

МЕХАНИЗМЫ НОРМАЛЬНОГО И ПАТОЛОГИЧЕСКОГО РАЗВИТИЯ ТКАНЕЙ

УДК 577.152.211

ЭПИГЕНЕТИЧЕСКИЙ МУТАГЕНЕЗ¹ КАК ПРОГРАММА ВОЗРАСТНОЙ ДИСФУНКЦИИ БЕЛКОВ И СТАРЕНИЯ

© 2015 г. Г. А. Романов*, **, В. С. Суховеров***, Б. Ф. Ванюшин*

*Институт физико-химической биологии им. А.Н. Белозерского МГУ им. М.В. Ломоносова
119991 Москва, Ленинские горы

**Институт физиологии растений им. К.А. Тимирязева Российской академии наук
127276, Москва, ул. Ботаническая 35, тел. (499)977-94-09, факс (499)977-80-18

***Институт проблем управления им. В.А. Трапезникова Российской академии наук
117997, Москва, ул. Профсоюзная, 65

E-mail: gromanov@yahoo.com; vanyush@genebee.msu.ru; suhoverv@ipu.ru

Поступила в редакцию 04.09.2014 г.

Окончательный вариант получен 29.09.2014 г.

Метилирование ДНК играет важную полифункциональную роль в онтогенезе человека и млекопитающих. Одним из следствий метилирования ДНК является резкое повышение вероятности мутационной замены в геноме динуклеотида CpG на динуклеотид TpG. Нами охарактеризован весь спектр (17) возможных мутаций ДНК и белков, обусловленных метилированием CpG-динуклеотида в ДНК; выделены 3 наиболее опасные мутации, способные приводить к инактивации белков. Создана компьютерная программа, позволяющая прогнозировать все наиболее вероятные мутации в анализируемом гене и кодируемом им белке. На примере генов человека и ряда млекопитающих показано, что число потенциально опасных сайтов эпигенетического мутагенеза в экзонах резко снижено в результате эволюции геномов. Но при этом установлены невынужденное сохранение таких сайтов и их персистентность, что указывает на наличие заложенной в геноме эпигенетической программы возрастной дисфункции белков, ведущей к апоптозу и старению; эта программа базируется на наборе и позиции метилируемых кодонов в экзонных участках генов. Предполагается, что программа эпигенетического мутагенеза ограничивает продолжительность жизни индивидуума, ускоряя освобождение популяции от особей-долгожителей, завершивших репродуктивный период.

Ключевые слова: ген, белок, эпигенетический мутагенез, метилирование, дезаминирование, геном, мутация, CpG-сайт.

DOI: 10.7868/S047514501502007X

ВВЕДЕНИЕ

Поиск и расшифровка молекулярных и физиологических механизмов финализации онтогенеза (старения) имеют большое научное и практическое значение (Анисимов, 2008). На молекулярном уровне старение объясняют главным образом с помощью одной из двух конкурирующих гипотез: либо стохастическим накоплением дефектов у макромолекул, в первую очередь ДНК, либо реализацией записанной в геноме программы, которую в принципе можно “отключить”, задерживая наступление старости и смерти (Longo et al., 2005; Trindale et al., 2013). В настоящей статье мы покажем, что обе эти конкурирующие гипотезы

могут реализоваться на базе единого механизма, связанного с эпигенетическим мутагенезом.

Известно, что ДНК большинства ныне живущих видов подвергается энзиматической модификации (Ванюшин, 2005). В частности, у млекопитающих и человека примерно каждый двадцатый остаток цитозина в ДНК метилирован с преобразованием в остаток 5-метилцитозина (m^5C). m^5C распределен в геноме не случайно, а находится главным образом в последовательности (5') m^5CpG (3'), которая в двутяжевой ДНК представляет собой динуклеотидный палиндром. В генах млекопитающих подавляющая часть динуклеотидов CpG метилирована (Романов, Ванюшин, 1980; Romanov, Vanyushin, 1981; Mazin, Vanyushin, 1987; Mazzio, Soliman, 2012). Это позволяет определить наиболее вероятную локализа-

¹ В данной работе термин “Эпигенетический мутагенез” означает мутагенез белков, обусловленный метилированием кодирующей ДНК.

