

## ГИПОТЕЗА О МЕХАНИЗМЕ ИНИЦИАЦИИ МАЛЫМИ РНК МЕТИЛИРОВАНИЯ ДНК DE NOVO И АЛЛЕЛЬНОГО ИСКЛЮЧЕНИЯ

© В. А. Галицкий

*Институт биохимии им. А. В. Палладина НАН Украины, Киев;  
электронный адрес: volha@biochem.kiev.ua*

Нами обнаружено, что встречаемость динуклеотидов 5'-CG-3' и тринуклеотидов 5'-CNG-3', содержащихся в неопубликованных последовательностях малых интерферирующих РНК и микроРНК, намного превышает их встречаемость в случайной последовательности ДНК. Данное обстоятельство указывает на то, что динуклеотид 5'-CG-3' и тринуклеотид 5'-CNG-3' в составе малых РНК должны играть существенную биологическую роль. Мы предполагаем, что малые РНК, содержащие указанные ди- и тринуклеотиды, участвуют в нанесении на хроматин меток эпигенетической информации, осуществляя высокоспецифический поиск подлежащих репрессии последовательностей ДНК и инициируя метилирование de novo комплементарных сайтов 5'-CG-3' и 5'-CNG-3' в пределах найденных малыми РНК участков ДНК. Если узнаваемый малой РНК мотив содержится в нескольких генах, они могут быть инактивированы одновременно. Аллельное исключение возникает, по нашему мнению, вследствие инициации малой РНК, предшественник которой транскрибируется с антипараллельной цепи аллелей, и метилирования de novo ДНК всех имеющихся в клетке аллелей, кроме одного. Инактивируются те аллели, антипараллельная цепь которых менее активно считывается РНК-полимеразой, освобождающей при транскрипции, как мы полагаем, ДНК от присоединившейся малой РНК. Количества же малой РНК, транскрибируемой только с одного аллеля, недостаточно для преодоления уровня, после которого инициируется процесс репрессии этого аллеля de novo.

Ключевые слова: эпигенетическая информация, аллельное исключение, метилирование ДНК, малые РНК, хроматин.

Принятые сокращения: m<sup>5</sup>C — 5-метилцитозин, N — нуклеотид, DNMT1 — DNA-methyltransferase 1, DNMT3a — DNA-methyltransferase 3a, DNMT3b — DNA-methyltransferase 3b, HDAC — histone deacetylase, HMT — histone methyltransferase, PcG — Polycomb group protein, dsRNA — double stranded RNA, RISC — RNA interfering silencing complex, siRNA — small interfering RNA, RdRP — RNA-directed RNA polymerase, RITS — RNA-induced transcriptional silencing complex.

Материальные носители эпигенетической информации. В процессе индивидуального развития организма экспрессия генов дифференцирующихся клеток подвергается изменениям, которые сохраняются в ряду клеточных поколений, но не связаны, как правило, с какими-либо нарушениями нативной нуклеотидной последовательности ДНК. Такие изменения называются эпигенетическими. Материальными носителями эпигенетической информации являются, с одной стороны, ковалентная модификация оснований ДНК, а именно метилирование цитозина в положении С-5 пиримидинового кольца с образованием 5-метилцитозина (m<sup>5</sup>C), а с другой — ковалентная модификация (ацетилирование, метилирование, фосфорилирование, убиквитилирование и ADP-рибозилирование) гистонов по определенным аминокислотным остаткам, чаще всего лизиновым (Habu et al., 2001; Martienssen, Colot, 2001; Moazed, 2001; Vaucheret, Fagard, 2001; Richards, Elgin, 2002).

Цитозин, как известно, метилируется ДНК-метилтрансферазами, главным образом в том случае, когда рядом с ним находится гуанин, образуя динуклеотид 5'-CG-3' или иногда тринуклеотид 5'-CNG-3'. Не все воз-

можные остатки цитозина в ДНК одновременно находятся в метилированном состоянии — рисунок метилирования/деметилирования ДНК является для клетки индивидуальным и несет эпигенетическую информацию, определяющую уровни экспрессии клеточных генов. Аналогично рисунок модификации гистонов нуклеосом также содержит подобную информацию; часто информация, закодированная посредством метилирования оснований ДНК, и информация, закодированная при помощи модификаций гистонов, дублируют друг друга (Stancheva, 2005).

Уровень, при котором проявляются эффекты, опосредованные метилированием ДНК, весьма низок — существенное подавление экспрессии гена наблюдается, когда 1 m<sup>5</sup>C приходится на 300 пар нуклеотидов (Hsieh, 1994), а репрессия клеточных промоторов наступает при достижении плотности 1 m<sup>5</sup>C на 100 пар нуклеотидов (Bird, 1995).

Сохранение эпигенетических меток требует функционирования особых механизмов. Так как ДНК-полимеразы включают в синтезируемую цепь ДНК лишь немодифицированные основания, после репликации в молекуле ДНК остается метилированной лишь одна из цепей (полу-

метилированное состояние). Рисунок метилирования ДНК восстанавливается вслед за ее репликацией благодаря способности ДНК-метилтрансферазы DNMT1 метилировать цитозин в тех динуклеотидах 5'-CG-3' новообразованной цепи, гуаниновый остаток которых спарен с m<sup>5</sup>C другой цепи (Araujo et al., 1998; Hermann et al., 2004). Указанный процесс носит название поддерживающего метилирования; благодаря ему эпигенетическая информация клетки сохраняется и передается дочерним клеткам. Уровень точности поддерживающего метилирования для большинства сайтов 5'-CG-3' составляет не менее 99 % (Riggs, Xiong, 2004).

Эпигенетические метки на ДНК, вероятно, используются клеточными механизмами также для коррекции и восстановления соответствующих модификаций связанных с нею гистонов. Это необходимо, если гистоны при репликации ДНК распределяются консервативно, оставаясь связанными с ДНК, содержащей ведущую цепь, а ДНК, содержащая отстающую цепь, связывается с новыми гистонами, лишаясь, таким образом, эпигенетической информации, заключенной в модификациях гистонов. Впрочем, недавно выдвинута гипотеза, согласно которой при движении репликационной вилки каждый нуклеосомный гистоновый октамер разделяется, предоставляя обоим дочерним цепям ДНК по димеру, состоящему из гистона H3 и гистона H4 (Tagami et al., 2004). Этот димер быстро превращается в (H3—H4)<sub>2</sub>-тетрамер, присоединяя новые, еще не модифицированные молекулы гистонов H3 и H4. Соответственно, нанесенные на гистоны эпигенетические метки могут наследоваться полуконсервативно и после репликации ДНК воссоздаваться при помощи особых ферментов на новых гистонах, включенных в состав нуклеосом.

Метилирование новых сайтов ДНК, т. е. задание и изменение эпигенетической информации, осуществляется в ходе другого процесса — метилирования *de novo*, за который отвечают ДНК-метилтрансферазы DNMT3а и DNMT3б, не имеющие в отличие от DNMT1 особого сродства к полуметилированной ДНК (Okano et al., 1998). С этим процессом тесно связана и модификация *de novo* гистонов — их деацетилирование и метилирование, также способствующие подавлению экспрессии генов. Механизм метилирования ДНК *de novo* в настоящее время интенсивно изучается; в настоящей работе будет рассмотрено его возможное функционирование с учетом выявленных нами некоторых особенностей нуклеотидных последовательностей малых РНК.

