствительные к ионам железа (iron responsive element – IRE), контролирующие метаболизм железа, обнаружены в 5'- и 3'-UTR 10 генов человека, в частности, генов легких (FTL) и тяжелых (AEP1) полипептидов ферритина, рецептора трансферрина 1 и др. Для IRE-ПП длиной 26–30 нт характерна вторичная структура типа «стебель–петля», с которой взаимодействуют два белка, регулирующие метаболизм железа – (iron regulatory protein 1 – IRP1) и IRP2. IRP1 связывается с 5'-UTR и подавляет трансляцию, предотвращая соединение малой и большой субчастиц рибосом. IRP2, взаимодействуя с 3'-UTR, стабилизирует мРНК [68].

ПП 3'-UTR, содержащая элемент, обеспечивающий вставку селеноцистеина (selenocysteine insertion sequence element – SECIS), образует структуру типа «стебель–петля», которая распознается белками, обеспечивающими декодирование стоп-кодона UGA, как кодона незаменимой аминокислоты селеноцистеина [69].

В заключение следует отметить, что многие 3'-UTR человека содержат ретроэлементы *Alu*, с которыми взаимодействуют lncРНК, содержащие аналогичную комплементарную ПП. В результате создается участок двухцепочечной РНК, с которым связывается белок Staufen-1 (Stau1), что направляет мРНК-мишень на путь деградации. Кроме того, сайт, формирующий в РНК внутримолекулярную двухцепочечную структуру, с которым взаимодействует Stau1, может присутствовать и в самой 3'-UTR [70].

Насыщенность 3'-UTR функциональными элементами делает их, как и промоторы, весьма чувствительными к мутационным заменам нуклеотидов, которые сопровождаются развитием многих патологических состояний организма человека, включая дисфункции иммунной системы и врожденные сердечнососудистые заболевания [66, 60]. Например, протромбин (фактор II свертывания крови) является предшественником тромбина. Последний играет главную роль в превращении фибриногена в фибрин при образовании тромба и других реакциях, обеспечивающих функционирование системы гемостаза. Мутация G20210A в гене протромбина затрагивает его 3'-UTR, что сопровождается увеличением содержания протромбина в плазме в 1,5–2 раза. Это, в свою очередь, повышает риск развития тромбоза в 3 раза [71].

Полиморфизм С/Т в 3'-UTR гена рецептора эстрогена 1 (ESR-1) затрагивает сайт связывания микроРНК miR-453. Замена С на Т уменьшает эффективность связывания miR-453 с данным сайтом. В результате уровень экспрессии гена ESR-1 повышается, что ассоциировано с риском развития рака молочной железы [72].

Участки метилирования ДНК

Остатки цитозина в ДНК человека могут быть метилированы с высокой специфичностью по положению 5 пиримидинового кольца с образованием 5-метицитозина (5mC), главным образом, в последовательностях CpG, которые в геноме распределены неслучайным образом (обзор [73]). Метильные группы 5mC располагаются в большой бороздке ДНК, что делает их доступными для распознавания специфическими белками, которые обладают метил-CpG-связывающими доменами (methyl-CpG binding domain – MBD). Это, в свою очередь, создает условия для взаимодействия с данными белками других белковых комплексов, подавляющих транскрипцию с участием разных механизмов, в том числе и через гетерохроматизацию соответствующих участков ДНК. Кроме того, наличие 5mC может препятствовать связыванию регуляторными участками ДНК факторов транскрипции, хотя это и не является распространенным механизмом подавления экспрессии генов [74]. В свою очередь, взаимодействие факторов транскрипции с регуляторными участками ДНК в ряде случаев является необходимым и достаточным условием для поддержания CpG-динуклеотидов в таких ПП в неметилированном состоянии и создает в геноме целые области с низким уровнем метилирования (low-methylated regions – LMRs) [75].

У метилирования ДНК имеются и другие функции. Репрессия транскрипции генов транспозонов метилированием ДНК предотвращает их мобилизацию, что является важным механизмом поддержания генома в стабильном состоянии [76]. Кроме того, метилирование ПП внутри активных генов, по-видимому, может оказывать регуляторное влияние на сплайсинг и выбор альтернативных промоторов [77]. В соматических тканях метилировано до 80% динуклеотидов CpG. Картирование участков метилирования ДНК в геноме высокопроизводительными методами показало, что, в соответствии с вышеупомянутыми функциями, в соматических клетках метилированными оказываются сателлитные и другие повторяющиеся ПП, включая транспозоны и их остатки, уникальные межгенные ПП и другие некодирующие ПП, а также экзоны генов. Уровень и паттерны метилирования в этих ПП, соответствующие встречаемости в них CpG-динуклеотидов, могут сильно различаться у разных индивидуумов одного биологического вида, представляя геномные полиморфизмы и эпимутации [78, 79].

Поддерживающие метилтрансферазы, такие как Dnmt1 (от DNA methyltransferase 1), сохраняют характерные паттерны метилирования ДНК (метилом клеток) на протяжении многих клеточных поколений. Инактивация этого фермента генным нокаутом в эмбриональных стволовых клетках мышей сопровождается глобальным деметилированием ДНК и летальна для развивающегося зародыша. Метилом, как существенная часть эпигенома клеток, высоко динамичен в жизненном цикле живого организма, тканеспецифичен и может изменяться вследствие ферментативного метилирования и деметилирования ДНК *de novo* [80, 73].

Исключение из глобально метилированных ПП генома составляют, так называемые, CpG-островки (CpG island – CGI), CG-богатые ПП длиной 0.2–2 т.п.н., расположенные вблизи промоторов и 5'-концевых ПП генов, а также в межгенных участках генома (обзор [81]). Функциональное значение имеют, по-видимому, ПП, удаленные от CGI до 2 т.п.н (так называемые «берега» – shores) и 2–4 т.п.н. («шельф» – shelves), которые дифференциально метилированы в клетках разных тканей [82, 83]. Около 60% генов млекопитающих обладают CGI-промоторами, и их CGI остаются неметилированными в большинстве соматических тканей, в клетках зародышевой линии и раннем эмбриогенезе. Неметилированное состояние CGI, как полагают, поддерживается при участии CpG-связывающего белка CFP1 (CxxC finger protein 1) и деметилазы гистонов KDM2A, обладающих характерными CXXC-доменами типа «цинковых

пальцев». Эти белки специфически связываются CpG-динуклеотидами, изменяют характер метилирования гистонов, ассоциированных с данными участками генома, и активируют соответствующие промоторы [84]. В нуклеосомах, ассоциированных с CGI, преобладают триметилированные формы гистонов H3K4me3, а гистоны H3K36me2 деметилированы, что характерно для активно транскрибируемых генов. В геномах млекопитающих обнаружено до 20 000 CGI, что составляет ~5% от геномных CpG-динуклеотидов или ~1% всего генома [85].

В отличие от CGI активных генов, глобальное метилирование CGI сопровождается инактивацией одной из X-хромосом у самок млекопитающих, необходимую для дозовой компенсации уровней экспрессии генов половых хромосом [86]. При сравнении ДНК из клеток с одной активной X-хромосомой (кариотип 45,X – синдром Тернера, X_a) и нормальных клеток с одной активной и одной инактивированной X-хромосомой (X_i) было установлено, что в X_i только 7% CGI демонстрирует понижение уровня метилирования, тогда как метилирование остальных CGI было значительно выше, чем в контроле. Как межгенные, так и внутригенные CGI демонстрировали высокий уровень метилирования, который был максимальным на промоторах инактивированных генов. Помимо метилирования в инактивации X-хромосом большую роль играют и некодирующие (nc) РНК.

Один из ярких примеров функционального значения метилирования дает родительский импринтинг, в результате которого часть генов одного из родителей передается потомству в неактивном состоянии [87]. Реализация этого механизма необходима для правильного эмбрионального и неонатального развития млекопитающих. В настоящее время у мышей известно около 150 импринтных генов, и они расположены кластерами на 17 хромосомах. Экспрессия каждого кластера (всего или его части) регулируется специальными ПП, получившими название областей, контролирующих импринтинг (imprinting control region – ICR), а также lncРНК, гены которых находятся рядом с ICR и экспрессируются только на гомологичных хромосомах, не подверженных импринтингу. В соответствии с этим, в большинстве случаев ICR положительно регулирует экспрессию lncРНК. В целом, в соответствии с современными моделями, ICR в импринтинге выполняет функции инсулятора, предотвращая взаимодействие энхансера с промоторами импринтных генов кластера, а lncРНК обеспечивает гетерохроматизацию не экспрессирующихся при импринтинге генов. (Подробнее об инсуляторах и lncРНК см. ниже в соответствующих разделах обзора).

ICR разделяют на два класса – ICR клеток зародышевой линии (germ-line ICR – gICR) и соматические ICR (sICR) [87a]. Аллель-специфическое метилирование первых происходит во время гаметогенеза и поддерживается на протяжении эмбрионального развития. Напротив, статус метилирования sICR устанавливается в процессе развития организма и часто носит тканеспецифический характер. В геноме мышей идентифицировано 55 ICR, длина которых варьирует в пределах 1–5 т.п.н. [87].

Метилирование ДНК в CGI промоторов типа II и III (см. выше), для которых характерно наличие CGI, сопровождается подавлением транскрипции прилегающих к ним генов. В то же время, CpG-метилирование промоторов типа I не сказывается сильно на синтезе соответствующих РНК.

Метилирование ДНК у млекопитающих не всегда приводит к подавлению транскрипции. Недавние исследования показали, что специфическое метилирование ДНК в межгенных областях генома, а также в интронах, может сопровождаться активацией генов. Например, гиперметилирование первого интрона гена *EGR2* оказывает стимулирующее влияние на его экспрессию, тогда как гипометилирование указанной ПП связано с подавлением транскрипции. Механизм этого явления в настоящее время не изучен [88]. Неактивное состояние онкогена *BCL6* поддерживается фактором транскрипции CTCF, который блокирует действие энхансера, взаимодействуя с CGI интрона 1, когда он находится в неметилированном состоянии. При метилировании этой ПП в опухолях происходит активация транскрипции онкогена из-за возникающего препятствия связыванию с CTCF [88].

Некодирующие ПП генома, подвергающиеся специфическому метилированию, выполняют важные регуляторные функции. На клеточном уровне метилирование обеспечивает выполнение программы дифференцировки клеток и поддержание стабильности генома путем подавления активности МГЭ. На молекулярном уровне функции 5mC зависят от многих параметров, основным из которых является геномный контекст, в котором находятся метилированные ПП, в частности, в промоторах, интронах, экзонах или межгенных областях генома. Такие участки генома маркированы специфическими ПП, в частности, CGI в окрестностях генов, ICE-элементами, контролирующими импринтинг, которые в целом составляют существенную часть некодирующих ПП генома млекопитающих. Действительно, хотя в настоящее время размер эпигенома человека точно не известен, он колоссален. В диплоидном геноме содержится $>10^8$ потенциально

метилируемых остатков С, из которых 10^7 находится в составе CpG-динуклеотидов, и значительная их часть локализована в некодирующих частях генома. Кроме того, в нем имеется $>10^8$ потенциально модифицируемых концевых аминокислотных ПП гистонов [89].

Эпигенетические модификации ДНК и гистонов наследуются в ряду клеточных поколений и могут быть нарушены в результате эпимутаций, что сопровождается развитием патологических состояний. В то же время, заболевания часто сопровождаются изменениями эпигенома, и бывает трудно определить, что является причиной, а что следствием. Вариации по метилированию CpG-сайтов, получившие название вариабельных положений метилирования (methylation variable position – MVP), являются эпигенетическими аналогами SNP. Изменение характера метилирования нескольких CpG-динуклеотидов подряд называют дифференциально метилированными областями (differentially methylated region – DMR). Такие изменения особенно характерны для онкологических заболеваний, при которых часто наблюдают aberrантное метилирование CGI, утрату импринтинга и эпигенетические перестройки повторов, особенно сателлитной ДНК [Feber 2011]. Разработаны технологии, позволяющие осуществлять полноэпигеномный поиск ассоциаций между изменениями эпигенома и заболеваниями человека (epigenome-wide association studies – EWAS) [89]. Использование этой группы методов обещает в скором времени углубить наши знания о влиянии MVP и DMR на здоровье человека и уже способствует получению новых данных [90].

Интроны

Интроны представляют собой транскрибируемые внутригенные ПП, которые, как правило, не включаются в зрелые мРНК и удаляются из их предшественников в результате сплайсинга [91]. Поскольку большинство интронов не несут информации об аминокислотных ПП, их можно отнести к некодирующей части эукариотического генома. Гены человека в среднем содержат по 8–9 интронов и являются одними из наиболее насыщенных этими ПП по сравнению с генами других биологических видов. В частности, ген титина (коннектина), гигантского мышечного белка, построенного из 244 белковых доменов, содержит 362 интрона [92], а длина самых больших интронов человека приближается к 1 м.п.н. Лишь ~300 аннотированных генов человека являются безинтронными, из которых половина имеет отношение к системам передачи сигналов, а пятая часть кодирует гистоны [93]. Суммарная доля интронов в геноме человека составляют ~24% [94, 95].

Для исчерпывающего ответа на вопрос о функциональном значении интронов необходимо было бы выяснить причины их эволюционного происхождения и широкого распространения среди эукариотических организмов. 20-летние дебаты между сторонниками раннего или позднего происхождения интронов могут завершиться их примирением [96]. Хотя самоудаляющиеся в результате аутосплайсинга интроны, по-видимому, могли появиться рано в развивающемся мире древних РНК, современные интроны, вероятно, возникли поздно в процессе эукариогенеза, и их глобальное функциональное значение для эукариот на сегодняшний день остается не до конца понятным [97].

Популярным объяснением эволюционного закрепления интронов в геноме, является предположение об облегчении рекомбинационного обмена между отдельными экзонами, объединение которых в новые ПП в разных сочетаниях могло бы ускорить появление новых генов [98]. В пользу этой гипотезы свидетельствует, в частности, функционирование у эукариот широко распространенного механизма альтернативного сплайсинга, в результате которого происходит объединение на уровне процессируемой РНК экзонов и их частей в разных сочетаниях с образованием изоформ исходного полипептида с разными активностями [99]. Появление этого механизма значительно увеличивает информационную емкость эукариотического генома: у человека 95% мультиэкзонных генов поддерживают альтернативный сплайсинг с образованием в среднем по 10–11 изоформ мРНК/ген [100]. При этом одна или две изоформы мРНК преобладают в конкретной линии клеток, а следовательно их экспрессия является тканеспецифической. В целом 20 687 генов человека направляют образование ~100 000 белков, что свидетельствует о 5-кратном увеличении информационной емкости генома за счет альтернативного сплайсинга [101]. Кроме этого альтернативный сплайсинг может участвовать в негативной регуляции экспрессии генов через образование нефункциональных или быстро деградирующих РНК [102], а также обеспечивать внутриклеточную локализацию транскриптов посредством введения в зрелые мРНК ПП, необходимых для правильного транспорта макромолекул [103]. Преимущества, которые предоставляет эукариотам альтернативный сплайсинг, не объясняют высокой гетерогенности интронов по размерам. Кроме того, многие альтернативные транскрипты остаются функционально неохарактеризованными, и вопрос об уровне «информационного шума» от aberrантного альтернативного сплайсинга остается открытым [104, 105].