цию m^5C в генах, если известна их нуклеотидная последовательность.

Роль метилирования ДНК в процессах онтогенеза многообразна, одной из важнейших функций этой модификации ДНК считается регуляция активности генов и локусов генома (Ванюшин, 2005). Метилирование ДНК существенно изменяется при старении и злокачественном перерождении клеток (Romanov, Vanyushin, 1981; Mazin, 2009; Calvanese et al., 2009). При этом роли метилирования ДНК в накоплении конкретных мутаций в геноме и связанных с этим возрастных дисфункциях уделялось недостаточное внимание, хотя предположение о возможном участии метилирования ДНК в запрограммированной дезинтеграции генома при апоптозе и старении ранее уже выдвигалось (Mazin, 2009). Как правило, исследования биологической роли метилирования ДНК фокусировались на регуляторных, но не структурных участках генома.

В данном исследовании рассматривается возможная роль метилирования кодирующей (экзонной) области гена в процессе возрастных изменений генома и кодируемых белков.

ДНК — генетическая матрица — является одной из относительно стабильных биологических макромолекул, однако и она постоянно испытывает структурные превращения при воздействии разнообразных повреждающих факторов. В числе таких повреждений — однонитевые разрывы, депуринизация и депиримидизация, окисление оснований и др. (Анисимов, 2008). Большинство повреждений ДНК обратимо, так как исправляется той или иной системой репарации. Однако есть и исключения. Среди спонтанных повреждений ДНК отмечено дезаминирование остатков цитозина, которое, по расчетам, случается в среднем около 200 раз на клетку в день (Bernstein, Bernstein, 1991). Цитозин при дезаминировании превращается в урацил, который затем в норме энзиматически выщепляется из ДНК с восстановлением ее исходной структуры. Однако если дезаминирование затрагивает m^5C , то в результате возникает не урацил, а тимин — типичное основание ДНК. В этом случае система репарации часто срабатывает неверно и сохраняет возникший тимин в динуклеотиде (5') TrG (3'), “исправляя” противоположащий динуклеотид CpG в комплементарной цепи на (5') CpA (3'). Таким образом, на месте (метилированного) цитозина в цепи ДНК становится тимин, а в комплементарной цепи — аденин вместо гуанина (Романов, Ванюшин, 1980). Такие точечные замены, естественно, могут приводить к определенным мутациям в белках, анализ которых и составляет предмет настоящего исследования.

МАТЕРИАЛ И МЕТОДИКА

Последовательности генов и белков извлекали для анализа из базы NCBI (<http://www.ncbi.nlm.nih.gov/nucscore/>). Для оценки вероятности возрастной дисфункции того или иного белка определяли состав и распределение кодонов экзонной области соответствующего гена как в исходном состоянии, так и в результате метилирования и последующего дезаминирования остатка цитозина. Для биоинформатического анализа больших массивов экзонных последовательностей ДНК нами была разработана специальная компьютерная программа, позволяющая анализировать переходы CpG → TpG в каждой из комплементарных цепей ДНК и возникающие при этом аминокислотные замены в кодируемом белке. Оконный интерфейс программы содержит меню и две панели, на которые выводятся последовательности нуклеотидов: исходная и измененная. Последовательности разбиты на кодоны и имеют подстрочки аминокислот. Меню программы позволяет задавать обусловленные метилированием замены нуклеотидов в последовательности, определять мутации, возникшие при этом в кодируемом белке, и отображать изменения на панелях. Для наглядности отражения результатов мутационных замен на панелях цветом выделяются сайты изменений, произошедшие в аминокислотном составе кодируемого белка, и отдельно маркируются те из них, которые представляют наибольшую опасность для функционирования белка. После выполнения мутационных замен и визуализации изменений на рабочих панелях программа позволяет выполнить определенную статистическую обработку данных и записать полученные результаты — сводные и отдельно по каждой цепи ДНК — в таблицы Excel. Алгоритм программы обеспечивает сравнение последовательностей генов — до и после мутационных изменений — в общей сложности по 80 параметрам, обозначенным в таблицах. Доступ к таблицам возможен через экранный интерфейс Excel.