Эпигенетическая регуляция активности генов. Транскрипционно активные гены, как правило, находятся в составе эухроматина, неактивные — в составе гетерохроматина. Считается, что в первом случае к ним легко могут получить доступ транскрипционные факторы и РНК-полимераза, во втором — такой доступ затруднен. В «открытом» (т. е. активном) хроматине в гистоне H3 ацетилированы остаток лизина-9 (H3K9-Ас), лизина-14 (H3K14-Ас) и лизина-18 (H3K18-Ас), а также в гистоне H4 остаток лизина-5 (H4K5-Ас) и лизина-8 (H4K8-Ас). Деацетилирование этих остатков аминокислот, а также метилирование в гистоне H3 остатков лизина-9 (H3K9-Ме) и лизина-27 (H3K27-Ме) имеют место в «закрытом» конденсированном хроматине и связаны с подавлением экспрессии генов; в случае стабильной репрессии остаток лизина-9 может подвергаться триметилированию (Felsenfeld, Groudine, 2003; Peters et al., 2003). Известно также, что белки-репрессоры привлекают к своим генам-мишеням гистон-деацетилазы (HDAC), а также ги-

стон-метилтрансферазы (HMT) и ДНК-метилтрансферазы, инактивируя гены (Di Croce et al., 2002; Шиф, 2005). Метилированные мотивы гистонов высокоизбирательно распознаются хромодоменами белков PcG и HP1, в результате чего нуклеосомы вместе со связанными с ними участками ДНК конденсируются в гетерохроматин (Lachner et al., 2001; Чуриков, 2005).

Модифицированные по остатку K9 молекулы гистонов H3 узнаются также белковыми комплексами, содержащими ДНК-метилтрансферазы, в результате чего ДНК, находящаяся в составе модифицированных нуклеосом, метилируется. В свою очередь метил-ДНК-связывающие белки, распознавая метилированные участки ДНК, способны присоединять гистондеацетилазы, тем самым вызывая удаление из гистонов меток активного хроматина (Jones et al., 1998; Nan et al., 1998; Wade et al., 1999). Таким образом, метилирование ДНК, определяющее нетранскрибируемый статус генов, способно вызывать модификации гистонов, характерные для неактивного хроматина, и наоборот. Этим устойчиво поддерживается неактивное состояние хроматина и обеспечивается восстановление информации в случае полной или частичной потери эпигенетических меток одной из сторон данного информационного взаимодействия — ДНК или гистонами. Из этого также вытекает, что для устойчивой активации экспрессии гена, находящегося в составе неактивного хроматина, необходимо удаление эпигенетических меток как с ДНК, так и с гистонов, тогда как для сайленсинга должно быть достаточно инициации соответствующей ковалентной модификации лишь одной из упомянутых сторон.

Исследования последних лет показали, что подавление экспрессии гена может быть вызвано не только репрессорами белковой природы, и не всегда на уровне транскрипции. Двухцепочечная РНК (dsRNA) также способна вызывать сайленсинг гена. В клетке она распознается и расщепляется эндорибонуклеазой III Dicer на фрагменты длиной около 22 нуклеотидов (Bernstein et al., 2001), которые попадают в комплексы RISC (RNA interfering silencing complex) (Zamore, 2002; Tijsterman, Plasterk, 2004). Там dsRNA расплетается при помощи белка Argonaute, и одна из цепей dsRNA остается в дальнейшем связанной с этим белком, выполняющим, очевидно, относительно нее ту же функцию, что и шапероны относительно отдельных белков, — удержания в развернутой конформации (что должно способствовать гибридизации с нуклеотидной последовательностью-мишенью). Упомянутые одноцепочечные олигонуклеотиды (аптамеры) являются малыми интерферирующими РНК (siRNA). Связываясь с комплементарными участками молекул РНК, siRNA иницируют их расщепление, в результате чего экспрессия генов прекращается на посттранскрипционной стадии. Также известен класс клеточных шпилькообразующих микроРНК (miRNA), синтезируемых в виде РНК-предшественников (pri-miRNA). Последние образуют несовершенные шпильки, которые сначала преобразуются ядерной рибонуклеазой III Drosha в двухцепочечные pre-miRNA длиной около 60 п. н. (Lee et al., 2003). В цитоплазму pre-miRNA попадают благодаря активному транспорту при участии экспортина 5 (Lund et al., 2004) и в дальнейшем процессируются, подобно siRNA, рибонуклеазой Dicer (He, Hannon, 2004). Затем зрелые miRNA подвергаются расплетанию в комплексе RISC и участвуют в регуляции экспрессии генов эукариотических клеток, в том числе в течение процесса дифференцировки (Hutvagner, Zamore, 2002; Schwarz, Zamore, 2002).

Кроме упомянутого посттранскрипционного механизма сайленсинга siRNA могут инициировать и подавление самой транскрипции генов. Показано, что воздействие искусственной siRNA, направленной против 5'-CG-3'-содержащих промоторов, вызывает в клетках человека метилирование ДНК и остатка K9 гистона H3 de novo, зависящее от одновременного присутствия ДНК-метилтрансфераз DNMT1 и DNMT3b, но не DNMT2 (Kawasaki, Taira, 2004). В настоящее время предложены два варианта механизма РНК-опосредованного транскрипционного сайленсинга (Verdel et al., 2004). Считается, что последний иницируется комплексом RITS (RNA-induced transcriptional silencing complex), состоящим у *Schizosaccharomyces pombe* из удерживающего молекулу siRNA белка Ago1 из семейства Argonaute, хромодомен-содержащего белка Chp1 и белка Tas3. siRNA узнает комплементарный мотив или непосредственно в одной из цепей молекулы ДНК, временно расплетенной в ходе репликации, или же в синтезируемом первичном РНК-транскрипте, еще до его отделения от РНК-полимеразы (Verdel et al., 2004). Комплекс RITS, связавший последовательность-мишень, рекрутирует РНК-зависимую РНК-полимеразу (RdRP), которая достраивает комплементарную цепочку на распознанном молекулой siRNA транскрипте (образовывающаяся в результате этого dsRNA служит субстратом для фермента Dicer). Этот факт, впрочем, не указывает на непосредственное взаимодействие siRNA с первичным транскриптом, так как RdRP не зависит от праймера (Motamedi et al., 2004), кроме того, RdRP присутствует у растений и дрожжей, но не у млекопитающих или дрозофилы. Высказывалось предположение о том, что функционирование RITS должно облегчаться вследствие метилирования гистона H3 по остатку K9 гистон-метилтрансферазой Clr4, хотя механизм привлечения последней в хроматин не был ясен (Verdel, 2005).