Trans-сплайсинг можно отнести к одной из разновидностей альтернативного сплайсинга, в результате которого происходит объединение экзонов пре-мРНК двух разных генов в одну химерную молекулу [106]. У человека наиболее часто встречаются химерные мРНК и белки, объединяющие ПП расположенных друг за другом генов. Это происходит в результате прохождения транскрибирующей РНК-полимеразой без терминации терминатора транскрипции, что приводит к образованию длинного химерного транскрипта. Последующий межгенный сплайсинг формирует химерные мРНК, объединяющие экзоны разных генов, а их трансляция сопровождается образованием химерных белков [107]. У человека в разных тканях обнаружены сотни случаев межгенного сплайсинга, а сам процесс образования химерных РНК носит регулируемый тканеспецифический характер. В результате могут образовываться новые бифункциональные белки, новые РНК, трансляция которых регулируется по типу одного из генов в зависимости от включившейся в транскрипт ПП 5'- или 3'-UTR, а также подавление экспрессии выше расположенного гена по механизму распада, опосредованного нонсенс-мутациями (Nonsense Mediated Decay – NMD) в том случае, если слияние ПП приводит к образованию нонсенс-кодона в новой рамке считывания.

Trans-сплайсинг между РНК далеко расположенных генов встречается у человека не так часто. Многие кажущиеся химерными молекулы РНК являются артефактами, которые возникают вследствие геномных перестроек или копирования с переменной матрицы при обратной транскрипции [108]. Тем не менее, наличие в клетках человека химерных РНК, возникших в результате *trans*-сплайсинга РНК далеко отстоящих друг от друга генов в настоящее время доказано. В частности, в нормальных клетках молочной железы человека обнаружены химерные транскрипты генов, расположенных на разных хромосомах [109]. Межхромосомный *trans*-сплайсинг отмечен между экзогенной бактериальной РНК и эндогенной клеточной РНК человека [110]. В транскриптоме плюрипотентных эмбриональных стволовых клеток человека, как природных, так и индуцированных, достоверно обнаружено несколько *linc*РНК с высоким внутриклеточным содержанием, возникших в результате *trans*-сплайсинга [111]. Подавление экспрессии этих РНК с помощью синтетических коротких РНК (shРНК) сопровождалось утратой клетками способности поддерживать плюрипотентность. В целом, существование механизма межгенного сплайсинга значительно расширяет кодирующий потенциал генома млекопитающих и повышает разнообразие его протеома [112–114].

Еще одним известным феноменом, указывающим на глобальную функциональную значимость интронов, является опосредованное интронами усиление экспрессии генов (Intron-Mediated Enhancement – IME) на уровне транскрипции. Наличие интронов в генах стимулирует их экспрессию у представителей далеко отстоящих друг от друга таксономических групп, включая животных и растения, что указывает на фундаментальное значение этого явления [91]. Многие гены с интактными промоторами в отсутствие интронов *in vivo* вообще не экспрессируются. Как правило наибольшее влияние на экспрессию оказывают интроны, локализованные в 5'- или 3'-UTR, которые сильно различаются по способности стимулировать транскрипцию. Способность интронов обеспечивать в мейозе эффективную рекомбинацию аллельных генов, кодирующих белки, рассматривают как один из факторов, ускоряющих эволюцию белков [115]. Протяженные ПП интронов повышают вероятность рекомбинации между аллельными мутантными экзонами и создают белки с новыми сочетаниями мутаций, которые далее испытываются организмом на функциональность.

Интроны являются местом локализации многих регуляторных элементов ДНК. Кроме обычных ПП, необходимых для прохождения сплайсинга (5'- и 3'-концевые сайты сплайсинга, сайт ветвления, полипиримидиновый тракт, интронные энхансеры и сайленсеры сплайсинга [116]), интроны могут содержать ПП, регулирующие транскрипцию (альтернативные промоторы и терминаторы транскрипции, транскрипционные энхансеры и сайленсеры), а также гены miРНК и lncРНК. Скорость сплайсинга как таковая может быть определяющим фактором в регуляции экспрессии генов. В частности, редкие U12-интроны, в сплайсинге которых участвуют сплайсеосомы, содержащие snРНК U12, удаляются из пре-мРНК человека медленнее, чем обычные интроны класса U2, что сопровождается сильным подавлением биосинтеза соответствующих белков [117]. Во многих случаях регуляторные функции интронов распространяются на конкретные классы генов и даже индивидуальные гены.

Как следует из вышеизложенного, интроны выполняют многочисленные важные функции в регуляции экспрессии генов. Поэтому не удивительно, что мутации в интронах часто сопровождаются тяжелыми патологическими последствиями. Это является дополнительным указанием на важную функциональную роль данных некодирующих ПП. Особенно часто встречаются мутационные замены нуклеотидов, нарушающие сплайсинг, которые изменяют донорные и акцепторные сайты сплайсинга [118]. Мутации в полипиримидиновом тракте и

точках ветвления редки. В целом интронные мутации составляют ~10% от всех известных к настоящему времени патогенных мутаций у человека и сопровождаются разными перестройками его зрелых мРНК.

Энхансеры транскрипции

Энхансеры представляют собой регулирующие транскрипцию ПП, расположенные как вблизи, так и на значительном (до нескольких м.п.н.) расстоянии от промоторов регулируемых ими генов. В некоторых случаях они обеспечивают межхромосомную передачу сигналов *in trans* (например [119]). По размерам различают типичные энхансеры (длина их ПП <1 т.п.н.), как правило, расположенные вблизи генов домашнего хозяйства, и суперэнхансеры (3–50 т.п.н.), локализованные преимущественно рядом с ключевыми генами, контролирующими развитие (обзор [120]).

Основными белками, функционально взаимодействующими с энхансерами, являются факторы транскрипции, которые могут быть как активаторами, так и репрессорами [121]. В частности, с активными энхансерами *in vivo* часто ассоциирован белок-коактиватор глобальной транскрипции, ацетилтрансфераза p300, субъединицы Медиатора, ДНК-связывающий белок 7 хромодоменной хеликазы, когезин, фактор CTCF и РНК-полимераза II [122]. Кроме того, контактирующий с энхансерами гистон H3 обнаруживает характерные модификации – моно- и/или диметилирование аминокислотного остатка Lys в положении 4 (H3K4me¹ и H3K4me²), а также ацетилирование Lys-27 (H3K27ac), тогда как свойственное активным промоторам триметилирование H3K4me³ незначительно [5, 123–125]. Кроме того ПП активных энхансеров обнаруживают гиперчувствительность к ДНКазе I (HS-сайты), что характерно для «открытого» состояния хроматина. Такого рода молекулярные маркеры энхансеров создают предпосылки для их идентификации в функционирующем геноме. Паттерны модификаций гистонов в хроматине энхансеров в клетках разных тканей строго коррелирует с тканеспецифической экспрессией генов [126]. По оценкам, полученным с использованием иммунопреципитации хроматина антителами к белку p300 и методологии NGS, общее число энхансеров в геноме человека приближается к 10⁶, и их ПП могут составить до 3% от всех нуклеотидов генома [126, 127]. На расположение энхансеров в геноме человека в среднем через каждые 3–30 т.п.н. указывают и данные, полученные консорциумом ENCODE [5]. Таким образом, в геноме человека число энхансеров значительно превышает число

промоторов, причем примерно половина из них располагается внутри генов [123].

Для энхансеров характерно модульное строение: они заключают в себе несколько (иногда более 10) коротких, длиной 6–12 п.н., сайтов связывания (мотивов) факторов транскрипции, расположенных кластерами [128]. Осуществляя взаимодействие *trans*-действующих белковых факторов транскрипции с регуляторными ПП промоторов, энхансеры вносят основной вклад в обеспечение тканеспецифической транскрипции генов и развитие многоклеточных организмов. Взаимодействие факторов транскрипции с энхансерами находится под строгим контролем и, само по себе, без вспомогательных белков, может не сопровождаться видимыми функциональными последствиями [129]. Модульное строение энхансеров обеспечивает функциональное объединение факторов транскрипции в разных комбинациях, что существенно расширяет их регуляторные возможности. Один и тот же промотор в разных тканях и в разное время может находиться под регуляторным воздействием разных регуляторных элементов энхансера.

Некоторые энхансеры присутствуют в геноме в виде двух и более удаленных друг от друга копий, которые обладают одинаковыми или сильно перекрывающимися активностями по отношению к регулируемому ими гену. Первичный энхансер, как правило, располагается близко к промотору гена-мишени, другие – на некотором расстоянии, часто вблизи другого гена. Такие дальние энхансеры, не отличимые по своим свойствам от первичного, получили название теневых [130]. Теневые энхансеры, впервые обнаруженные у дрозофилы [131], функционируют и в геноме человека, располагаясь, в частности, в интронах гена фактора транскрипции GLI3 [132], или перед геном ренина [132a]. Кажущаяся функциональная избыточность таких энхансеров повышает устойчивость системы к мутационным изменениям и в ряде случаев к экстремальным факторам окружающей среды, а также обеспечивает более точное выполнение программы раннего развития организма [130].

Стимулирующее действие энхансеров на транскрипцию, осуществляемую РНК-полимеразой II, не зависит от их ориентации в молекулах ДНК, а также положения относительно регулируемого промотора [133]. Они одинаково эффективно осуществляют свои функции находясь перед геном, за геном, внутри него или на значительном (до нескольких сотен т.п.н.) удалении, что является одним из критериев, используемых при экспериментальном выявлении энхансеров. При этом независимость от положения на молекуле ДНК распространяется на энхансеры как целое, но не их отдельные модули.

Вопрос о молекулярных механизмах, лежащих в основе передачи регуляторных сигналов от энхансеров к промоторам на большие расстояния в настоящее время решается в пользу «выпетливания» участков ДНК, расположенных между энхансерами и промоторами генов и прямого взаимодействия обоих регуляторных элементов [134]. Модель получает убедительное подтверждение в многочисленных опытах с применением высокопроизводительных подходов к анализу конформаций хроматина в интерфазном ядре на основе 3С-метода захвата конформации хромосом (Chromosome Conformation Capture), с помощью которого взаимодействие между участками хромосом определяют путем введения между ними поперечных сшивок, с последующей рестрикцией ДНК, лигированием сближенных фрагментов и секвенированием образовавшихся продуктов лигирования [135]. Факторы транскрипции и их комплексы являются основными участниками взаимодействия энхансеров с промоторами. Эти белково-нуклеиновые комплексы, получившие название энхансеосом, также содержит белок когезин, который, кроме того, участвует в удержании сестринских хроматид хромосом в митозе и мейозе (см. обзор [125]).

С энхансерами взаимодействует и основной коактиватор транскрипции СВР, привлекающий к энхансерам РНК-полимеразу II, которая у эукариот осуществляет транскрипцию большинства генов, кодирующих белки [136]. В соответствии с этим, недавно было обнаружено, что многие энхансеры сами обладают свойствами промоторов, с которых РНК-полимераза II синтезирует lncРНК, получившие название eРНК [137–140]. Синтез eРНК может быть как однонаправленным (1D-eРНК), так и двунаправленным (2D-eРНК), а сами транскрипты эспированы на 5'-концах. При этом однонаправленные транскрипты полиаденилированы, что нехарактерно для большинства двунаправленных, которыми, в основном, представлены eРНК [126]. Постепенно становится ясно, что eРНК играют важную роль в инициации образования и стабилизации промоторно-энхансерных петель (см. обзоры [138, 140]). В пользу этого свидетельствуют, например, данные о нарушении функционирования энхансеров в результате направленного разрушения eРНК под действием искусственных siРНК или антисмысловых олигонуклеотидов [141, 142]. Поскольку транскрипция энхансеров, в свою очередь, регулируется разнообразными сигналами, то это создает новый уровень регуляции глобальной транскрипции у эукариот.

Важнейшая роль энхансеров в выполнении эмбриональными стволовыми клетками программы развития организма реализуется путем их подготовки к активации через взаимодействие с белковыми

факторами транскрипции, которые поддерживают энхансеры в открытом для присоединения других факторов (компетентном) состоянии. Появление в клетках новых факторов транскрипции приводит к замещению ими старых, активации транскрипции соответствующих генов и продолжению дифференцировки клеток в нужном направлении [143]. Помимо факторов транскрипции на некоторых энхансерах обнаруживают и белки, обычно входящие в состав комплексов, подавляющих транскрипцию, которые в данном случае выступают в необычной для них роли активаторов. Недавнее открытие таких двойственных функций белков репрессоров-активаторов транскрипции явилось важным достижением современных исследований транскрипции у эукариот (см. обзор [144]).

Замечательным свойством некоторых энхансеров является их высокая эволюционная и функциональная консервативность. Например, среди 167 исследованных ультраконсервативных энхансерных ПП человека почти половина поддерживала тканеспецифическую экспрессию гена-репортера в эмбриогенезе трансгенных мышей и рыб [145, 146].

Исключительная важность энхансеров в поддержании паттернов тканеспецифической экспрессии генов на разных этапах индивидуального развития предполагает тяжелые функциональные последствия их мутационных изменений. Действительно, описаны многочисленные мутации, приводящие к развитию онкологических заболеваний, глухоте, системной красной волчанке, множественному склерозу, болезни Крона и другим патологическим состояниям организма человека [147].

Локус-контролирующие области

Многие кластеры генов, дифференциально экспрессирующихся при развитии организмов и в разных тканях, координированно регулируются локус-контролирующими областями (locus control region – LCR) (обзоры [148, 149]). Этот тип *cis*-действующих позитивных регуляторных элементов, близких по своим свойствам к энхансерам, в отличие от них обеспечивает устойчивую к эффекту положения транскрипцию контролируемых ими трансгенов, уровень которой также зависит и от числа копий трансгенов в геноме. Известно, что экспрессия трансгенов, стабильно интегрированных в геном, зависит от места их интеграции в хромосому (так называемый мозаичный эффект положения), и их транскрипция подавлена при встраивании в гетерохроматизированные участки хромосом [150]. LCR, расположенные вблизи промотора, обеспечивают функцио-

нальность регулируемого гена независимо от его положения в хромосоме. Это указывает на то, что среди функциональных модулей LCR присутствуют инсультаторы (см. ниже в соответствующем разделе обзора), что и было подтверждено экспериментально [151].