Двухцепочечная структура ДНК временно нарушается в ходе таких процессов, как репликация, транскрипция, репарация и рекомбинация, поэтому молекулы малой РНК теоретически могут непосредственно взаимодействовать с комплементарной им последовательностью ДНК. Исходя из данных соображений мы решили изучить, могут ли быть мишенями siRNA и miRNA непосредственно участки ДНК, содержащие потенциальные сайты метилирования. Так как взаимодействие РНК с ДНК комплементарно, данная задача сводится к поиску сайтов, комплементарных сайтам метилирования, т.е. динуклеотидов 5'-CG-3' и тринуклеотидов 5'-CNG-3', в последовательностях самих siRNA и miRNA. Также заметим, что в ходе эволюции большинство остатков цитозина в динуклеотидах 5'-CG-3' подверглось транзиции из-за замены цитозина на тимин вследствие дезаминирования m<sup>5</sup>C. В настоящее время, например, в геноме человека динуклеотид 5'-CG-3' встречается с частотой 1% (Scarano et al., 1967; Jabbari, Bernardi, 2004), что в 4 с лишним раза меньше частоты его ожидаемой встречаемости в случайной последовательности ДНК (так как геном человека содержит 42% пар G-C, частота встречаемости динуклеотида 5'-CG-3' должна была бы равняться  $0.21 \cdot 0.21 = 0.0441 = 4.41\%$ ).

## Материал и методика

Последовательности siRNA человека, мыши и крысы (всего 599 последовательностей) были извлечены из базы данных «siRNA Database and Resources for

RNA Interference Studies», находящейся по адресу: <http://www.rnainterference.org/Sequences.html>. Последовательности зрелых miRNA человека, мыши и крысы (всего 1083 последовательности) были извлечены из базы данных miRBase, находящейся по адресу: <http://microrna.sanger.ac.uk/> (Griffiths-Jones et al., 2006). Заметим, что в последней базе данных содержится также информация о последовательностях miRNA многих других организмов, в том числе кабана, овцы, быка, опоссума, собаки, макаки-резуса, гориллы, орангутана, дрозофилы, пчелы, шелкопряда, планарии, риса, кукурузы, люцерны, сахарного тростника, сои, *Xenopus laevis*, *X. tropicalis*, *Danio rerio*, *Tetraodon nigroviridis*, *Fugu rubripes*, *Gallus gallus*, *Caenorhabditis elegans*, *Caenorhabditis briggsae*, *Anopheles gambiae*, *Arabidopsis thaliana*, *Populus trichocarpa*, *Sorghum bicolor* и *Physcomitrella patens* (всего 4167 последовательностей miRNA). Для поиска сайтов 5'-CG-3', 5'-CNG-3' и 5'-GC-3' в нуклеотидных последовательностях использовали программные пакеты Discovery Studio Gene, версия 1.5, производство Accelrys Inc., США; Vector NT1 Advance 10, производство Invitrogen Corp., США; LaserGene 6, производство DNASTAR Inc., США.

## Результаты

Распространение сайтов 5'-CG-3' и 5'-CNG-3' в молекулах siRNA и miRNA. Динуклеотид 5'-CG-3' в составе последовательностей siRNA, опубликованных в базе данных «siRNA Database and Resources for RNA Interference Studies», встретился 361 раз; в среднем на каждые 100 молекул siRNA приходится 60.3 сайта 5'-CG-3' (табл. 1). В составе исследованных последовательностей зрелых miRNA, опубликованных в базе данных miRBase, динуклеотид 5'-CG-3' встретился 539 раз; таким образом, на каждые 100 молекул miRNA приходится 49.8 сайта 5'-CG-3' (табл. 1). Но, как показали наши исследования базы данных, последовательности около 10% siRNA и miRNA содержат два динуклеотида 5'-CG-3' или больше, поэтому последние должны встречаться в составе несколько меньшего количества малых РНК.

Так как в последовательности из  $n$  нуклеотидов содержится  $n - 1$  возможных динуклеотидов, средняя частота встречаемости динуклеотида 5'-CG-3' составила 2.85% в последовательностях siRNA и 2.39% в последовательностях зрелых miRNA. Видовые показатели частоты встречаемости динуклеотидов 5'-CG-3', а также интересующих нас сайтов 5'-CNG-3' и 5'-GC-3' (см. далее) в siRNA и miRNA приведены в табл. 2. Так как частота встречаемости сайта 5'-CG-3' в геноме составляет у человека и мыши 1%, у крысы 1.2% (Jabbari, Bernardi, 2004), выглядит весьма интригующим то, что в последовательностях siRNA и miRNA у данных организмов этот уровень является намного большим.

Тринуклеотид 5'-CNG-3' (буквой N обозначен любой из нуклеотидов) был обнаружен нами в исследованных последовательностях siRNA 734 раза; таким образом, на каждые 100 молекул siRNA приходится 122.5 триплета 5'-CNG-3'. В составе подвергшихся изучению последовательностей зрелых miRNA тринуклеотид 5'-CNG-3' встретился нам 1364 раза; таким образом, на каждые 100 молекул miRNA в среднем приходится 125.9 триплета 5'-CNG-3' (табл. 1). При подсчете количества тринуклеотидов 5'-CNG-3' мы принимали, что тетрануклеотид

Таблица 1

## Количество исследованных молекул siRNA и miRNA и обнаруженных сайтов 5'-CG-3', 5'-CNG-3' и 5'-GC-3'

Вид	Количество последовательностей		Суммарное количество нуклеотидов в последовательностях		Ожидаемая частота встречаемости сайтов в случайной последовательности					
	siRNA	miRNA	siRNA	miRNA	5'-CG-3'		5'-CNG-3'		5'-GC-3'	
					siRNA	miRNA	siRNA	miRNA	siRNA	miRNA
<i>Homo sapiens</i>	468	470	10 323	10 292	286 (61.1)	237 (50.4)	563 (120.3)	561 (119.4)	538 (115.0)	560 (119.1)
<i>Mus musculus</i>	93	375	2030	8174	56 (60.2)	192 (51.2)	119 (128.0)	492 (131.2)	119 (128.0)	438 (116.8)
<i>Rattus norvegicus</i>	38	238	895	5182	19 (50.0)	110 (46.2)	52 (136.8)	311 (130.7)	37 (97.4)	261 (109.7)
Всего	599	1038	13 248	23 648	361 (60.3)	539 (49.8)	734 (122.5)	1364 (125.9)	694 (115.9)	1259 (116.3)

Примечание. В скобках указано количество сайтов, приходящееся на 100 последовательностей.

Таблица 2

## Частота встречаемости сайтов 5'-CG-3', 5'-CNG-3' и 5'-GC-3' в исследованных молекулах siRNA и miRNA, %

Вид	Содержание GC в геноме, % <sup>a</sup>	Ожидаемая частота встречаемости сайтов 5'-CG-3', 5'-GC-3' или 5'-CNG-3' в случайной последовательности	Сайт 5'-CG-3'		Сайт 5'-CNG-3'				Сайт 5'-GC-3'		
			среднее по геному <sup>a</sup>	siRNA	miRNA	siRNA	miRNA	среднее по геному <sup>a</sup>	siRNA	miRNA	
											miRNA
<i>Homo sapiens</i>	42.00	4.41	1.00	2.90	2.41	6.20	6.13	4.30	5.46	5.70	
<i>Mus musculus</i>	42.20	4.45	1.00	2.89	2.46	6.67	6.82	3.90	6.14	5.62	
<i>Rattus norvegicus</i>	43.90	4.82	1.20	2.22	2.22	6.47	6.71	4.10	4.32	5.28	
В среднем	—	—	—	2.85	2.39	6.29	6.49	—	5.49	5.58	

<sup>a</sup> По: Jabbari, Bernardi, 2004.