LCR рассматривают как сложные энхансеры, составленные из нескольких энхансерных модулей. Расстояние LCR до регулируемых промоторов может составлять несколько десятков т.п.н. В соответствии с этим, для LCR характерно наличие нескольких (до десяти) HS-сайтов, маркирующих индивидуальные энхансеры. Эти ПП, как и в обычных энхансерах, маркированы ацетилированными формами гистона H3, характерными для «открытого» состояния хроматина. Кроме того, LCR и регулируемые ими гены в активном состоянии пространственно сближены и непосредственно взаимодействуют друг с другом, что сопровождается выпетливанием ДНК, разделяющей эти последовательности. Более того, выявлены межхромосомные взаимодействия промоторов и LCR: например, в Т-хелперах мышей промотор гена интерферона γ в локусе T_H1 , расположенном на хромосоме 10, взаимодействует с LCR локуса T_H2 , расположенного на хромосоме 11 [152]. Еще большее функциональное сходство LCR с обычными энхансерами придает межгенная транскрипция, инициация которой происходит в LCR, и которая сопровождается образованием lncРНК [153]. Появление этих некодирующих транскриптов коррелирует со структурными изменениями хроматина, деметилированием ДНК, ацетилированием гистонов и может играть важную роль в регуляции этих процессов. Подробнее о lncРНК см ниже в нашем обзоре.

Помимо непосредственного участия в активации генов, энхансеры, в том числе и в составе LCR, играют важную роль в поддержании пространственной организации хроматина в интерфазных ядрах. Например, активно транскрибируемые гены часто выдвигаются в составе петель хроматина за пределы своих хромосомных территорий, и на этот процесс оказывают влияние, в частности, LCR β -глобинового локуса [154]. При дифференцировке клеток эритроидного ряда происходит перемещение того же локуса от периферии ядра к его центру и ассоциация с «фабриками транскрипции» или другими самоорганизующимися субструктурами типа спеклов, участвующих в сплайсинге, что сопровождается активацией экспрессии генов, и этот процесс также требует участия LCR [155, 156]. Энхансер E_{μ} требуется для миграции локуса IgH от периферии ядра к его центру [157]. Необходимость правильной локализации экспрессирующихся генетических локусов внутри интерфазных ядер, указывает, на наш взгляд, на еще одну общую функцию межгенных некодирующих ПП в

эукариотическом геноме: они обеспечивают требуемую подвижность локусов внутри ядра, сохраняя их связанными в единое целое.

Инсуляторы хроматина

Регуляторные ПП, получившие название инсуляторов (от англ. *insulate* – защищать, изолировать), играют ключевую роль в поддержании внутриядерной пространственной организации эукариотического генома и формировании топологически ассоциированных доменов хромосом (*topologically associating domain* – TAD), длина которых находится в пределах 1 м.п.н. TAD являются универсальными структурными элементами пространственной организации геномов человека и животных в интерфазных ядрах, и их расположение в геноме высококонсервативно [158]. Кроме этого, возникающие в результате действия инсуляторов многочисленные внутригеномные контакты сопровождаются пространственным сближением промоторов и энхансеров, что необходимо для их адекватного функционирования. В результате инсуляторы предотвращают перекрестное неспецифическое действие энхансеров и сайленсеров на промоторы, изолируя друг от друга функциональные домены хроматина [обзоры 159, 160]. В реализацию такого действия инсуляторов вовлечены два основных механизма: блокировка действия энхансеров и создание барьера на пути распространения гетерохроматина на соседние эухроматиновые участки хромосом. В генно-инженерных экспериментах первый механизм проявляется в инактивации трансгена инсулятором, помещенным между энхансером и промотором, а второй – в защите трансгена, фланкированного инсуляторами, от эффекта положения. Кроме этого, недавно продемонстрирована способность инсуляторов направлять специфическое сближение энхансеров и промоторов, сопровождающееся активацией соответствующих генов, а также их участие в демаркации границ между участками хроматина, находящимися в различных эпигенетических состояниях. У позвоночных сфера влияния инсуляторов не ограничивается только транскрипцией. Например, недавно было продемонстрировано их участие в регуляции V(D)J-рекомбинации в иммуноглобулиновых локусах [160].

В хроматине инсуляторы часто маркированы специфическими белками: фактором транскрипции CTCF, высоко консервативным белком, обладающим 11 доменами типа «цинковых пальцев», и/или фактором транскрипции РНК-полимеразы III TFIIIS, многосубъединичным белком, взаимодействующим с В-боксом промотора генов тРНК (обзор [161]). Оба белка контактируют с когезиновым комплексом, который, связываясь с ДНК, стабилизирует взаимодействие удален-

ных участков хроматина друг с другом при формировании оснований петель доменов хроматина. Таким образом, ПП инсуляторов животных могут содержать сайт связывания CTCF, представляющий собой вырожденную ПП длиной до 50 п.н., или же ген тРНК.

Методом ChIP с последующим NGS в геноме человека было обнаружено до 30 000 сайтов связывания CTCF, а, следовательно, и потенциальных инсуляторов, из которых 43% было локализовано в межгенных областях, 7% – в 5'-UTR, 3% – в экзонах и 29% – в интронах генов, 2% – в 3'-UTR, а также 16 % – в окрестностях TSS [162]. При использовании того же подхода в геноме человека было обнаружено в дополнение к промоторам генов тРНК несколько тысяч коротких ПП, с которыми взаимодействовал TFIIIC [163]. Эти дополнительные ПП, получившие название ETC (от англ. Extra TFIIIC), обладали способностью связывать TFIIIC независимо от промотора РНК-полимеразы III, который содержит ПП А- и В-боксов. В отличие от этих промоторов, ETC-сайты содержали или В-бокс, или консервативный GC-богатый 16-нуклеотидный мотив.

S/MAR-последовательности

В пространственной организации функциональных доменов хромосом в интерфазных ядрах наряду с инсуляторами важную роль играют ПП ДНК, которые обеспечивают прикрепление оснований петель к ядерному матриксу (scaffold/matrix attachment regions – S/MAR). Эти AT-богатые (>70%) фрагменты ДНК остаются ассоциированными с ядерным матриксом после экстракции ядер детергентами или буфером с высокой ионной силой [164], и присутствуют как в генах, так и межгенных областях, вблизи инсуляторов, энхансеров, *cis*-действующих регуляторных элементов, а также вблизи областей начала репликации. Кроме того S/MAR часто маркируют границы между конденсированным и деконденсированным хроматином [165]. В процессе дифференцировки клеток имеет место внутриядерное перераспределение петель хроматина, контактирующих с ядерным матриксом, в котором участвуют S/MAR, что связано с изменением профилей экспрессии генов [166, 167]. Исследования S/MAR не выявили для них консенсусной ПП, что, как предполагают, можно объяснить распознаванием взаимодействующими с ними белками матрикса особенностей пространственной, но не первичной структуры ДНК. Недавно с помощью гибридизации на биочипах были картированы в клетках HeLa в ПП ДНК длиной в 30 м.п.н., представляющей 1% генома человека, 453 S/MAR, средняя длина которых составляла ~5 т.п.н. [168]. Большинство S/MAR было лока-

лизовано вблизи концевых ПП экспрессирующихся генов и были ассоциированы с РНК-полимеразой II, а также сайтами связывания факторов транскрипции. В то же время, около 40% S/MAR было расположено в межгенных областях или в окрестностях неактивных генов и могли представлять собой классические пограничные ПП, а также участвовать в организации общей пространственной укладки генома. Расстояние между соседними S/MAR в клетках HeLa составляет 80–90 т.п.н. [169] и может варьировать в разных участках генома [168].

НЕКОДИРУЮЩИЕ РНК

Современный анализ транскриптома млекопитающих высокопроизводительными методами с применением NGS показал, что его значительная часть представлена ncРНК. Все РНК, не кодирующие белки, попадают в эту группу, которая включает как хорошо изученные рРНК, тРНК, малые ядерные (sn) и малые ядрышковые (sno) РНК, так и недавно обнаруженные короткие и длинные ncРНК, функции большинства которых в настоящее время неизвестны. Именно эта последняя группа ncРНК будет рассмотрена в данном разделе нашего обзора.

Короткие некодирующие РНК

Короткие некодирующие РНК, обнаруженные в организме млекопитающих, принято разделять на три класса: микроРНК (miРНК), короткие интерферирующие РНК (short interfering RNA – siРНК) и РНК, взаимодействующие с PIWI (piРНК) (обзоры [170, 171]). РНК всех трех классов представляют собой одноцепочечные олигонуклеотиды длиной ~22 нт, реализующие свое регуляторное действие посредством образования специфических комплексов с мРНК-мишенями на основе комплементарности, но различающиеся биогенезом.

miРНК, которых у человека известно в настоящее время >3000, кодируются генами, рассеянными по геному, и транскрибируются РНК-полимеразой II (реже РНК-полимеразой III) с образованием предшественников pri-miРНК с одним или несколькими элементами вторичной структуры типа «стебель-петля», в стебле которых находятся ПП miРНК. Гены miРНК могут располагаться как в межгеномном пространстве геномной ДНК, часто кластерами (50% всех генов), так и в интронах (40%) и некодирующих экзонах (10%) генов белков или некодирующих РНК [172]. Межгенные miРНК транскрибируются с собственных промоторов их генов, тогда как интронные miРНК – чаще всего с промоторов генов, в которых они локализованы и в ~1/3 слу-

чаев с собственных промоторов [173]. Интересной разновидностью интронных miРНК являются митроны – первые этапы биогенеза таких miРНК осуществляются в результате сплайсинга [174]. Кроме того, miРНК могут быть заключены внутри более длинных некодирующих РНК, например, snoРНК [175]. Транскрипция с собственных промоторов обеспечивает дополнительные возможности регуляции экспрессии miРНК с использованием классических механизмов. Pri-miРНК, синтезированные РНК-полимеразой II, кэпированы и полиаденилированы.

Процессинг pri-miРНК продолжается в ядре с участием белкового комплекса, получившего название микропроцессора, основными компонентами которого являются белки Drosha (обладает активностью РНКазы III), DGCR8, а также хеликазы p68 и p72. Образовавшаяся в итоге pre-miРНК, сохранившая структуру «стебель–петля», экспортируется из ядра в цитоплазму через ядерный поровый комплекс, где «стебель» отщепляется другой РНКазой III, названной Dicer, с образованием miРНК-дуплекса, одна цепь которого представляет собой miРНК, а другая – неактивную miРНК* (цепь–пассажир). Процессинг pre-miРНК может протекать с отклонениями, приводящими к образованию большого числа изоформ miРНК (Isomir), которые различаются по размерам и первичной структуре, а также, как следствие, по функциональной активности [176].

В процессе формирования комплекса miRISC (RNA-induced silencing complex) miРНК-дуплекс включается в белковый комплекс AGO, который распознает 5'-монофосфат цепи miРНК с предпочтением к остаткам А и U в качестве 5'-концевых нуклеотидов. Вытеснением цепи–пассажира и ее последующей деградацией завершается образование miRISC [177]. Предполагается, что при выборе цепи РНК-дуплекса в качестве miРНК реализуется правило асимметрии, в соответствии с которым включенной в комплекс оказывается цепь, 5'-концевая часть которой менее стабильно связана с комплементарной ей цепью дуплекса, чем 3'-концевая. Реализация такого механизма выбора цепей miРНК-дуплекса должна быть высоко чувствительна к изменениям его первичной структуры в результате включения альтернативных нуклеотидов SNP или под действием мутаций.

Зрелый miRISC-комплекс взаимодействует с mРНК-мишенью, которой полностью или частично комплементарна включенная в него miРНК. Примерно в половине исследованных случаев одним из требований к этому взаимодействию является наличие в 5'-концевой части miРНК затравочного («seed») участка длиной в 2–7

нт, полностью комплементарного мРНК [178]. Однако недавно в затравочной области комплекса были обнаружены неканонические взаимодействия miРНК с мРНК [179], а также кооперативно действующие множественные затравочные участки [180], что вместе с известным фактом регуляторного действия отдельных miРНК на многие мРНК-мишени указывает на гибкость процесса распознавания miRISC-комплексом мРНК-мишеней. Участки взаимодействия мРНК с miРНК часто расположены в 3'UTR, однако они обнаружены также в 5'UTR и их кодирующих частях.

Основным механизмом регуляторного действия miРНК является подавление трансляции мРНК через ее декэпирование и деаденилирование с последующей деградацией. Кроме того, взаимодействие мРНК с miRISC может сопровождаться ингибированием синтеза белка на уровне инициации или элонгации. В редких случаях отмечена стимуляция трансляции под действием miРНК. Недавно было обнаружено участие ядерных miРНК в позитивной и негативной регуляции транскрипции путем прямого взаимодействия с промоторами (обзор [181]). Кроме того, все этапы биогенеза miРНК и их взаимодействия с мишенями дополнительно регулируются (обзоры [170, 182]). Все это указывает на miРНК как ключевого регулятора экспрессии генов у млекопитающих, а их гены составляют существенную часть некодирующих ПП генома.

Биогенез эндогенных siРНК незначительно отличается от такового miРНК [183]. После объединения в цитоплазме РНК со своим антисмысловым партнером образовавшаяся дцРНК становится субстратом для Dicer с последующим включением siРНК в RISC-комплекс, в составе которого она реализует свое действие на мРНК-мишень. Основными источниками эндогенных siРНК у млекопитающих являются двунаправленные транскрипты транспозонов и перекрывающиеся участки мРНК и антисмысловых РНК генов, образовавшиеся в результате конвергентной или дивергентной транскрипции, а также при транскрипции псевдогенов.

Как следует из названия, малые piРНК при формировании RISC-комплекса взаимодействуют с белками PIWI, составляющими подсемейство Piwi семейства белков Argonaute и экспрессирующимися преимущественно в клетках зародышевой линии, хотя белки PIWI обнаружены в эмбриональных стволовых клетках и клетках разных соматических тканей (обзоры [172, 184, 185]). Длина piРНК составляет 26–31 нт, и общее число их видов приближается к 100 000. Основная функция piРНК – предотвращение транскрипции транспозонов и их мобилизации. В отличие от обсуждавшихся выше

малых РНК, предшественники рiРНК являются одноцепочечными, и в их процессинге не участвует Dicer. Описаны два пути биогенеза рiРНК – первичный и вторичный. При реализации первичного пути предшественники рiРНК длиной 1–100 т.п.н. транскрибируются РНК-полимеразой II с генных кластеров. Транскрипты являются антисмысловыми по отношению к РНК транспозонов. Многие ферменты, участвующие в биогенезе рiРНК, еще точно не идентифицированы. Предполагается, что после синтеза предшественник фрагментируется эндонуклеазой, и образовавшиеся фрагменты РНК с 5'-концевыми U взаимодействуют с белками Piwi, после чего они укорачиваются с 3'-конца до конечного размера и метилируются. Далее нуклеопротеиновые комплексы переносятся в ядро, где связываются с растущими цепями мРНК транскрибируемых транспозонов и, привлекая соответствующие белки, запускают метилирование или гетерохроматизацию ДНК, приводящие к подавлению транскрипции МГЭ. Вторичный путь биогенеза рiРНК обеспечивает их амплификацию на матрице мРНК транспозонов по пинг-понговому механизму.