Таблица 3

## Количество исследованных молекул siRNA и miRNA, полностью лишенных сайтов 5'-CG-3' и 5'-CNG-3'

Вид	siRNA		miRNA	
	абсолютное значение	доля от общего количества, %	абсолютное значение	доля от общего количества, %
<i>Homo sapiens</i>	75	16.03	101	21.49
<i>Mus musculus</i>	14	15.05	75	20.00
<i>Rattus norvegicus</i>	5	13.16	51	21.43
Всего	94	15.69	227	20.96

Примечание. Общее количество исследованных последовательностей указано в табл. 1.

5'-CCGG-3', который может читаться как две разные последовательности 5'-CNG-3' (5'-CCG-3' и 5'-CGG-3'), соответствует одному триплету 5'-CNG-3'.

Частоту встречаемости сайтов 5'-CNG-3' рассчитывали исходя из того соображения, что в последовательности из  $n$  нуклеотидов содержится  $n - 2$  возможных тринуклеотидов (в этом случае тетрануклеотид 5'-CCGG-3' мы считали содержащим два возможных тринуклеотида 5'-CNG-3'). Средняя частота встречаемости сайта 5'-CNG-3' составила 6.29 % в последовательностях siRNA и 6.49 % в последовательностях зрелых miRNA, видо-

вые показатели также приведены в табл. 2. Заметим, что в случайной последовательности ДНК ожидаемая частота встречаемости тринуклеотида 5'-CNG-3' должна быть той же, что и динуклеотида 5'-CG-3' (см. выше), в частности у человека  $0.21 \cdot (0.21 + 0.21 + 0.29 + 0.29) \cdot 0.21 = 0.21 \cdot 1 \cdot 0.21 = 0.0441 = 4.41$  %, где 0.21 — доля цитозина, а также гуанина, 0.29 — доля аденина, а также тимина в геноме человека. Таким образом, и тринуклеотид 5'-CNG-3' в последовательностях siRNA и miRNA обнаруживали значительно чаще, чем он должен встречаться в случайной последовательности.

Заметим также, что если в одной молекуле малой РНК обнаруживали несколько триплетов 5'-CNG-3', то нуклеотид N у них часто оказывался одним и тем же. Чем это может быть вызвано, пока неизвестно.

В упомянутых базах данных нами были найдены всего 94 (15.69 %) последовательности siRNA и 227 (20.96 %) последовательностей зрелых miRNA, вообще не содержащих динуклеотидов 5'-CG-3' или тринуклеотидов 5'-CNG-3' (табл. 3). В то же время примерно в половине таких последовательностей обнаруживали один или несколько сайтов 5'-GC-3'.

Распространение динуклеотидов 5'-GC-3' в молекулах siRNA и miRNA. Частота встречаемости динуклеотида 5'-GC-3' привлекла наше внимание, поскольку известно, что молекулы нуклеиновых кислот физически способны к образованию не только антипараллельных, но и параллельных дуплексов; есть указания на то, что у прокариот сайленсинг вызывает siRNA, комплементарная РНК-мишень не в антипараллельном, а в параллельном направлении (Tchurikov et al., 2000). Тем не менее частота встречаемости динуклеотида 5'-GC-3' составила 5.58 % в последовательностях miRNA и 5.49 % в последовательностях siRNA (табл. 2). Данный уровень несколько превышает среднюю частоту встречаемости динуклеотида 5'-GC-3' в геноме, колеблющуюся у позвоночных организмов, малые РНК которых подвергались исследованию, в пределах от 3.9 (мышь) до 4.3 (человек) % (Jabbari, Bernardi, 2004), но это превышение выражено не столь резко, как в случае частоты встречаемости сайтов 5'-CNG-3' и особенно 5'-GC-3'. Поэтому мы не будем рассматривать возможности сайленсинга, связанные с динуклеотидами 5'-GC-3'.

## Обсуждение

Гипотеза о механизме метилирования ДНК *de novo*, иницируемого малыми РНК. Полученные результаты свидетельствуют о том, что динуклеотид 5'-CG-3' в siRNA и miRNA обнаруживается значительно чаще, чем он должен встречаться в случайной последовательности, и имеется примерно у половины малых РНК. У подавляющего большинства малых РНК имеется и тринуклеотид 5'-CNG-3', который также обнаруживается чаще, чем должен встречаться в случайной последовательности. В то же время лишь пятая часть miRNA и меньше шестой части siRNA полностью лишены сайтов 5'-GC-3' или 5'-CNG-3'. Данное обстоятельство указывает на то, что динуклеотиды 5'-GC-3' и тринуклеотиды 5'-CNG-3' в малых РНК должны играть некую существенную биологическую роль. Это позволяет нам ввести в предложенную гипотезу о сайленсинге, опосредованном siRNA-ДНК-взаимодействием (Verdel et al., 2004), другой молекулярный механизм сайленсинга.

По нашему мнению, комплексы малых РНК и белка Argonaute сканируют нуклеотидную последовательность не первичных транскриптов, а именно цепей молекулы ДНК по мере ее расплетания РНК-полимеразой II при транскрипции (последним обстоятельством объясняется зависимость транскрипционного сайленсинга от функционирования РНК-полимеразы II). Скорее всего, малая РНК сканирует цепь ДНК, не являющуюся матрицей для транскрипции (поэтому малые РНК комплементарны также последовательности транскрипта и могут связываться с ним в ходе обычного посттранскрипционного сай-

ленсинга). Обнаружив комплементарный участок ДНК, siRNA или miRNA связывается с ним и, как мы предполагаем, привлекает клеточные ДНК-метилтрансферазы, которые метилируют в ДНК цитозин динуклеотидов 5'-GC-3' и тринуклеотидов 5'-CNG-3', оказавшихся связанными с аналогичными сайтами в составе малой РНК (рис. 1). Комплементарная цепь ДНК метилируется, возможно, поддерживающей метилазой DNMT1 уже после вытеснения малой РНК и восстановления двухцепочечной структуры ДНК. Возможно также, что в ходе этих событий привлекаются гистон-деацетилазы и гистон-метилазы, соответственно удаляющие с гистонов метки активного хроматина и наносящие на них метки нетранскрибируемого хроматина, но вероятнее, что белки, модифицирующие гистоны, рекрутируются опосредованно, распознав метильные метки на ДНК, нанесенные в участке ее гибридизации с siRNA или miRNA (рис. 1).

Заметим, что функционирование данного механизма должно непосредственно приводить к метилированию лишь тех сайтов, ДНК которых подвергается расплетанию РНК-полимеразой II, т. е. расположенных практически в пределах транскрибируемого участка гена. Как же метилируются регуляторные элементы, находящиеся в нетранскрибируемых участках гена? Можно предположить, что эпигенетические метки способны распространяться на весь транскриптон, в частности из-за того, что процессы деацетилирования и метилирования гистонов могут захватывать не только те нуклеосомы, которые связаны с первым метилированным *de novo* сайтом ДНК, но и соседствующие с ним (рис. 1). Тогда ДНК-метилтрансферазы, привлеченные к метилированным гистонам, метилируют связанную с ними ДНК, что снова ведет к модификации близлежащих гистонов и т. д.