Повсеместное участие miРНК в регуляции экспрессии генов предполагает их важную роль в развитии патологических процессов. Действительно, исследование полиморфизма и мутаций miРНК, их мишеней, а также ферментов и вспомогательных белков, участвующих в их процессинге и реализации регуляторного действия, выявило след этих вездесущих небольших молекул во всех исследованных патологических процессах, протекающих в организме человека [186].

Длинные некодирующие РНК

lncRNA, к которым относят ncРНК, длина которых превышает 200 нт, представляют наиболее многочисленный класс ncРНК с общим числом кодирующих их генов у человека приближающимся к 10 000 [187]. В этом недавно составленном каталоге представлено более 15 000 транскриптов, функции большинства из которых в настоящее время неизвестны. На основании локализации их ПП в геноме по отношению к кодирующим генам lncRNA разделяют на межгенные (intergenic – lincRNA) и внутригенные: антисмысловые, интронные, экзонные и перекрывающиеся. Большинство lncRNA представляют собой независимые единицы транскрипции, синтезируемые РНК-полимеразой II, 40% которых имеют сигнал полиаденилирования. При этом 98% lncRNA подвержены сплайсингу по каноническим сайтам (GT/AG) и, как правило, содержат только два экзона. Около четверти lncRNA претерпевают альтернативный сплайсинг и существуют в виде

нескольких изоформ. 58% lncРНК представлены молекулами небольшой длины (200~950 нт), у 40% длина составляет 950~4,800 нт, и 2% представлены транскриптами большей длины. При этом наибольший размер был зафиксирован для продукта одноэкзонного гена *NEAT1* – 22.7 т.н., который участвует в образовании ядерных параспеклов. Для генов lncRNA характерен более низкий по сравнению с кодирующими генами уровень экспрессии и более тканеспецифический ее характер. При этом для них отмечена выраженная ядерная локализация.

Известные функции lncRNA исключительно разнообразны и распространяются на все этапы экспрессии генов (обзоры [188, 189]). Большинство ядерных lncRNA направляют белки-модификаторы хроматина к соответствующим генетическим локусам [190]. Результатом может быть репрессия или активация генов соответствующего локуса через модификации ДНК или гистонов, сопровождающиеся гетерохроматизацией хроматина или изменением его конформации. Кроме того, выше уже упоминалось об участии ядерных lncРНК в дозовой компенсации при инактивации X-хромосом, установлении импринтинга, регуляции активности энхансеров и других некодирующих регуляторных элементов генома.

В цитоплазме lncРНК могут изменять стабильность мРНК, а также подавлять или активировать ее трансляцию. Эти активности lncРНК часто реализуются через комплементарные взаимодействия с мРНК-мишенью. Как уже упоминалось, антисмысловые lncРНК могут направлять образование siРНК. Транскрипты псевдогенов выступают в роли ловушек miРНК, которые регулируют экспрессию соответствующих генов. Такого рода lncРНК получили название конкурирующих эндогенных РНК (competing endogenous) ceРНК, разновидностью которых являются недавно обнаруженные и широко представленные кольцевые РНК (circРНК) [191, 192]. Линейные ceРНК обладают малой стабильностью, тогда как circРНК более стабильны и содержат регуляторные сайты связывания miРНК, с участием которых происходит их регулируемая инактивация.

ПСЕВДОГЕНЫ

Псевдогеном называют копию гена, утратившую способность продуцировать функциональный белок. В зависимости от происхождения, псевдогены классифицируют на процессированные и непроцессированные. Последние, в свою очередь, разделяют на одиночные и дублированные (обзоры [193–195]). Отдельную группу составляют ядерно-митохондриальные псевдогены, NUMT-псевдогены (NUMT – nuclear mitochondrial) [196].

Процессированные псевдогены образуются в результате встраивания продуктов обратной транскрипции мРНК соответствующих генов в новый участок генома. В дальнейшем различные повреждающие мутации способствуют тому, что эти копии генов утрачивают свой кодирующий потенциал. Отличительной особенностью процессированных псевдогенов является отсутствие интронов и наличие поли-А-тракта на 3'-конце. Часто процессированные псевдогены не содержат промоторов и их экспрессия может осуществляться с участием других регуляторных элементов. Например, псевдогены, расположенные в интронах других генов, используют аппарат транскрипции гена-хозяина. Приблизительно 10% генов имеют процессированные псевдогены. Такие псевдогены наиболее характерны для генов «домашнего хозяйства».

Дуплицированные псевдогены образуются в результате тандемной дупликации или кроссинговера. Последующие мутации делают эти копии генов функционально неактивными. Такие псевдогены сохраняют интрон-экзонную структуру. В отличие от процессированных псевдогенов, которые могут располагаться в различных участках генома, дуплицированные псевдогены находятся на тех же хромосомах, что и гены-предшественники. Одиночные псевдогены являются результатом повреждающей мутации в единственной копии исходного гена. Такие псевдогены не имеют функционально активного «родительского» гена.

NUMT-псевдогены представляют собой фрагменты ДНК митохондрий (мтДНК), встроенные в различные участки ядерного генома. Такие псевдогены обнаружены у самых разных эукариотических организмов, в том числе у человека в количестве 286 [197]. В настоящее время установлено, что в ядерном геноме человека присутствуют последовательности всех генов митохондрий, которые почти равномерно распределены по всем хромосомам. NUMT-псевдогены могут представлять собой как отдельные митохондриальные гены, так и фрагменты мтДНК, включающие два и более соседних гена. Многие NUMT-псевдогены имеют различные мутации: замены, вставки, делеции, дупликации. Механизмы проникновения мтДНК в ядерный геном до конца не изучен. Предполагается, что поврежденная различными эндогенными или экзогенными факторами мтДНК может удаляться из митохондрий, попадать в цитоплазму клеток, затем проникать в ядро и встраиваться в ядерную ДНК. Как считают многие исследователи, встраивание фрагментов мтДНК в ядерный геном может происходить в процессе репарации двунитевых разрывов

по механизму негомологичного соединения концов ДНК (NHEJ). Для некоторых генов (например, гена *PCNA* человека) описано более одного псевдогена [198].

Долгое время псевдогены считались функционально неактивными и рассматривались как эволюционные обломки генов. Однако исследования последнего десятилетия показали важную регуляторную роль этих участков генома. Прежде всего, следует отметить, что значительная часть псевдогенов транскрибируется [199]. Образующиеся при этом РНК реализуют свои регуляторные функции с помощью нескольких механизмов. Недавно было показано, что благодаря высокой гомологии транскрипты псевдогена *PTENP1* и «родительского» гена *PTEN* могут конкурировать друг с другом за регуляторные miРНК, которые связываются со своими сайтами, локализованными в 3'-UTR [200]. Концепция функционального взаимодействия генов и их транскрибируемых псевдогенов на уровне мРНК была сформулирована в общем виде, и было введено понятие конкурирующих эндогенных мРНК (сеРНК).

Рассмотренный выше механизм реализуется в том случае, если в результате транскрипции псевдогена образуется смысловая РНК по отношению мРНК «родительского» гена. Однако существуют псевдогены, которые являются источниками антисмысловых транскриптов. Подобные РНК образуют дуплексы с транскриптами генов-предшественников и могут далее расщепляться соответствующими ферментными системами с образованием эндогенных siРНК [201, 202]. siРНК участвуют в регуляции экспрессии генов-предшественников по механизму РНК-интерференции (подробнее см. выше раздел о miРНК). Например, siРНК, возникающая при участии РНК псевдогена *OCT4-pg5* приводит к подавлению экспрессии гена *OCT4* [203].

Как уже отмечалось выше, процессированные псевдогены могут располагаться в различных участках генома: в межгенных областях, а также экзонах и интронах других генов. В последнем случае последовательность псевдогена может стать новым экзоном гена-хозяина. Подобный процесс появления новых экзонов получил название «экзонизация». В результате транскрипции псевдогена в составе гена-хозяина может образоваться химерная мРНК. Белок, транслируемый с такой мРНК, по своим свойствам будет отличаться от белкового продукта гена-хозяина [195]. То есть, в процессе экзонизации изменяются функции как псевдогена, так и гена-хозяина.

Геном человека содержит ~18 000 псевдогенов, из которых 2/3 – процессированные (<http://www.pseudogene.org>, [204]). При

этом большинство псевдогенов ассоциировано с ограниченным числом семейств активно транскрибируемых генов. Так, гены 79 рибосомных белков человека связаны с 20% всех его псевдогенов, а у гена глицеральдегид-3-фосфатдегидрогеназы таких псевдогенов 62 [205]. Такое большое превышение числа псевдогенов над их генами объясняют недавней вспышкой активности ретротранспозонов в геноме человека. С другой стороны, учитывая описанное выше функционально активное состояние многих псевдогенов, можно предположить, что они могли бы непосредственно участвовать в регуляции экспрессии этих активно транскрибируемых генов. В целом, исследования последних лет приподняли завесу и над этим обширным классом геномных ПП, долго считавшихся некодирующими, обнаружив их разнообразное регуляторное влияние на экспрессию генов.

ПОВТОРЯЮЩИЕСЯ ПОСЛЕДОВАТЕЛЬНОСТИ

Повторяющиеся ПП занимают большую часть генома млекопитающих [206]. За исключением центромерных и теломерных ПП функции остальных повторов все еще остаются загадкой. Прогресс в понимании роли повторяющихся ПП в функционировании генома стал заметен только после разработки новых высокопроизводительных методов анализа ПП и пространственной структуры ДНК и РНК и проведения глобальных исследований генома и транскриптома.

Мобильные генетические элементы

МГЭ, иначе называемые транспозонами, относятся к умеренно повторяющимся ПП генома эукариот. Они составляют 45% генома человека и 40% генома мыши. Свое название МГЭ получили от способности менять свое местоположение в геноме соматических клеток и клеток зародышевой линии. Кроме истинных МГЭ в геноме встречаются их многочисленные фрагменты и копии, инактивированные мутациями. Поскольку цельные МГЭ содержат гены, обеспечивающие их подвижность и выживаемость в геноме, отнесение таких элементов генома к некодирующим последовательностям в данном обзоре условно и отражает наше незнание основных функций транспозонов в эукариотических клетках.

В соответствии с молекулярными механизмами, которые МГЭ используют для своего перемещения в геномных ПП, их разделяют на два больших класса, а именно, ретроэлементы (класс I) и ДНК-транспозоны (класс II) [206]. Ретроэлементы используют для своей мобилизации механизмы, в которых центральную роль играет

обратная транскрипция. На основании особенностей структуры и механизмов репликации ретроэлементы разделяют, в свою очередь, на LTR-содержащие элементы: ретротранспозоны и эндогенные ретровирусы, а также ретроэлементы, не содержащие LTR: длинные диспергированные элементы (LINE – long interspersed elements) и короткие диспергированные элементы (SINE – short interspersed elements) [207, 208]. Перемещение в геноме этих МГЭ происходит в результате их транскрипции, синтеза кДНК на матрице образовавшихся РНК обратной транскриптазой и интеграции кДНК в новый генетический локус (механизм копирования и вставки – copy and paste). В результате таких транспозиций происходит увеличение числа копий ретроэлементов в геноме.

Ретротранспозоны по своей структуре и механизмам репликации напоминают экзогенные ретровирусы. Особенностью их структуры является наличие на концах длинных концевых повторов (LTR – long terminal repeats), которые содержат ПП, участвующие в регуляции их транскрипции и репликации. *LINE-элементы*, иначе называемые длинными ретропозонами, обладают теми же генами, что и ретротранспозоны, однако в отличие от них не имеют LTR. Тем не менее, у них имеются промоторы РНК-полимеразы II, которая осуществляет транскрипцию генов LINE. Поскольку ретротранспозоны и LINE-элементы обладают всем необходимым для перемещения в геноме, их называют автономными транспозонами. *SINE-элементы* (короткие ретропозоны) не являются автономными и для своей транспозиции требуют наличия белковых продуктов экспрессии генов автономных транспозонов. Вблизи 5'-конца они содержат внутренний промотор РНК-полимеразы III, которая осуществляет их транскрипцию.

В отличие от ретротранспозонов, **ДНК-транспозоны** перемещаются в геноме по механизму вырезания и вставки (cut and paste) с участием транспозазы – фермента класса рекомбиназ [206]. В результате происходит вырезание транспозона, сопровождающееся дубликацией (удвоением) короткой ПП нуклеотидов в старом сайте интеграции, и встраивание в новое место генома, как правило, недалеко от старого.

Чрезвычайно высокая представленность МГЭ среди геномных ПП, свидетельствует об их важной роли в эволюции геномов эукариот и позволяет рассматривать их в качестве молекулярных эндосимбионтов эукариотических клеток [209, 210].

В последнее время становится ясно, что МГЭ могут оказывать большое влияние на экспрессию обычных генов эукариотического генома. Прежде всего, в этих процессах участвуют промоторы ретро-транспозонов и ассоциированные с ними регуляторные ПП. Глобальный анализ экспрессии ретро-транспозонов в геномах человека и мыши методом CAGE выявил соответственно ~275 000 и ~44 000 TSS в повторяющихся ПП их геномов, что составило ~31% и ~18% от всех известных TSS этих организмов, хотя уровни их активности значительно ниже, чем TSS обычных генов [40]. Транскрипция, инициированная на повторяющихся ПП, носит тканеспецифический характер. Так, в эмбриональных тканях человека до 30% всех TSS обнаруживается в повторяющихся ПП (16% в ретро-транспозонах, 10% – в сателлитах и 5% в простых повторах). При этом LINE-ПП вносят основной вклад в этот синтез РНК. Транскрипция простых повторов четко преобладает примерно в половине исследованных тканей. Использование ~35% всех TSS, ассоциированных с ретро-транспозонами, регулируется в онтогенезе. При анализе LTR-ретро-транспозонов мышей, принадлежащих семейству VL30, также был обнаружен четкий тканеспецифический характер их транскрипции, которая не происходит в мозге, гипоталамусе и эмбриональных тканях. Синтез большинства обнаруженных в этой работе матричных и антисмысловых транскриптов инициировался на неизвестных ранее промоторах.

Повсеместно распространенная транскрипция повторяющихся ПП оказывает влияние на транскриптом генов, кодирующих белки [40]. Оказалось, что 144 промотора ретро-транспозонов или их фрагментов мышей и 576 аналогичных промоторов человека используются в качестве альтернативных при транскрипции известных генов. Кроме того, ретро-транспозоны, встречающиеся в 3'-UTR более чем четверти генов, экспрессируются и понижают уровень их транскрипции. Двухсторонняя транскрипция, сопровождаемая образованием как обычных, так и антисмысловых РНК, также часто начинается в пределах ретро-транспозонов и, как уже обсуждалось нами выше, обеспечивает поддержание эпигенома и установление пограничных барьеров между функциональными доменами хромосом. Описаны случаи включения энхансеров и инсуляторов транспозонов в транскрипционные сети человека, животных и растений (см. недавний обзор [211]). Транспозоны активны в нормальных тканях мозга млекопитающих и могут оказывать влияние на его метаболизм [212]. LTR используется в качестве альтернативного промотора

для гена эритроидного фактора транскрипции *Pu.1* мышей, где он играет важную функциональную роль в эритропоэзе [213]. Большое число ретротранспозонов активируется деметилированием и экспрессируется в раннем эмбриогенезе млекопитающих, где они могут обеспечивать зиготную индукцию экспрессии генов развивающегося зародыша [214]. Действительно, экспериментальное подавление транскрипции эндогенных ретровирусов и LINE1-элементов сопровождается снижением компетентности эмбрионов к развитию [215]. Таким образом, результаты недавних исследований подтверждают ранее предполагаемую важную роль МГЭ в филогенезе и онтогенезе эукариот, однако их глобальное значение для эукариотического генома еще предстоит выяснить.