В результате этих событий ген, распознанный малой РНК, перенаправляется в гетерохроматин и оказывается стойко инактивированным. Кроме того, если узнаваемый малой РНК мотив содержится в нескольких генах, все они могут быть выключены одновременно. В связи с этим мы допускаем, что генные модули (группы генов, предположительно являющиеся элементарными единицами генной сети клеточной дифференцировки) содержат гены специфических miRNA, активирующиеся в определенные моменты времени с целью стойкого наследуемого прекращения деятельности других генных модулей, выполняющих свои задания в процессе клеточной дифференцировки или отвечающих за ее альтернативные направления.

Формирование комплекса RITS. Мы предполагаем также, что комплекс RITS, открытый у некоторых организмов (Verdel et al., 2004; Verdel, 2005), формируется уже после указанных событий, и его предназначение — иницировать уничтожение РНК-транскрипта гена, подвергнувшегося распознаванию молекулой siRNA. Также мы считаем, что последовательность формирования комплекса RITS несколько иная, чем предложили открывшие его исследователи. Заметим, что в RITS входит белок Chp1, содержащий хромодомен, а последний, как известно, связывается с метилированным по остатку K9 гистонам H3 (Sadaie et al., 2004); также установлено, что H3K9-метилирование сохраняется в отсутствие Chp1 или механизмов РНК-интерференции (Sadaie et al., 2004). Это может означать, что лизин-9 в данном гистоне должен быть модифицирован прежде, чем начнется формирование комплекса RITS, а не после его присоединения, как предполагали Бюлер и соавторы (Bühler et al., 2006).

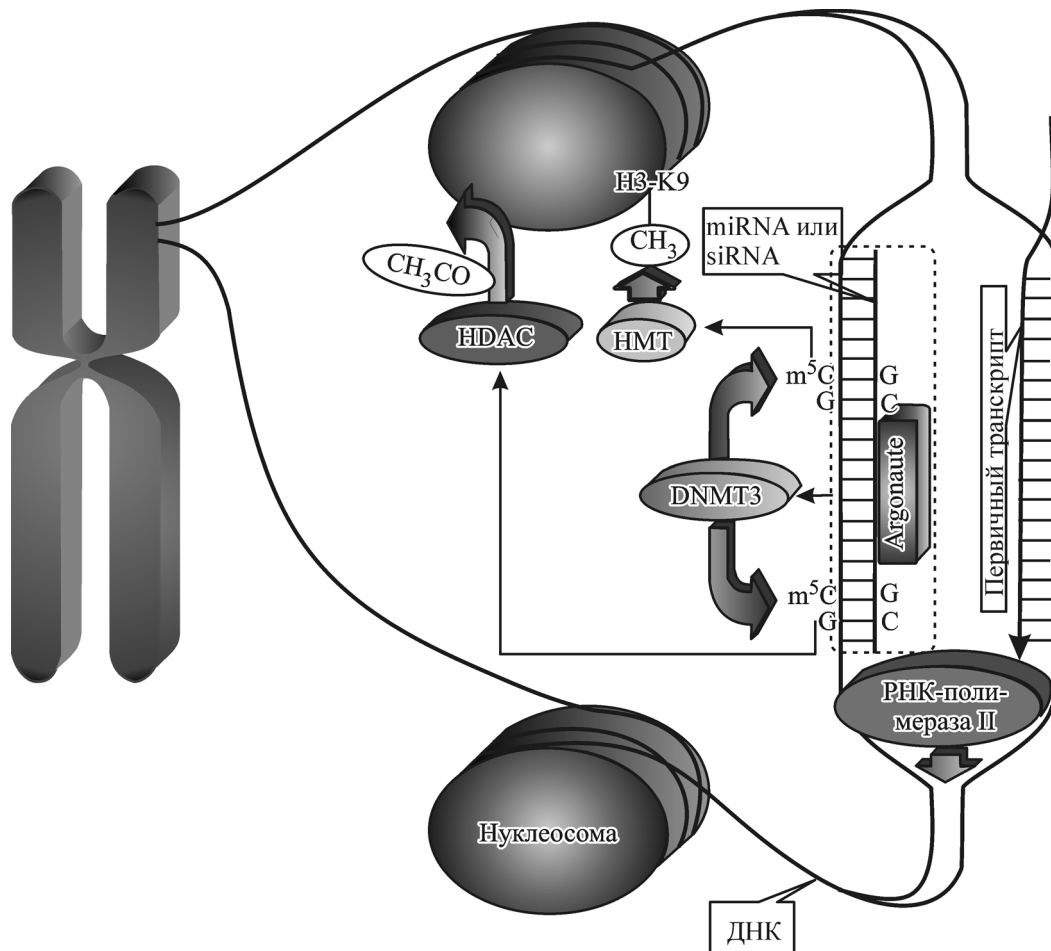


Рис. 1. Инициация малыми РНК метилирования ДНК de novo.

$m^5C$  — 5-метилцитозин, С — цитозин, G — гуанин, DNMT3 — ДНК-метилтрансфераза 3, HDAC — гистон-деацетилаза, HMT — гистон-метилтрансфераза, H3-K9 — остаток лизина-9 в молекуле гистона H3,  $CH_3$  — метильная группа,  $CH_3CO$  — ацетильная группа, miRNA — микроРНК, siRNA — малая интерферирующая РНК.

По нашему мнению, первым к белку Argonaute, удерживающему молекулу siRNA, связавшуюся с комплементарной последовательностью ДНК, присоединяется — уже после метилирования ДНК и гистона H3 — содержащий хромодомен белок Chp1, а к ним — белок Tas3; так как он способен связывать РНК (Bühler et al., 2006), мы считаем, что его задача — найти и присоединить транскрипт, в процессе синтеза которого произошло узнавание последовательности гена молекулой siRNA. Формирование комплекса RITS на этом завершается, после чего он присоединяет RdRP и нацеливает ее на образование dsRNA на первичном транскрипте, все еще связанном с РНК-полимеразой (рис. 2). Таким образом, для инициации деградации первичного транскрипта (и остальной РНК, ранее считанной с этого гена) посредством РНК-интерференции должно быть необходимо, во-первых, распознавание молекулой siRNA комплементарного мотива ДНК и, во-вторых, взаимодействие хромодомена белка Chp1 с метильной меткой на остатке K9 гистона H3; этим может предотвращаться несанкционированная деградация транскриптов. Для того чтобы выяснить, действительно ли имеет место указанная последовательность событий, необходимы исследования связывания белков, которые в них участвуют, друг с другом, с siRNA, РНК и ДНК в процессе транскрипционного сайленсинга.

Можно также предположить, что молекулы siRNA и miRNA, вообще не содержащие динуклеотидов 5'-CG-3' или тринуклеотидов 5'-CNG-3', принимают участие лишь в обычном, посттранскрипционном, сайленсинге.