Теломеры и центромеры

Значительную долю некодирующих ПП генома млекопитающих составляют простые повторы, из которых построены теломерные и центромерные участки хромосом, выполняющие важные биологические функции.

Теломеры – специализированные генетические локусы, организованные в большие нуклеопротеиновые комплексы, расположенные на концах эукариотических хромосом, обеспечивают их полноценную репликацию и стабильность [216]. Существование линейных хромосом в клетке приводит к необходимости решения, по крайней мере, двух связанных с этим проблем. Во-первых, воспроизведение линейных молекул ДНК в ряду клеточных поколений без привлечения специальных молекулярных механизмов непременно сопровождалось бы недорепликацией их концов и уменьшением размеров молекул в каждом клеточном цикле. Действительно, поскольку репликативные ДНК-полимеразы осуществляют синтез ДНК только в направлении 5'→3', они не могут заполнить одноцепочечную брешь, которая бы возникла после удаления последнего 5'-концевого РНК-прайма. Во-вторых, концы хромосом должны быть защищены от ошибочного распознавания системами репарации, которые в отсутствие такой защиты объединяли бы хромосомы по концам друг с другом. Решение этих двух проблем обеспечивают теломеры.

Теломерная ДНК (тДНК) млекопитающих построена из тандемно повторяющихся гексануклеотидов TTAGGG, и ее суммарная длина у человека составляет 10–15 т.п.н., а у грызунов может достигать 20–50 т.п.н. В основном двухцепочечная тДНК содержит G-богатый 3'-выступающий одноцепочечный конец, с которым взаимодействуют

теломерные белки, и который служит праймером для теломеразы. Одноцепочечный конец может существовать, по крайней мере, в двух альтернативных конформациях, образуя так называемые t-петли и G-квадруплексы [217, 218]. Структура типа t-петли образуется с участием белковых факторов в результате внедрения концевой одноцепочечного участка между цепями тДНК ближайших к нему TTAGGG-повторов. G-квадруплексы формируются стопками из четырех расположенных в одной плоскости остатков G, объединенных Хугстиновскими связями. G-квадруплексы обнаружены *in vivo* и, как предполагается, могут ограничивать элонгацию тДНК теломеразой.

На тДНК формируется белковый комплекс, состоящий из шести полипептидов, получивший название шелтерина. Он предохраняет концы хромосом от слияния друг с другом, что доказано экспериментальной инактивацией отдельных компонентов комплекса [219]. Другим важнейшим компонентом теломерного комплекса является теломераза – обратная транскриптаза, обеспечивающая вместе со вспомогательными белками образование и поддержание тДНК [220]. Интегральной частью теломеразы является теломеразная РНК (TR), которая объединена с каталитической субъединицей (TERT), использующей TR в качестве матрицы при синтезе тДНК. Затравкой при этом служит одноцепочечный 3'-конец тДНК. TR выполняет функции не только матрицы, но и направляет сборку вспомогательного дискеринового белкового комплекса, обеспечивающего стабильность TR и функционирование теломеразы *in vivo*, которая у человека способна добавлять к теломерам более 100 нт за один клеточный цикл. Образовываясь в результате одноцепочечная тДНК превращается в двухцепочечную репликативными ДНК-полимеразами.

Если TR синтезируется повсеместно, то экспрессия TERT строго регулируется в онтогенезе и характерна, в частности, для клеток зародышевой линии, а также эмбриональных стволовых, но не соматических клеток. Вследствие этого после каждого деления соматических клеток длина тДНК в них сокращается из-за вышеупомянутой проблемы репликации концов ДНК, что в конце концов приводит к остановке клеточных делений. Важную роль в регуляции активности теломеразы, играет тримерный белковый комплекс CST (CST1–STN1–TEN1), ограничивающий ее процессивность [221]. Несмотря на гетерохроматизацию, тДНК транскрибируется РНК-полимеразой II с субтеломерных и теломерных III с образованием экпированных и полиаденилированных lncРНК длиной 0.1–9 т.н., которые получили название РНК, содержащие теломерные повторы (Telomeric Repeat containing RNA – TERRA) [222]. При этом только

неполиаденилированные TERRA оказались ассоциированными с хроматином. TERRA частично комплементарна TR, а также может непосредственно взаимодействовать с TERT и подавлять активность теломеразы. Кроме того, она участвует в гетерохроматизации хроматина теломер [223].

Наличие связи между теломерами и пролиферацией клеток четко подтверждается при онкологических заболеваниях. В 90% опухолей наблюдают повышенный уровень активности теломеразы, что способствует иммортализации малигнизированных клеток через увеличение длины тДНК. Длина тДНК у человека наследуется и, по-видимому, может оказывать влияние на продолжительность жизни индивидуума, а также его репродуктивные функции [224]. Многочисленные патологические состояния, связанные с дисфункцией теломеразы и теломеразного комплекса, классификация которых затруднена из-за их большой гетерогенности, получили название теломеропатий [225].

Центромеры представляют собой генетические локусы эукариотических хромосом, контролирующие их расхождение в дочерние клетки в митозе и мейозе (обзоры [226, 227]). Каждая центромера является местом сборки мультибелкового комплекса – кинетохора – обеспечивающего прикрепление хромосомы к микротрубочкам и ее перемещение вдоль веретена деления при кариокинезе. Центромеры человека и приматов образованы ПП α -сателлитной ДНК (сатДНК), построенной из tandemно повторяющихся («голова к хвосту») мономеров длиной 171 п.н. Индивидуальные мономеры обнаруживают между собой 50–70% гомологии и объединяются в новую повторяющуюся единицу – повтор высокого порядка (higher order repeat – HOR) – длиной в 1–3 т.п.н., который, непрерывно повторяясь, образует центромерный локус. Длина организованной таким образом сатДНК составляет 0.25–5 м.п.н. В общей сложности ПП центромер составляют ~5% от всех ПП генома человека.

Каждая индивидуальная хромосома содержит уникальную ПП α -сателлитов, в которых мультимеры HOR содержат свое число tandemных мономеров, что позволяет отличать индивидуальные хромосомы друг от друга. При этом общий размер центромер различается даже у гомологичных хромосом одного человека и между индивидуумами. Полиморфизм характерен и для ПП индивидуальных мономеров HOR: некоторые из них содержат специфический элемент, получивший название CENP-B-бокса, который распознается специфическим ДНК-связывающим белком центромер CENP-B. Другие мономеры маркированы SNP.

Важное функциональное значение сатДНК человека для активности центромер было установлено в опытах по созданию искусственных хромосом. При конструировании минихромосом из X- и Y-хромосом методом «сверху-вниз» в гибридных клетках человек-мышь и человек-курица происходило последовательное уменьшение размеров хромосом вследствие их гомологичной рекомбинации с плазмидной ДНК, содержащей теломерные ПП и селективируемые маркеры. В образовавшихся в итоге минимальных конструкциях, сохранивших способность правильно сегрегировать в дочерние клетки, центромеры были построены из сатДНК. При подходе «снизу-вверх» в дрожжевые или бактериальные искусственные хромосомы встраивали синтетические или клонированные сатДНК, которые также придавали искусственным хромосомам требуемые свойства. Однако не все сатДНК обладали способностью формировать центромеры. Их функциональность проявлялась только при наличии в них нативного CENP-B-бокса.

В свете недавно обнаруженной всепроникающей транскрипции уже не кажется удивительным, что весь центромерный локус, включая перицентромерные области, транскрибируется (обзор [228]). Транскрипты, в том числе и полиаденилированные, обнаруживаются как в ядре, так и в цитоплазме. Перицентромерная сатДНК активно транскрибируется в эмбриогенезе мышей и участвует в гетерохроматизации этих областей хромосом. Разрушение этих транскриптов с использованием антисмысловых технологий приводит к остановке развития [229]. Дерепрессия транскрипции сатДНК обнаружена во многих опухолях эпителия человека, хотя пока остается неясным является ли данная транскрипция причиной или следствием развития опухолевого процесса [230]. Появляются данные об участии транскриптов сатДНК в сборке и поддержании кинетохор [231]. Все это указывает на важные и многогранные функции простых повторяющихся ПП сатДНК в геноме млекопитающих.

III. ЗАКЛЮЧЕНИЕ

Некодирующие белки ПП геномов человека и животных насыщены разнообразными регуляторными элементами и генами ncРНК (см. таблицу). К таким наиболее известным и хорошо изученным элементам относятся, прежде всего, фланкирующие гены *cis*-действующие регуляторные ПП: промоторы, а также ПП, представляющие 5'- и 3'-UTR. Хотя они занимают лишь небольшую часть обсуждаемых геномов, их роль в регуляции экспрессии генов исключительно

Таблица
Содержание известных и предполагаемых функциональных некодирующих последовательностей ДНК в геноме человека

Элементы ДНК	Размер, т.п.н.	Всего в геноме*		Присутствующие функциональные элементы и/или функции	
		Нуклеотиды, м.п.н.	Доля, %		
Мобильные генетические элементы	<1–25	1395	45	Тканеспецифическая регуляция транскрипции генов, кодирующих белки; поддержание эпигенома и установление пограничных барьеров между функциональными доменами хромосом.	
Интроны	<0.1–1000	744	24	5-Кратное повышение информационной емкости генома за счет альтернативного, в том числе межгенного сплайсинга; IME; рекомбинация аллельных генов. Могут содержать промоторы, терминаторы, энхансеры и сайленсеры транскрипции.	
Консервативные ПП, эволюционирующие				Экзоны (30%), интроны (30%) и межгенные ПП (40%), в том числе сайты, гиперчувствительные к ДНКазе, сайты связывания факторов транскрипции, промоторы, UTR, энхансеры, инсуляторы, lncРНК	
		– медленно	130		4.2
		– быстро	254		8.2
satДНК центромер	250–5000	155	5	Место сборки кинетохор, участие транскриптов satДНК в гетерохроматизации хроматина и регуляции развития	
Энхансеры	<1–50	93	3	Сборка белковых комплексов, активирующих или ингибирующих транскрипцию, в том числе и тканеспецифическую	
СрG-островки и ICR	0.2–2	31	1	Регуляция транскрипции генов через метилирование-деметилирование СрG и их окрестностей, в том числе, при импринтинге	
5'-UTR	0.02–3 (0.21**)	4	<0.1	Регуляция трансляции	
3'-UTR	1.3**		<0.1	Регуляция экспрессии генов на посттранскрипционном уровне и трансляции.	
tДНК теломер	10–15	0.23–0.35	<0.1	Поддержание целостности хромосом, регуляция числа делений клеток	
Псевдогены	0.83**	11.9	9	Регуляция транскрипции генов, кодирующих белки (их РНК могут быть «ловушками» miРНК, а также источниками siРНК)	
Инсуляторы	1**	<0.1	<0.1	Предотвращение неспецифического действия энхансеров на промоторы; разделение функциональных доменов хромосом; регуляция V(D)J-рекомбинации в иммуноглобулиновых локусах.	
S/MAR	5	<0.1	<0.1	Организация функциональных доменов хромосом в интерфазном ядре	
Промоторы		<0.1	<0.1	Регуляция транскрипции	
Гены некодирующих РНК		<0.1–23	>90 ?	Регуляция экспрессии генов на всех уровнях	

* – Размер гаплоидного генома человека – 3100 м.п.н.;
 ** – средний размер.

высока. Многие гены используют для регуляции своей экспрессии альтернативные регуляторные ПП, поэтому общее число промоторов и UTR заметно превышает число генов, аннотированных в геномах.

Более массивными регуляторными элементами эукариотических геномов являются энхансеры, локус-контролирующие области, инсуляторы и S/MAR-ПП. Доля энхансеров вместе с суперэнхансерами в геноме человека приближается к 3%. Эти элементы играют ключевую роль в установлении высокоупорядоченной транскрипции в разных тканях организма и в процессе его онтогенеза. Инсуляторы и S/MAR-ПП, хотя и занимают небольшую (<0.1%) часть генома человека, замечательны тем, что пространственно и функционально организуют протяженные (~1 м.п.н.) домены хромосом, экспрессирующие гены, в интерфазных ядрах. Наблюдаемые перемещения генетических локусов от периферии к центру ядер и обратно, связанные с их активацией-деактивацией, указывают на еще одну редко обсуждаемую функцию некодирующих ПП, которые могут составлять значительную часть таких доменов. На наш взгляд, эволюционное включение некодирующих ПП в промежутки между экспрессирующимися генами, придает хромосомным доменам гибкость, требуемую для поддержания динамического внутриядерного состояния работающих генетических локусов, их подвижности внутри хромосомных территорий. Выполнение такой линкерной функции уже не накладывает жестких требований к первичной структуре этих некодирующих ПП, а под давлением отбора оказываются лишь их линейные размеры. Если это предположение верно, то трудно ожидать от меггенных линкерных ПП значительной консервативности, поскольку таковой не требуется для выполнения ими своих функций.

Значительную часть (~1%) генома человека составляют и геномные сайты CGI, маркирующие участки метилирования ДНК, локализованные в межгенных областях и вблизи 5'-концов многих генов, а также внутри самих генов, и участвующие в регуляции их транскрипции. С учетом ПП, фланкирующих CGI (так называемых «берегов» – shores и «шельфа» – shelves), которые распространяются на ~4 т.п.н. в каждую сторону, их доля в геноме может быть еще выше. Недавно Г.А. Романовым и др. была разработана новая концепция, с неожиданной стороны освещающая роль метилирования ДНК в онтогенезе эукариотических организмов [232]. Авторами было установлено, что метилируемые CpG-динуклеотиды неслучайным образом входят в состав разных кодонов в экзонах генов человека и животных. Спонтанное дезаминирование 5-mC, происходящее на протяжении жизни особи, сопровождается образованием T и воз-

никновением соответствующих мутаций. Превращение в результате таких мутаций осмысленного кодона в стоп-кодон или кодон, кодирующий неблагоприятную для белка аминокислоту, приводит к инактивации соответствующего белка или фермента. Такие кодоны авторы называют «опасными». В соответствии с предложенной моделью, частоты опасных кодонов специфичны для организмов разных таксономических групп и негативно коррелируют с продолжительностью жизни особей, а паттерны метилирования ДНК в опасных кодонах представляют своеобразный код старения биологического вида.

Псевдогены – еще один обширный класс некодирующих ПП, которые совсем недавно считались нефункциональными обломками соответствующих генов. В настоящее время убедительно показано, что те из них, которые транскрибируются в прямом или противоположном направлении с образованием соответственно мРНК или антисмысловой РНК, являются источниками ссРНК или siРНК, участвующих в регуляции экспрессии породивших их генов на уровне трансляции.

Исследования последних лет, проведенные на интронах, которые составляют ~24% всех ПП генома человека, также обнаружили их разнообразные функции в регуляции экспрессии генов. Альтернативный сплайсинг и *trans*-сплайсинг предшественников мРНК, ставшие возможными благодаря наличию интронов, в 5 раз повышает информационную емкость генома человека. Глобальная функциональная значимость интронов также следует из опосредованного интронами усиления экспрессии генов (феномен IME). Кроме того интроны являются местом расположения многих регуляторных элементов ДНК – альтернативных промоторов и терминаторов транскрипции, транскрипционных энхансеров и сайленсеров, а также генов miРНК и lncРНК.

Обнаруженная недавно повсеместная транскрипция, охватывающая 99% ПП генома человека, открывает новые горизонты в исследовании функциональной роли образующихся в итоге многочисленных ncРНК. Уже установлено, что присутствующие среди транскриптов малые и длинные некодирующие РНК оказывают регуляторное действие на всех уровнях экспрессии эукариотических генов. Выявленные в ходе недавних исследований регуляторные сети, формирующиеся с участием клеточных РНК, приоткрыли вершину нового неисследованного айсберга регуляторных механизмов, управляющих экспрессией эукариотических генов.

Наконец, повторяющиеся ПП, которые составляют основную часть генома млекопитающих, больше не представляются пустыней в геномном ландшафте. Простые повторы, образующие центромеры и теломеры хромосом, транскрибируются и играют важную роль в функционировании этих генетических локусов и клеток в целом. Транспозоны активно участвуют в регуляции экспрессии обычных генов, а также в формировании и поддержании эпигенома и пограничных барьеров между функциональными доменами хромосом.

Несмотря на разнообразие функций некодирующих ПП генома эукариот, которое мы продемонстрировали в данном обзоре, значение большинства из них для жизни клетки и организма все-таки остается непонятным. Установленная нами ранее защитная функция некодирующих ПП в отношении генов от эндогенных химических мутагенов могла бы быть первичной по отношению ко всем остальным (обзор [14] и ссылки в нем). Действительно, современные эукариотические организмы живут и эволюционируют в атмосфере кислорода. Образующиеся в процессе жизнедеятельности активные формы кислорода непрерывно нарушают целостность их ДНК, ежедневно вызывая до 200 000 повреждений в каждой клетке, часть которых, в отсутствие репарации, может превращаться в мутации. Поскольку некодирующие ПП представляют большую часть генома млекопитающих, такого рода мутации происходят преимущественно в некодирующих ПП без вредных последствий для клеток и организма в целом. После эволюционного приобретения геномом современного размера с определенным соотношением кодирующих и некодирующих ПП, в организме устанавливается равновесие, при котором система репарации повреждений ДНК и защита генов некодирующими ПП обеспечивают приемлемый уровень эндогенного мутагенеза, совместимый с жизнью многоклеточного организма. Вместе с тем, внутригеномная экспансия некодирующих ПП создает условия для их дальнейшего эволюционирования, сопровождаемого появлением новых функций, большую часть которых еще предстоит установить в ходе дальнейших исследований.

Благодарность. Авторы благодарят профессора Г.А. Романова за критическое прочтение рукописи и ценные замечания.

ЛИТЕРАТУРА

1. Griffiths, P.E., Stotz, K. (2006) Genes in the postgenomic era, *Theor. Med. Bioeth.*, **27** (6), 499–521.
2. El-Hani, C.N. (2007) Between the cross and the sword: The crisis of the gene concept, *Genet. Mol. Biol.*, **30** (2), 297–307.
3. Boyle, A.P., Araya, C.L., Brdlik, C., Cayting, P., Cheng, C., Cheng, Y., Gardner, K., Hillier, L. W., Janette, J., Jiang, L. et al. (2014) Comparative analysis of regulatory information and circuits across distant species, *Nature*, **512**, 453–456.
4. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012) An integrated encyclopedia of DNA elements in the human genome, *Nature*, **489**, 57–74.
5. The ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature*, **447**, 799–816.
6. 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A. et al. (2010) A map of human genome variation from population-scale sequencing, *Nature*, **467** (7319), 1061–73.
7. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes, *Nature*, **491** (7422), 56–65.
8. Roberts, N.J., Vogelstein, J.T., Parmigiani, G., Kinzler, K.W., Vogelstein, B., Velculescu, V.E. (2012) The predictive capacity of personal genome sequencing, *Sci. Transl. Med.*, **4** (133).
9. Gjuvsland, A.B., Vik, J., Beard, D.A., Hunter, P.J. and Omholt, S.W. (2013) Bridging the genotype–phenotype gap: what does it take? *J. Physiol.*, **591** (8), 2055–2066.
10. Van der Sijde, M.R., Ng, A., Fu, J. (2014) Systems genetics: From GWAS to disease pathways, *Biochim. Biophys. Acta*, **1842** (10), 1903–1909.
11. Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A. et al. (2011) The Reality of Pervasive Transcription, *PLoS Biol.*, **9** (7).
12. Harmston, N., Baresic, A., Lenhard, B. (2013) The mystery of extreme non-coding conservation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **368**, 1471–2970.
13. Gregory, T.R. (2005) The C-value enigma in plants and animals: a review of parallels and an appeal for partnership, *Ann. Bot.*, **95** (1), 133–46.
14. Патрушев Л.И., Минкевич И.Г. (2007) Проблема размера генома эукариот, *Успехи биологической химии*, **47**, 293–370.
15. Kellis, M., Wold, B., Snyder, M.P., Bernstein, B.E., Kundaje, A., Marinov, G.K., Ward, L.D., Birney, E., Crawford, G.E., Dekker, J. et al. (2014) Defining functional DNA elements in the human genome, *Proc. Natl. Acad. Sci. USA*, **111** (17), 6131–6138.
16. Ponting, C.P., Hardison, R.C. (2011) What fraction of the human genome is functional? *Genome Res.*, **21**, 1769–1776.
17. Lunter, G., Ponting, C.P., Hein, J. (2006) Genome-wide identification of human functional DNA using a neutral indel model, *PLoS Comput Biol.*, **2**, e5.
18. Chiaromonte, F., Weber, R.J., Roskin, K.M., Diekhans, M., Kent, W.J., Haussler, D. (2003) The share of

- human genomic DNA under selection estimated from human-mouse genomic alignments, *Cold Spring Harb. Symp. Quant. Biol.*, **68**, 245–54.
19. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E. et al. (2011) A high-resolution map of human evolutionary constraint using 29 mammals, *Nature*, **478** (7370), 476–482.
 20. Rands, C.M., Meader, S., Ponting, C.P., Lunter, G. (2014) 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage, *PLoS Genet.*, **10** (7).
 21. Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P. and Gingeras, T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22, *Science*, **296**, 916–919.
 22. Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P. M., Nelson, F.K., Miller, P., Gerstein, M., Weissman, S. and Snyder, M. (2003) The transcriptional activity of human chromosome 22, *Genes Dev.*, **17**, 529–540.
 23. Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M. and Snyder, M. (2004) Global identification of human transcribed sequences with genome tiling arrays, *Science*, **306**, 2242–2246.
 24. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tamma, H., Helt, G. et al. (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution, *Science*, **308**, 1149–1154.
 25. Johnson, J.M., Edwards, S., Shoemaker, D., Schadt, E.E. (2005) Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.*, **21** (2), 93–102.
 26. Kapranov, P., Sementchenko, V.I., Gingeras, T.R. (2003) Beyond expression profiling: next generation uses of high density oligonucleotide arrays, *Brief. Funct. Genomic Proteomic*, **2**, 47–56.
 27. Wang, Z., Gerstein, M., Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics, *Nat Rev Genet.*, **10** (1), 57–63.
 28. Mutz, K.O., Heilkenbrinker, A., Lönne, M., Walter, J.G., Stahl, F. (2013) Transcriptome analysis using next-generation sequencing, *Curr. Opin. Biotechnol.*, **24** (1), 22–30.
 29. Okoniewski, M.J., Miller, C.J. (2006) Hybridization interactions between probe sets in short oligo microarrays lead to spurious correlations, *BMC Bioinformatics*, **7**, 276.
 30. Royce, T.E., Rozowsky, J.S., Gerstein, M.B. (2007) Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification, *Nucleic Acids Res.*, **35** (15).
 31. Pareek, C.S., Smoczynski, R., Tretyn A. (2011) Sequencing technologies and genome sequencing, *Appl. Genetics*, **52**, 413–435.
 32. Schadt, E.E., Turner, S., Kasarskis, A. (2010) A window into third-generation sequencing, *Hum. Mol. Genet.*, **19**, 227–240.
 33. Costa, V., Angelini C., De Feis, I., Ciccociola, A. (2010) Uncovering the complexity of transcriptomes with RNA-seq, *J. Biomed. Biotechnol.*, **2010**, 853916.
 34. Mortazavi, A., Williams, B. A., Mc Cue, K., Schaeffer, L., Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nature Methods*, **5**, 621–628.
 35. Harbers, M., Carninci, P. (2005) Tag-based approaches for transcriptome

- research and genome annotation, *Nature Methods*, **2** (7), 495–502.
36. Jacquier, A. (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs, *Nature Rev. Genet.*, **10**, 833–844.
 37. Kapranov, P., St. Laurent, G. (2012) Dark matter RNA: existence, function, and controversy. *Front. Genet.*, **3** (60).
 38. Kapranov, P., St. Laurent, G., Raz, T., Ozsolak, F., Reynolds, C.P., Sorensen, P.H., Reaman, G., Milos, P., Arceci, R.J., Thompson, J.F. and Triche, T.J. (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is «dark matter» unannotated RNA. *BMC Biol.*, **8**, 149.
 39. Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P. and Hayashizaki, Y. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage, *Proc. Natl. Acad. Sci. USA*, **100**, 15776–15781.
 40. Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe A.L., Lassmann, T. et al. (2009) The regulated retrotransposon transcriptome of mammalian cells, *Nature Genet*, **41**, 563–571.
 41. Maston, G.A., Evans, S.K., Green, M.R. (2006) Transcriptional regulatory elements in the human genome, *Annu. Rev. Genomics Hum. Genet.*, **7**, 29–59.
 42. Lenhard, B., Sandelin, A., Carninci, P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation, *Nature Rev. Genet.*, **13**, 233–245.
 43. Kadonaga, J.T. (2012) Perspectives on the RNA polymerase II core promoter, *Wiley Interdiscip. Rev. Dev. Biol.*, **1** (1), 40–51.
 44. Thomas, M.C., Chiang C.M. (2006) The general transcription machinery and general cofactors, *Crit. Rev. Biochem. Mol. Biol.*, **41**, 105–178.
 45. Massari, M.E., Murre, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucariotic organisms, *Mol. Cell. Biol.*, **20** (2), 429–440.
 46. Nakahata, Y., Yoshida, M., Takano, A., Soma, H., Yamamoto, T., Yasuda, A., Nakatsu, T. and Takumi, T. (2008) A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes, *BMC Mol. Biol.*, **9**, 1.
 47. Santoro, N., Johansson, N. and Thiele, D.J. (1998) Heat shock element architecture is an important determinant in the temperature and transactivation domain requirements for heat shock transcription factor, *Mol. Cell. Biol.*, **18** (11), 6340–6352.
 48. Perry, R.P. (2005) The architecture of mammalian ribosomal protein promoters, *BMC Evol Biol.*, **5**, 15.
 49. Панкратова Е.В. (2008) Альтернативные промоторы в реализации генетической информации, *Молекулярная биология*, **42** (3), 422–433.
 50. Bestor, T.H. (2000) The DNA methyltransferases of mammals, *Hum. Mol. Genet.*, **9** (16), 2395–2402.
 51. Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otilar, R.P., Myers, R.M. (2004) An abundance of bidirectional promoters in the human genome, *Genome Research*, **14** (1), 62–66.
 52. Yang, M.Q and Elnitski, L. L. (2008) Diversity of core promoter elements comprising human bidirectional promoters, *BMC Genomics*, **9** (Suppl. 2), S3.

53. Rachakonda, P.S., Hosen I., de Verdeir, P.J. Fallah, M., Heidenreich, B., Ryk, C., Wiklund, N.P., Steineck, G., Schadendorf, D., Hemminki, K. and Kumar, R. (2013) TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism, *Proc. Natl. Acad. Sci. USA*, **110** (43), 17426–17431.
54. Савинкова Л.К., М. П. Пономаренко М.П., Пономаренко П.М., Драчкова И.А., Лысова М.В., Аршинова Т.В., Колчанов Н.А. (2009) Полиморфизмы ТАТА-боксов промоторов генов человека и ассоциированные с ними наследственные патологии, *Биохимия*, **74** (2), 149–163.
55. Mignone, F., Gissi, C., Liuni, S. and Pesole, G. (2002) Untranslated regions of mRNAs, *Genome Biol.*, **3** (3).
56. Zhang, T., Haws, P., Wu, Q. (2004) Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation, *Genome Res.*, **14** (1), 79–89.
57. Iakova, P., Wang, G.L., Timchenko, L., Michalak, M., Pereira-Smith, O.M., Smith, J.R. and Timchenko, N.A. (2004) Competition of CUGBP1 and calreticulin for the regulation of p21 translation determines cell fate, *The EMBO Journal*, **23** (2), 406–417.
58. Araujo, P.R., Yoon, K., Ko, D., Smith, A.D., Qiao, M., Suresh, U., Burns, S.C., Penalva, L.O.F. (2012) Before it gets started: Regulating translation at the 5'UTR, *Compar. Funct. Genomics*, **2012** (475731).
59. Wethmar, K., Smink, J.J., Leutz, A. (2010) Upstream open reading frames: molecular switches in (patho)physiology, *Bioessays*, **32** (10), 885–893.
60. Chatterjee, S., Pal, J.K. (2009) Role of 5'- and 3'-untranslated regions of mRNAs in human diseases, *Biol. Cell*, **101** (5), 251–62.
61. Matoulkova, E., Michalova, E., Vojtesek, B., Hrstka, R. (2012) The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells, *RNA Biol.*, **9** (5), 563–76.
62. Zhao, W., Blagev, D., Pollack, J.L., Erle, D.J. (2011) Toward a systematic understanding of mRNA 39 untranslated regions, *Proc. Am. Thorac. Soc.*, **8**, 163–166.
63. Smith, R.W.P., Blee, T.K.P., Gray, N.K. (2014) Poly(A)-binding proteins are required for diverse biological processes in metazoans, *Biochem. Soc. Trans.*, **42**, 1229–1237.
64. Chen, J., Kastan, M.B. (2010) 5'-3'-UTR interactions regulate p53 mRNA translation and provide a target for modulating p53 induction after DNA damage, *Genes Dev.*, **24**, 2146–2156.
65. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., Burge, C.B. (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites, *Science*, **320**, 1643–1647.
66. Hitti, E., Khabar, K.S.A. (2012) Sequence variations affecting AU-rich element function and disease. *Front. Biosci.*, **17**, 1846–1860.
67. Halees, A.S., Hitti, E., Al-Saif, M., Mahmoud, L., Vlasova-St. Louis, I.A., Beisang, D.J., Bohjanen, P.R. and Khabar, K. (2011) Global assessment of GU-rich regulatory content and function in the human transcriptome, *RNA Biol.*, **8**, 681–691.
68. Hentze, M.W., Muckenthaler, M.U., Andrews, N.C. (2004) Balancing acts: molecular control of mammalian iron metabolism, *Cell*, **117**, 285–297.
69. Chavatte, L., Brown, B.A., Driscoll, D.M. (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes, *Nature Struct. Mol. Biol.*, **12**, 408–16.