Аллельное исключение и эффект дозы гена. Объяснение механизма аллельного исключения и других феноменов, где проявляет себя эффект дозы гена, представляет собой весьма непростую задачу. Далеко не все гены подвержены аллельному исключению, но если оно для какого-либо гена характерно, то будет неизменно устанавливаться при самом широком диапазоне колебаний уровня его экспрессии и регуляторных влияний. Независимо от количества аллелей в результате аллельного исключения в клетке останется активным не больше и не меньше одного из них.

Тем не менее если учесть, что транскрипции по крайней мере иногда подвергается не только цепь ДНК, содержащая ген, кодирующий белок, подверженный аллельному исключению, но и антипараллельная цепь, и предположить, что ее транскрипт представляет собой предшественник miRNA, зрелая форма которой впоследствии может с этой цепью связаться, то механизм аллельного исключения оказывается на удивление простым. Будем считать, что количества малой РНК, считываемой только с одного аллеля, в любом случае недостаточно для

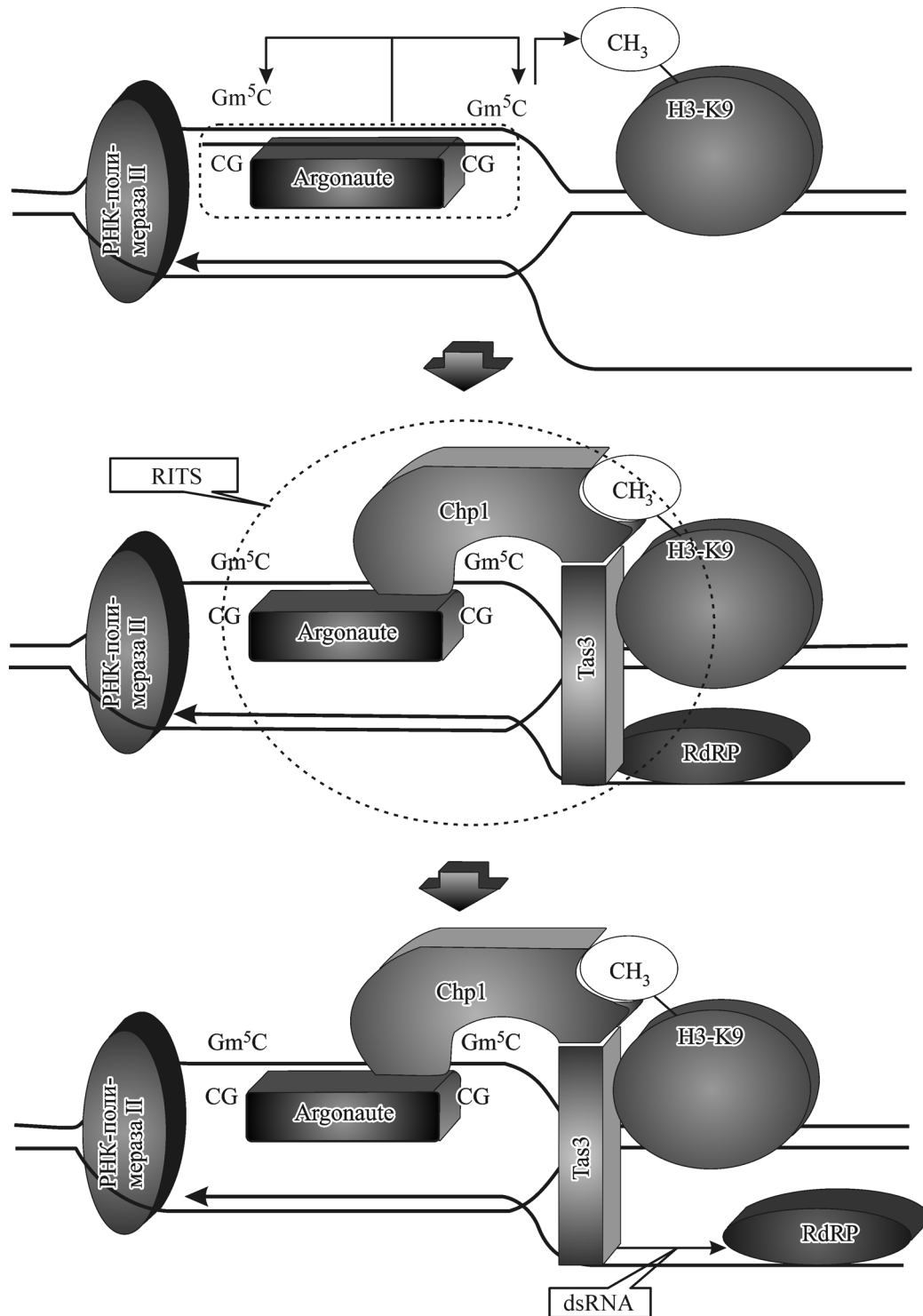


Рис. 2. Формирование комплекса RITS и его нацеливание на транскрипт.

m<sup>5</sup>C — 5-метилцитозин, dsRNA — двухцепочечная РНК, RdRP — РНК-зависимая РНК-полимераза; остальные обозначения те же, что и на рис. 1.

преодоления того порога, после которого инициируется процесс метилирования и репрессии этого аллеля *de novo*, предложенный выше. Существование данного порогового уровня можно объяснить тем, что РНК-полимераза, транскрибирующая антипараллельную цепь ДНК, отделяет или расщепляет связавшиеся с последней (напомним, при считывании белок-кодирующего гена) молекулы miRNA (возможно, что это отделение происходит с определенной задержкой, так как гибрид miRNA—ДНК не-

сколько устойчивее, чем обычная двухцепочечная ДНК). Тогда малые РНК просто не успевают инициировать и(или) поддерживать эпигенетические изменения. Аллельное же исключение возникает вследствие того, что рано или поздно уровень miRNA в клетке, где по каким-то причинам функционируют два аллеля гена или больше, преодолевает указанный порог. При этом РНК-полимераза, транскрибирующая антипараллельную цепь одного из аллелей, не успевает сбросить присоединившуюся miRNA