70. Park, E., Maquat, L.E. (2013) Staufen-mediated mRNA decay, *Wiley Interdiscip. Rev. RNA*, **4** (4), 423–435.
71. Nguyen, A. (2000) Prothrombin G20210A polymorphism and thrombophilia, *Mayo Clin Proc.*, **75**, 595–604.
72. Varol, N., Conac, E., Gurocak, S., Sozen, S. (2011) The realm of micro RNAs in cancers, *Mol. Biol. Rep.*, **38**, 1079–1089.
73. Li, E., Zhang, Y. (2014) DNA methylation in mammals, *Cold Spring Harb. Perspect Biol.*, **6** (5).
74. Medvedeva, Y.A., Khamis, A.M., Kulakovskiy, I.V., Ba-Alawi, W., Bhuyan, M.S.I., Kawaji, H., Lassmann, T., Harbers, M., Forrest, A.R., Bajic, V.B., FANTOM consortium. (2013) Effects of cytosine methylation on transcription factor binding sites, *BMC Genomics*, **15**, 119.
75. Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K. and Schübeler, D. (2011) DNA binding factors shape the mouse methylome at distal regulatory regions, *Nature*, **480**, 490–495.
76. Bestor, T.H., Bourchis, D. (2004) Transposon silencing and imprint establishment in mammalian germ cells, *Cold Spring Harb. Symp. Quant. Biol.*, **69**, 381–387.
77. Maunakea, A.K., Nagarajan, R.P., Bilienky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G. et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters, *Nature*, **466** (7303), 253–257.
78. Plongthongkum, N., van Eijk, K.R., de Jong, S., Wang, T., Sul, J.H., Boks, M.P., Kahn, R.S., Fung, H.L., Ophoff, R.A. and Zhang, K. (2014) Characterization of Genome-Methylome Interactions in 22 Nuclear Pedigrees, *PLoS ONE*, **9** (7).
79. Paul, D.S., Beck, S. (2014) Advances in epigenome-wide association studies for common diseases. *Trends Mol. Med.*, pii: S1471–4914(14)00115–4.
80. Hackett, J.A., Surani, M.A. (2012) DNA methylation dynamics during the mammalian life cycle, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **368** (1609).
81. Illingworth, R.S., Bird, A.P. (2009) CpG islands – «A rough guide», *FEBS Letters*, **583**, 1713–1720.
82. Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M. et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores, *Nature Genet.*, **41**, 178–186.
83. Kozlenkov, A., Roussos, P., Timashpolsky, A., Barbu, M., Rudchenko, S., Bibikova, M., Klotzle, B., Byne, W., Lyddon, R., Di Narzo, A.F. et al. (2014) Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. *Nucleic Acids Research*, **42** (1), 109–127.
84. Blackledge, N.P., Thomson, J.P. and Skene, P.J. (2013) CpG Island Chromatin Is Shaped by Recruitment of ZF-CxxC Proteins, *Cold Spring Harb. Perspect. Biol.*, **5**, a018648
85. Vinson, C., Chatterjee, R. (2012) CG methylation, *Epigenomics*, **4** (6), 655–663.
86. Sharp, A. J., Stathaki, E., Migliavacca, E., Brahmachary, M. Montgomery, S.B., Dupre, Y., Antonarakis, S.E. (2011) DNA methylation profiles of human active and inactive X chromosomes, *Genome Res.*, **21**, 1592–1600.
87. Barlow, D.P., Bartolomei, M.S. (2014) Genomic Imprinting in Mammals. *Cold Spring Harb. Perspect. Biol.*, **6**, 018382.

- 87a. Wang L., Zhang J., Duan J., Gao X., Zhu W., et al. (2014) Programming and inheritance of parental DNA methylomes in mammals. *Cell*, **157**, 979–991.
88. Unoki, M., Nakamura, Y. (2003). Methylation at CpG islands in intron 1 of EGR2 confers enhancer-like activity, *FEBS Letters*, **554**, 67–72.
89. Rakyan, V.K., Down, T.A., Balding, D.J., Beck, S. (2011) Epigenome-wide association studies for common human diseases, *Nature Rev. Genet.*, **12**, 529–541.
90. Dick, K., Nelson, C.P., Tsaprouni, L., Sandling, J.K., Aïssi, D., Wahl, S., Meduri, E., Morange, P.E., Gagnon, F., Grallert, H., Waldenberger, M., Peters, A., Erdmann, J., Hengstenberg, C., Cambien, F., Goodal, I.A.H., Ouwehand, W.H., Schunkert, H., Thompson, J.R., Spector, T.D., Gieger, C., Tréguët, D.A., Deloukas, P. and Samani, N.J. (2014) DNA methylation and body-mass index: a genome-wide analysis, *Lancet*, **383**, 1990–1998.
91. Rose, A.B. (2008) Intron-mediated regulation of gene expression, *Curr. Top. Microbiol. Immunol.*, **326**, 277–290.
92. Bang, M.L., Centner, T., Fornoff, F., Geach, A.J., Gotthardt, M., McNabb, M., Witt, C.C., Labeit, D., Gregorio, C.C., Granzier, H. and Labeit, S. (2001) The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system, *Circ.Res.*, **89**, 1065–1072.
93. Grzybowska, E.A. (2012) Human intronless genes: Functional groups, associated diseases, evolution, and mRNA processing in absence of splicing, *Biochem. Biophys. Res. Commun.*, **424**, 1–6.
94. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome, *Nature*, **409**, 860–921.
95. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001) The sequence of the human genome, *Science*, **291**, 1304–1351.
96. Koonin, E.V. (2006) The origin of introns and their role in eukaryogenesis: A compromise solution to the intron-early versus intron-late debate? *Biol Direct.*, **1**, 22.
97. Irimia, M., Roy, S.W. (2014) Origin of spliceosomal introns and alternative splicing, *Cold Spring Harb. Perspect. Biol.*, **6** (6).
98. Gilbert, W. (1987) The exon theory of genes, *Cold Spring Harb Symp Quant Biol.*, **52**, 901–905.
99. Elliott, D.J. (2014) Illuminating the Transcriptome through the Genome, *Genes*, **5**, 235–253.
100. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F. et al. (2012) Landscape of transcription in human cells, *Nature*, **489**, 101–108.
101. Harrison, P.M., Kumar, A., Lang, N., Snyder, M., Gerstein, M. (2002) A question of size: The eukaryotic proteome and the problems in defining it, *Nucleic Acids Res.*, **30**, 1083–1090.
102. Amrani, N., Ganesan, R., Kerestini, S., Mangus, D.A., Ghosh, S. and Jacobson, A. (2004) A faux 30-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay, *Nature*, **432**, 112–118.
103. Buckley, P.T., Lee, M.T., Sul, J.Y., Miyashiro, J.Y., Bell, T.J., Fisher, S.A., Kim, J. and Eberwine, J. (2011) Cytoplasmic intron sequence retaining transcripts can be dendritically targeted via ID element retrotransposons, *Neuron*, **69**, 877–884.

104. Sorek, R., Shamir, R., Ast, G. (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet.*, **20**, 68–71.
105. Roy, S.W., Irimia, M. (2008) Intron mis-splicing: No alternative? *Genome Biol.*, **9** (2), 208.
106. Lasda, E.L., Blumenthal T. (2011) Trans-splicing, *WIREs RNA*, **2**, 417–434.
107. Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A. and Sorek, R. (2006) Transcription-mediated gene fusion in the human genome, *Genome Res.*, **16**, 30–36.
108. McManus, C.J., Duff, M.O., Eipper-Mains, J., Graveley B.R. (2010) Global analysis of trans-splicing in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, **107**, 12975–12979.
109. Fang, W., Wei, Y., Kang, Y., Landweber, L.F. (2012) Detection of a common chimeric transcript between human chromosomes 7 and 16, *Biology Direct.*, **7**, 49.
110. Hu, G.-J., Chen, J., Zhao, X.-N., Xu, J.-J., Guo, D.-Q., Lu, M., Zhu, M., Xiong, Y., Li, Q., Chang, C.C. et al. (2013) Production of ACAT1 56-kDa isoform in human cells via trans-splicing involving the ampicillin resistance gene, *Cell Research*, **23**, 1007–1024.
111. Wu, C.-S., Yu, C.-Y., Chuang, C.-Y., Hsiao, M., Kao, C.F., Kuo, H.C., Chuang, T.J. (2014) Integrative transcriptome sequencing identifies trans-splicing events with important roles in human embryonic stem cell pluripotency, *Genome Res.*, **24**, 25–36.
112. Gingeras, T.R. (2009) Implications of chimaeric non-co-linear transcripts, *Nature*, **461**, 206–211.
113. Frenkel-Morgenstern, M., Lacroix, V., Ezkurdia, I., Levin, Y., Gabashvili, A., Prilusky, J., Del Pozo, A., Tress, M., Johnson, R., Guigo, R. and Valencia, A. (2012) Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts, *Genome Res.*, **22**, 1231–1242.
114. Greger, L., Su, J., Rung, J., Ferreira, P.G., Geuvadis consortium, Lappalainen, T., Dermitzakis, E.T., Brazma, A. (2014) Tandem RNA Chimeras Contribute to Transcriptome Diversity in Human Population and Are Associated with Intronic Genetic Variants, *PLoS ONE*, **9** (8).
115. Fedorova, L., Fedorov, A. (2003) Introns in gene evolution, *Genetica*, **118** (2–3), 123–131.
116. Wang, Z., Burge, C.B. (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code, *RNA*, **14**, 802–813.
117. Patel, A.A., McCarthy, M., Steitz, J.A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression, *The EMBO Journal*, **21** (14), 3804–3815.
118. Lewandowska, M.A. (2013) The missing puzzle piece: splicing mutations, *Int. J. Exp. Pathol.*, **6** (12), 2675–2682.
119. Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J. and Axel, R. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell*, **126** (2), 403–413.
120. Mousavi, K., Zare, H., Koulis, M., Sartorelli, V. (2014) The emerging roles of eRNAs in transcriptional regulatory networks, *RNA Biol.*, **11** (2), 106–110.
121. Bulger, M., Groudine, M. (2010) Enhancers: the Abundance and Function of Regulatory Sequences Beyond Promoters, *Dev. Biol.*, **339** (2), 250–257. doi:10.1016/j.ydbio.2009.11.035
122. Zentner, G.E., Tesar, P.J., Scacheri, P.C. (2011) Epigenetic signatures distinguish multiple classes of en-

- hancers with distinct cellular functions, *Genome Res.*, **21**, 1273–1283.
123. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, *Nature Genet.*, **39**, 311–318.
 124. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers, *Nature*, **457**, 854–858.
 125. Ong, C.T., Corces, V.G. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression, *Nature Rev. Genet.*, **12** (4), 283–293.
 126. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression, *Nature*, **459**, 108–112.
 127. Cho, K.W.Y. (2012) Enhancers: WIRE Review, *WIREs Dev. Biol.*, **1** (4), 469–478.
 128. Spitz, F., Furlong, E.E.M. (2012) Transcription factors: from enhancer binding to developmental control, *Nature Rev. Genet.*, **13** (9), 613–626.
 129. Maston, G.A., Landt, S.G., Snyder, M., Green, M.R. (2012) Characterization of enhancer function from genome-wide analyses, *Annu. Rev. Genomics Hum. Genet.*, **13**, 29–57.
 130. Barolo, S. (2012) Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays*, **34** (2), 135–141.
 131. Hong, J.W., Hendrix, D.A., Levine, M.S. (2008) Shadow enhancers as a source of evolutionary novelty, *Science*, **321**(5894).
 132. Abbasi, A.A., Pappadopoulos, Z., Malik, S., Bangs, F., Schmidt, A., S. Koch, Lopez-Rios J. and Grzeschik, K.H. (2010) Human intronic enhancers control distinct sub-domains of Gli3 expression during mouse CNS and limb development, *BMC Devel. Biol.*, **10**, 44.
 - 132a. Zhou, X., Sigmund, C.D. (2008) The chorionic enhancer is dispensable for regulated expression of the human renin gene. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, **294** (2), R279–R287.
 133. Blackwood, E.M., Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action, *Science*, **281**, 60–63.
 134. Krivega, I., Dean, A. (2012) Enhancer and promoter interactions-long distance calls, *Curr. Opin. Genet. Devel.*, **22**, 79–85.
 135. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science*, **326**, 289–293.
 136. Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. et al. (2010) Widespread transcription at neuronal activity-regulated enhancers, *Nature*, **465**, 182–187.
 137. Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., Guigo, R. and Shiekhattar, R. (2010) Long noncoding RNAs with enhancer-like function in human cells, *Cell*, **143**, 46–58.

138. Ørom, U.A., Shiekhattar, R. (2013) Long Noncoding RNAs usher in a new era in the biology of enhancers. *Cell*, **154**, 1190–1193.
139. Kowalczyk, M.S., Hughes, J.R., Garrick, D., Lynch, M.D., Sharpe, J.A., Sloane-Stanley, J.A., McGowan, S.J., De Gobbi, M., Hosseini, M., Vernimmen, D. et al. (2012) Intragenic enhancers act as alternative promoters, *Mol. Cell*, **45**, 447–458.
140. Lam, M.T.Y., Li, W., Rosenfeld, M.G., Glass, C.K. (2014) Enhancer RNAs and regulated transcriptional programs, *Trends Biochem. Sci.*, **39**, 170–182.
141. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X. (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation, *Nature*, **498**, 516–520.
142. Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., de Laat, W. and Agami, R. (2013) eRNAs are required for p53-dependent enhancer activity and gene transcription, *Mol. Cell*, **49**, 524–535.
143. Xu, J., Watts, J.A., Pope, S.D., Gadue, P., Kamps, M., Plath, K., Zaret, K.S., and Smale, S.T. (2009). Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells, *Genes Dev.*, **23**, 2824–2838.
144. Reynolds, N., O’Shaughnessy, A., Hendrich, B. (2013) Transcriptional repressors: multifaceted regulators of gene expression, *Development*, **140**, 505–512.
145. Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences, *Nature*, **444**, 499–502.
146. Navratilova, P., Fredman, D., Hawkins, T.A., Turner, K., Lenhard, B., Becker, T.S. (2009) Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes, *Dev. Biol.*, **327**, 526–540.
147. Herz, H.-M., Hu, D., Shilatifard, A. (2014) Enhancer Malfunction in Cancer, *Mol. Cell*, **53** (6), 859–866.
148. Dean, A. (2006) On a chromosome far, far away: LCRs and gene expression, *TRENDS in Genet.*, **22** (1), 38–45.
149. Liang, S., Moghimi, B., Yang, T.P., Strouboulis, J., Bungert, J. (2008) Locus control region mediated regulation of adult β -globin gene expression, *J. Cell Biochem.*, **105** (1), 9–16.
150. Yankulov, K. (2013) Dynamics and stability: epigenetic conversions in position effect variegation, *Biochem. Cell Biol.*, **91**, 6–13.
151. Farrell, C.M., West, A.G., Felsenfeld, G. (2002) Conserved CTCF insulator elements flank the mouse and human b-globin loci, *Mol. Cell Biol*, **22**, 3820–3831.
152. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee G.R., Flavel, R.A. (2005) Interchromosomal associations between alternatively expressed loci, *Nature*, **435**, 637–645.
153. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. and Fraser, P. (2000) Intergenic transcription and developmental remodeling of chromatin subdomains in the human b-globin locus, *Mol. Cell*, **5**, 377–386.
154. Noordermeer, D., Branco, M.R., Splinter, E., Klous, P., van Ijcken,

- W., Swagemakers, S., Koutsourakis, M., van der Spek, P., Pombo, A., and de Laat, W. (2008). Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region, *PLoS Genet.*, **4**, e1000016.
155. Ragozy, T., Bender, M.A., Telling, A., Byron, R., and Groudine, M. (2006). The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation, *Genes Dev.*, **20**, 1447–1457.
156. Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A., Smith, A.J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R., et al. (2008) Association between active genes occurs at nuclear speckles and is modulated by chromatin environment, *J. Cell Biol.*, **182**, 1083–1097.
157. Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R. Oltz, E.M. and Sen, R. (2011) Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus, *Cell*, **147** (2), 332–343.
158. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions, *Nature*, **485**, 376–380.
159. Van Bortle, K., Corces, V.G. (2013) The role of chromatin insulators in nuclear architecture and genome function. *Curr. Opin. Genet. Dev.*, **23** (2), 212–218.
160. Chetverina, D., Aoki, T., Erokhin, M., Georgiev, P., Schedl, P. (2013) Making connections: Insulators organize eukaryotic chromosomes into independent cis-regulatory networks, *Bioessays*, **36**, 163–172.
161. Van Bortle, K., Corces, V.G. (2012) tDNA insulators and the emerging role of TFIIC in genome organization, *Transcription*, **3** (6), 277–284.
162. Jothi, R., Cuddapah, S., Barski, A., Cui, K., Zhao, K. (2008). Genomewide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.*, **36**, 5221–5231.
163. Moqtaderi, Z., Wang, J., Raha, D., White, R.J., Snyder, M., Weng Z., Struhl, K. (2010) Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells, *Nature Struct. Mol. Biol.*, **17**, 635–640.
164. Mirkovitch, J., Mirault, M.E., Laemmli, U.K. (1984) Organization of the higher order chromatin loop: specific DNA attachment sites on nuclear scaffold, *Cell*, **39**, 223–232.
165. Gerasimova, T.I., Byrd, K., Corces, V.G. (2000). A chromatin insulator determines the nuclear localization of DNA, *Mol. Cell*, **6**, 1025–1035.
166. Iarovaia, O.V., Akopov, S.B., Nikolaev, L.G., Sverdlov, E.D., Razin, S.V. (2005) Induction of transcription within chromosomal DNA loops flanked by MAR elements causes an association of loop DNA with the nuclear matrix, *Nucleic Acids Res.*, **33**, 4157–4163.
167. Shaposhnikov, S.A., Akopov, S.B., Chernov, I.P., Thomsen, P.D., Joergensen, C., Collins, A.R., Frengen, E., Nikolaev, L.G. (2007) A map of nuclear matrix attachment regions within the breast cancer loss-of-heterozygosity region on human chromosome 16q22.1, *Genomics*, **89**, 354–361.
168. Keaton, M.A., Taylor, C.M., Layer, R.M., Dutta, A. (2011) Nuclear Scaffold Attachment Sites within ENCODE Regions Associate with Actively Transcribed Genes, *PLoS ONE*, **6** (3).

169. Jackson, D.A., Dickinson, P., Cook, P.R. (1990) The size of chromatin loops in HeLa cells, *The EMBO Journal*, **9**, 567–571.
170. Libri, V., Miesen, P., van Rij, R.P., Buck, A.H. (2013) Regulation of microRNA biogenesis and turnover by animals and their viruses, *Cell Mol Life Sci.*, **70**, 3525–3544.
171. Hirose, T., Mishima, Y., Tomari, Y. (2014) Elements and machinery of non-coding RNAs: toward their taxonomy, *EMBO Reports*, **15** (5), 489–507.
172. Kim, V.N., Han, J., Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nature Rev. Mol. Cell Biol.*, **10**, 126–139.
173. Monteys, A.M., Spengler, R.M., Wan, J., Tecedor, L., Lennox, K.A., Xing, Y. and Davidson, B.L. (2010) Structure and activity of putative intronic miRNA promoters, *RNA*, **16** (3), 495–505.
174. Westholm, J.O., Lai, E.C. (2011) Mirtrons: microRNA biogenesis via splicing, *Biochimie*, **93**, 1897–1904.
175. Scott, M.S., Ono, M. (2011) From snoRNA to miRNA: Dual function regulatory non-coding RNAs, *Biochimie*, **93**, 1987–1992.
176. Neilsen, C.T., Goodall, G.J., Bracken, C.P. (2012) IsomiRs – the overlooked repertoire in the dynamic microRNAome, *Trends Genet.*, **28**, 544–549.
177. Kawamata, T., Tomari, Y. (2010) Making RISC, *Trends Biochem. Sci.*, **35**, 368–376.
178. Elkayam, E., Kuhn, C-D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J. and Joshua-Tor, L. (2012) The structure of human argonaute-2 in complex with miR-20a, *Cell*, **150**, 100–110.
179. Helwak, A., Kudla, G., Dudnakov, T., Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding, *Cell*, **153**, 654–665.
180. Broderick, J., Salomon, W.E., Ryder, S.P., Aronin, N. and Zamore, P.D. (2011) Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing, *RNA*, **17**, 1858–1869.
181. Salmanidis, M., Pillman, K., Goodall, G., Bracken, C. (2014) Direct transcriptional regulation by nuclear microRNAs, *Int. J. Biochem. Cell Biol.*, **54**, 304–311.
182. Stroynowska-Czerwinska, A., Fiszler, A., Krzyzosiak, W.J. (2014) The panorama of miRNA-mediated mechanisms in mammalian cells, *Cell. Mol. Life Sci.*, **71**, 2253–2270.
183. Okamura, K., Lai, E.C. (2008) Endogenous small interfering RNAs in animals, *Nature Rev. Mol. Cell Biol.*, **9** (9), 673–678.
184. Luteijn, M.J., Ketting, R.F. (2013) PIWI-interacting RNAs: from generation to transgenerational epigenetics, *Nature Rev. Genet.*, **14**, 523–534.
185. Ross, R.J., Weiner, M.M., Lin, H. (2014) PIWI proteins and PIWI-interacting RNAs in the soma, *Nature*, **505**, 353–359.
186. Sayed D., Abdellatif M. (2011) MicroRNAs in development and disease. *Physiol. Rev.*, **91**, 827–887.
187. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, *Genome Res.*, **22**, 1775–1789.
188. Fatica, A., Bozzoni, I. (2014) Long non-coding RNAs: new players in cell differentiation and development, *Nature Rev. Genet.*, **15**, 7–21.

189. Batista, P.J., Chang, H.Y. (2013) Long noncoding RNAs: cellular address codes in development and disease, *Cell*, **152**, 1298–1307.
190. Khalil, A. M. Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A. et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression, *Proc. Natl. Acad. Sci. USA*, **106**, 11667–11672.
191. Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges, *Nature*, **495**, 384–388.
192. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F. and Rajewsky, N. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency, *Nature*, **495**, 333–338.
193. Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., Carter, D.R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease, *RNA*, **17**, 792–798.
194. Yang, W., Wang, X. (2013). Pseudogenes: pseudo or real functional elements? *Journal of Genetics and Genomics*, **40** (4), 171–177.
195. Poliseno, L. (2012). Pseudogenes: newly discovered players in human cancer, *Science Signaling*, **5** (242).
196. Газиев А.И., Шайхаев Г.О. (2010) Ядерно-митохондриальные псевдогены, *Молекулярная биология*, **44** (3), 405–417.
197. Tourmen, Y., Baris, O., Dessen, P., Jacques, C., Malthiery, Y., Reynier, P. (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes, *Genomics*, **80** (1), 71–77.
198. Stoimenov, I., Lagerqvist, A. (2012) The PCNA pseudogenes in the human genome, *BMC Res. Notes.*, **5** (87).
199. Muro, E.M., Mah, N., Andrade-Navarro, M.A. (2011) Functional evidence of post-transcriptional regulation by pseudogenes, *Biochimie*, **93** (11), 1916–1921.
200. Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., Pandolfi, P.P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology, *Nature*, **465** (7301), 1033–1038.
201. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., Hannon, G.J. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes, *Nature*, **453**, 534–538.
202. Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., Sakaki, Y. and Sasaki, H. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes, *Nature*, **453**, 539–543.
203. Hawkins, P.G., Morris, K.V. (2010) Transcriptional regulation of Oct4 by a long noncoding RNA antisense to Oct4-pseudogene 5, *Transcription*, **1** (3), 165–175.
204. Zhang, Z., Gerstein, M., (2004) Large-scale analysis of pseudogenes in the human genome, *Curr. Opin. Genet. Dev.*, **14**, 328–335.
205. Liu, Y.J., Zheng, D., Balasubramanian, S., Carriero, N., Khurana, E., Robilotto, R. and Gerstein, M.B. (2009) Comprehensive analysis

- of the pseudogenes of glycolytic enzymes in vertebrates: The anomalously high number of GAPDH pseudogenes highlights a recent burst of retrotranspositional activity, *BMC Genomics*, **10**, 480.
206. López-Flores, I., Garrido-Ramos, M.A. (2012) The repetitive DNA content of eukaryotic genomes, *Genome Dyn.*, **7**, 1–28.
207. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhou, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H. (2007) A unified classification system for eukaryotic transposable elements. *Nature Rev. Genet.*, **8**, 973–982
208. Kramerov, D.A., Vassetzky, N.S. (2005) Short retroposons in eukaryotic genomes, *Int. Rev. Cytol.*, **247**, 165–221.
209. Orgel, L.E., Crick, F.H.C., Sapienza, C. (1980) Selfish DNA, *Nature*, **288** (5792), 645–646.
210. Georgiev, G.P. (1984) Mobile genetic elements in animal cells and their biological significance, *Eur. J. Biochem.*, **145**, 203–220.
211. De Souza, F.S.J., Franchini, L.F., Rubinstein, M. (2013) Exaptation of Transposable Elements into Novel Cis-Regulatory Elements: Is the Evidence Always Strong? *Mol. Biol. Evol.*, **30** (6), 1239–1251.
212. Abrusán, G. (2006) Somatic transposition in the brain has the potential to influence the biosynthesis of metabolites involved in Parkinson's disease and schizophrenia, *Biology Direct.*, **7**, 41.
213. Upton, K.R., Faulkner, G.J. (2014) Blood from «junk»: the LTR chimeric transcript Pu.2 promotes erythropoiesis, *Mobile DNA*, **5** (15).
214. Smith, Z.D., Chan, M.M., Mikelsen, T.S., Gu, H., Gnirke, A., Regev, A., Meissner, A. (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo, *Nature*, **484**, 339–344.
215. Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E. and Spadafora, C. (2006) Expression of LINE-1 retroposons is essential for murine preimplantation development, *Mol. Reprod. Dev.*, **73**, 279–287.
216. Nandakumar, J., Cech, T.R. (2013) Finding the end: recruitment of telomerase to the telomere, *Nature Rev. Mol. Cell Biol.*, **14** (2), 69–82.
217. Williamson, J.R., Raghuraman, M.K., Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model, *Cell*, **59**, 871–880.
218. Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H. and de Lange, T. (1999) Mammalian telomeres end in a large duplex loop, *Cell*, **97**, 503–514.
219. Sfeir, A., de Lange, T. (2012) Removal of shelterin reveals the telomere end-protection problem, *Science*, **336**, 593–597.
220. Зверева М.Э., Щербакова Д.М., Донцова О.А. (2010) Теломераза: структура, функции и пути регуляции активности. *Успехи биологической химии*, **50**, 155–202.
221. Chen, LY., Lingner, J. (2013) CST for the grand finale of telomere replication, *Nucleus*, **4** (4), 277–282.
222. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends, *Science*, **318**, 798–801.
223. Cifuentes-Rojas, C., Shippen, D.E. (2012) Telomerase regulation, *Mutation Res.*, **730**, 20–27.
224. Eisenberg, D.T., Hayes, M.G., Kuzawa, C.W. (2012) Delayed paternal age of reproduction in humans

- is associated with longer telomeres across two generations of descendants, *Proc. Natl. Acad. Sci. USA*, **109**, 10251–10256.
225. Holohan, B., Wright, W.E., Shay, J.W. (2014) Telomeropathies: An emerging spectrum disorder, *J. Cell Biol.*, **205** (3), 289–299.
226. Plohl, M., Meštrović, N., Mravinac, B. (2014) Centromere identity from the DNA point of view, *Chromosoma*, **123**, 313–325.
227. Aldrup-MacDonald, M.E., Sullivan, B.A. (2014) The past, present, and future of human centromere genomics. *Genes*, **5**, 33–50.
228. Gent, J.I., Dawe, R.K. (2012) RNA as a structural and regulatory component of the centromere, *Annu. Rev. Genet.*, **46**, 443–453.
229. Probst, A.V., Okamoto, I., Casanova, M., El Marjou, F., Le Vaccon, P., Almouzni, G. (2010) A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development, *Dev. Cell*, **19**, 625–638.
230. Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., Rivera, M.N., Bardeesy, N., Maheswaran, S. and Haber D.A. (2011) Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers, *Science*, **331**, 593–596.
231. Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W, Earle, E., Choo, K.H. and Wong, L.H. (2012) Active transcription and essential role of RNA polymerase II at the centromere during mitosis, *Proc. Natl. Acad. Sci. USA*, **109**, 1979–1984.
232. Романов, Г.А., Суховеров В.С., Ванюшин Б.Ф. (2015) Эпигенетический мутагенез как программа возрастной дисфункции белков и старения. *Онтогенез*, (в печати).