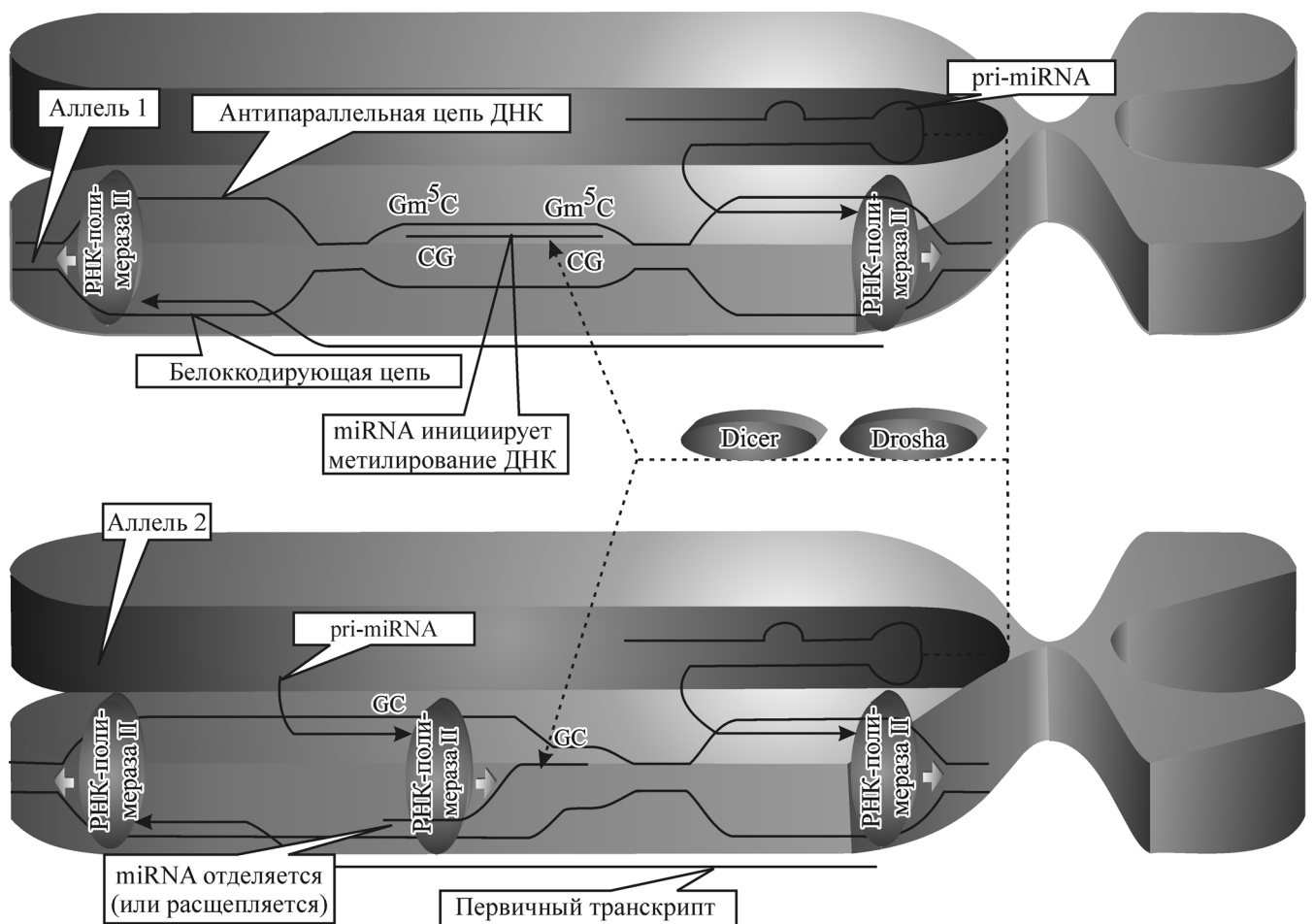


Рис. 3. Механизм аллельного исключения: аллель 1 подвергается репрессии, аллель 2 остается активным. pri-miRNA — предшественник miRNA; остальные обозначения те же, что и на рис. 1, 2.

до того, как последняя привлечет ДНК-метилтрансферазу и инициирует нанесение метильной метки на связанный участок ДНК, модифицирование близлежащих гистонов и дальнейшее ремоделирование хроматина (рис. 3). В итоге обе цепи маркированного аллеля перестают считываться, тогда как другой аллель сохраняет активный статус. Отметим также, что предшественник miRNA в ходе своего синтеза не должен сам вызывать нанесение эпигенетических маркеров, так как он стремится образовать шпильку, вследствие чего его связь с транскрибируемой целью ДНК быстро разрывается (рис. 3).

Очевидно, что репрессированным оказывается тот аллель, который при прочих равных условиях менее активно освобождается от присоединившихся miRNA, т. е. слабее транскрибируется в антипараллельном направлении (например, из-за каких-то особенностей содержащего его и близлежащего хроматина). Можно также предположить, что инициирующий репрессию порог концентрации miRNA для каждого из аллелей не является некоей стандартной величиной, а гибко настраивается, становясь тем выше, чем активнее транскрибируется его антипараллельная цепь и быстрее отделяются гибриды, образовавшиеся с последней малые РНК (рис. 3). Важным следствием этого есть то, что если в клетке начинает экспрессироваться второй аллель, то задающий репрессию порог концентрации miRNA для него может оказаться существенно ниже, чем для уже полностью активного аллеля, и в предельном слу-

чае концентрация малой РНК, считанной с активного аллеля, будучи подпороговой для репрессии его самого, в то же время уже оказывается надпороговой для аллелей, начинающих экспрессироваться, что ведет к их сайленсингу и возникновению аллельного исключения. С другой стороны, благодаря настраиваемому пороговому уровню аллельное исключение может устанавливаться при самых разных уровнях экспрессии подверженных ему генов.

Можно ожидать, что неслучайный рисунок исключения некоторых отцовских и материнских аллелей у межвидовых гибридов связан именно с изначально детерминированным неравенством уровней их экспрессии, вызванным, например, различиями в нуклеотидных последовательностях регуляторных элементов аллелей генов, принадлежащих разным видам.

В другом предельном случае, если все аллели гена примерно одинаково активны, и соответственно может быть одновременно начато их маркирование метками репрессированного хроматина, то на функционировании аллелей будет отражаться противостояние между влиянием факторов, активирующих экспрессию гена (напомним, они способны удалять маркеры молчащего хроматина), и влиянием механизмов транскрипционного сайленсинга, запущенных вследствие накопления miRNA. Уровень концентрации miRNA при этом, очевидно, должен стремиться к значению, околупороговому для устанавливаемой интенсивности транскрипции, но так как это состояние не-



стойчиво, а может быть, и недостижимо, случайные флуктуации при нанесении и удалении эпигенетических меток в конце концов определяют, какой из аллелей подвергнется репрессии, а какой останется активным.

Механизмы, которые опосредуют аллельное исключение и другие эффекты дозы гена, в процессе эволюции возникли, вероятно, с целью поддержания стабильности клеточного генома и противодействия горизонтальному переносу генов. Они подавляли при помощи miRNA функционирование транспозонов и защищали клетки от последствий вызываемой мобильными элементами генома чрезмерной копийности некоторых генов. Также сходные механизмы, используя siRNA, предотвращают размножение внедрившейся в клетку чужеродной генетической информации, вызывают ее сайленсинг или (если чужеродная информация представлена только в виде РНК) полное разрушение. Иногда мишенью влияния данных механизмов становятся гены, целенаправленно перенесенные в клетки при помощи биотехнологических манипуляций.

### Список литературы

- Чуриков Н. А. 2005. Молекулярные механизмы эпигенетики. Биохимия. 70 (4) : 493—513.
- Шуф М. 2005. Метилирование и деметилирование ДНК как мишени для противораковой терапии. Биохимия. 70 (5) : 651—669.
- Araujo F. D., Knox J. D., Szyf M., Price G. B. Zannis-Hadjopoulos M. 1998. Concurrent replication and methylation at mammalian origins of replication. Mol. Cell. Biol. 18 : 3475—3482.
- Bernstein E., Caudy A. A., Hammond S. M., Hannon G. J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 409 : 363—366.
- Bird A. P. 1995. Gene number, noise reduction and biological complexity. Trends Genet. 11 : 94—100.
- Bihler M., Verdel A., Moazed D. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell. 125 : 873—886.
- Di Croce L., Raker V. A., Corsaro M., Fazi F., Fanelli M., Farretta M., Fuks F., Lo Coco F., Kouzarides T., Nervi C., Minucci S., Pelicci P. G. 2002. Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. Science. 295 : 1079—1082.
- Felsenfeld G., Groudine M. 2003. Controlling the double helix. Nature. 421 : 448—453.
- Griffiths-Jones S., Grocock R. J., van Dongen S., Bateman A., Enright A. J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucl. Acids Res. 34 (Database Issue): D140—D144.
- Habu Y., Kakutani T., Paszkowski J. 2001. Epigenetic developmental mechanisms in plants: molecules and targets of plant epigenetic regulation. Curr. Opin. Genet. Develop. 11 : 215—220.
- He L., Hannon G. J. 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. 5 : 522—531.
- Hermann A., Goyal R., Jeltsch A. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J. Biol. Chem. 279 : 48 350—48 359.
- Hsieh C. L. 1994. Dependence of transcriptional repression on CpG methylation density. Mol. Cell. Biol. 14 : 5487—5494.
- Hutvagner G., Zamore P. D. 2002. RNAi: nature abhors a double-strand. Curr. Opin. Genet. Develop. 12 : 225—232.
- Jabbari K., Bernardi G. 2004. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene. 333 : 143—149.
- Jones P. L., Veenstra G. J. C., Wade P. A., Vermaak D., Kass S. U., Landsberger N., Strouboulis J., Wolffe A. P. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat. Genet. 19 : 187—191.
- Kawasaki H., Taira K. 2004. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature. 431 : 211—217.
- Lachner M., O'Carroll D., Rea S., Mechtler K., Jenuwain T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 410 : 116—120.
- Lee Y., Ahn C., Han J., Choi H., Kim J., Yim J., Lee J., Provost P., Radmark O., Kim S., Kim V. N. 2003. The nuclear RNase III Drosha initiates microRNA processing. Nature. 425 : 415—419.
- Lund E., Guttinger S., Calado A., Dahlberg J. E., Kutay U. 2004. Nuclear export of microRNA precursors. Science. 303 : 95—98.
- Martienssen R. A., Colot V. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science. 293 : 1070—1074.
- Moazed D. 2001. Common themes in mechanisms of gene silencing. Mol. Cell. 8 : 489—498.
- Motamedi M. R., Verdel A., Colmenares S. U., Gerber S. A., Gygi S. P., Moazed D. 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding genomic RNAs. Cell. 119 : 789—802.
- Nan X. S., Ng H. H., Johnson C. A., Laherty C. D., Turner B. M., Eisenman R. N., Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature. 393 : 386—389.
- Okano M., Xie S., Li E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat. Genet. 19 : 219—220.
- Peters A. H., Kubicek S., Mechtler K., O'Sullivan R. J., Derijck A. A., Perez-Burgos L., Kohlmaier A., Opravil S., Tachibana M., Shinkai Y., Martens J. H., Jenuwein T. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol. Cell. 12 : 1577—1589.
- Richards E. J., Elgin S. C. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell. 108 : 489—500.
- Riggs A. D., Xiong Z. 2004. Methylation and epigenetic fidelity. Proc. Nat. Acad. Sci. USA. 101 : 4—5.
- Sadaie M., Iida T., Urano T., Nakayama J. 2004. A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. EMBO J. 23 : 3825—3835.
- Scarano E., Iaccarino M., Grippo P., Parisi E. 1967. The heterogeneity of thymine methyl group origin in DNA pyrimidine isotopes of developing sea urchin embryos. Proc. Nat. Acad. Sci. USA. 57 : 1394—1400.
- Schwarz D. S., Zamore P. D. 2002. Why do miRNAs live in the miRNP? Genes Develop. 16 : 1025—1031.
- Stancheva I. 2005. Caught in conspiracy: cooperation between DNA methylation and histone H3K9 methylation in the establishment and maintenance of heterochromatin. Biochem. Cell. Biol. 83 : 385—395.
- Tagami H., Ray-Gallet D., Almouzni G., Nakatani Y. 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell. 116 : 51—61.
- Tchurikov N. A., Chistyakova L. G., Zavilgelsky G. B., Manukhov I. V., Chernov B. K., Golova Y. B. 2000. Gene-specific silencing by expression of parallel complementary RNA in *Escherichia coli*. J. Biol. Chem. 275 : 26 523—26 529.
- Tijsterman M., Plasterk R. H. A. 2004. Dicers at RISC: the mechanism of RNAi. Cell. 117 : 1—3.
- Vaucheret H., Fagard M. 2001. Transcriptional gene silencing in plants: targets, inducers and regulators. Trends Genet. 17 : 29—35.
- Verdel A., Jia S., Gerber S., Sugiyama T., Gygi S., Grewal S. I. S., Moazed D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. Science. 303 : 672—676.
- Verdel A., Moazed D. 2005. RNAi-directed assembly of heterochromatin in fission yeast. FEBS Lett. 579 : 5872—5878.
- Wade P. A., Geggion A., Jones P. L., Ballestar E., Aubry F., Wolffe A. P. 1999. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat. Genet. 23 : 62—66.
- Zamore P. D. 2002. Ancient pathways programmed by small RNAs. Science. 296 : 1265—1269.

HYPOTHESIS OF THE INITIATION OF DNA METHYLATION *DE NOVO* AND ALLELIC EXCLUSION BY SMALL RNA*V. A. Halytskiy*A. V. Palladin Institute of Biochemistry, National Academy of Sciences of Ukraine, Kiev;  
e-mail: volha@biochem.kiev.ua

We have discovered that 5'-CG-3' dinucleotide and 5'-CNG-3' trinucleotide are found in published sequences of small interfering RNA and microRNA more often than they should be found in a random sequence. This circumstance is evidence of an important biological purpose of 5'-CG-3' dinucleotides and 5'-CNG-3' trinucleotides in small RNA sequences. We suppose that small RNAs containing mentioned di- and trinucleotides participate in creation of chromatin marks of epigenetic information through high-specific search of DNA sequences liable to repression and through initiation of the methylation *de novo* of 5'-CG-3' and 5'-CNG-3' sites in DNA fragments, which appeared to be bound complementary with small RNA. Several genes can be inactivated simultaneously when they contain the motif which is recognized by small RNA. Allelic exclusion appears, to our opinion, as a result of initiation by small RNA of *de novo* DNA methylation of all alleles but one that exist in the cell. The predecessor of this small RNA is transcribed from the antiparallel allele chain. Those alleles are inactivated which antiparallel chain is less actively read by RNA-polymerase, which, as we suppose, releases DNA from attached to it small RNA in the process of transcribing. But the quantity of small RNA which is transcribed from just one allele is insufficient to overcome the level when the repression process of this allele *de novo* starts.

Key words: epigenetic information, allelic exclusion, DNA methylation, small RNA, chromatin.

---