Вероятности P нахождения тех или иных фрагментов ДНК (моно-, ди- или олигонуклеотидов) в произвольно выбранной позиции гена определялись на основе данных о нуклеотидном составе экзонной ДНК, как произведение: $P = (P_a)^k (P_c)^l (P_g)^m (P_t)^n$, где P_a, P_c, P_g, P_t — доли нуклеотидов в составе ДНК, а k, l, m, n — соответственно количества каждого из нуклеотидов в искомом фрагменте длиной N , где $N = k + l + m + n$. Для расчета статистически ожидаемого числа M содержания искомого фрагмента в полноразмерной кДНК длиной L н.о. (за вычетом стоп-кодона) результирующая формула имеет вид: $M = P(L - (N - 1))$. В случае динуклеотидов CpG ($N = 2$) формула упрощается до вида:

Таблица 1. Кодоны, подверженные эпигенетическому мутагенезу (подчеркнуты), в таблице универсального генетического кода. CpG-содержащие кодоны выделены жирным шрифтом. Кодоны с возможностью эпигенетически-обусловленных мутаций при их нахождении после цитозина выделены курсивом. Образуемый в результате эпигенетически-обусловленной мутации стоп-кодон выделен фоном. В скобках – обозначения аминокислот

		2-е положение в кодоне			
		A	T	G	C
1-е положение в кодоне	A	AAA (Lys)	ATA (Ile)	AGA (Arg)	ACA (Thr)
		AAG (Lys)	ATG (Met)	AGG (Arg)	ACG (Thr)
		AAT (Asn)	ATT (Ile)	AGT S	ACT (Thr)
		AAC (Asn)	ATC (Ile)	AGC S	ACC (Thr)
	T	TAA Stop	TTA (Leu)	TGA Stop	TCA (Ser)
		TAG Stop	TTG (Leu)	TGG (Trp)	TCG (Ser)
		TAT (Tyr)	TTT (Phe)	TGT (Cys)	TCT (Ser)
		TAC (Tyr)	TTC (Phe)	TGC (Cys)	TCC (Ser)
	G	<i>GAA</i> (Glu)	<i>GTA</i> (Val)	<i>GGA</i> (Gly)	<i>GCA</i> (Ala)
		<i>GAG</i> (Glu)	<i>GTG</i> (Val)	<i>GGG</i> (Gly)	GCG (Ala)
		<i>GAT</i> (Asp)	<i>GTT</i> (Val)	<i>GGT</i> (Gly)	<i>GCT</i> (Ala)
		<i>GAC</i> (Asp)	<i>GTC</i> (Val)	<i>GGC</i> (Gly)	<i>GCC</i> (Ala)
C	CAA (Gln)	CTA (Leu)	CGA (Arg)	CCA (Pro)	
	CAG (Gln)	CTG (Leu)	CGG (Arg)	CCG (Pro)	
	CAT (His)	CTT (Leu)	CGT (Arg)	CCT (Pro)	
	CAC (His)	CTC (Leu)	CGC (Arg)	CCC (Pro)	

$M = P_c P_g (L - 1)$. Например, статистически ожидаемое число CpG-динуклеотидов в кДНК гена каталазы GC-состава 49.37% и общей длиной 1581 н.о. определяется как $M = P_c P_g (L - 1) = 0.2494 \times 0.2443 \times 1580 = 96.27$, т.е. порядка 96.

РЕЗУЛЬТАТЫ

Мы провели исчерпывающий теоретический анализ всех возможных мутаций, возникающих в результате метилирования-дезаминирования динуклеотида CpG. В кодирующей цепи ДНК метилируемый цитозин в динуклеотиде CpG может находиться в одной из 3-х позиций кодона (табл. 1). При нахождении в 1-ой позиции (это кодоны аргинина CGN) после транзиции C → T возникают кодоны цистеина (TGT и TGC), триптофана (TGG) и один стоп-кодон (TGA). При нахождении во 2-ой позиции замена C → T приведет к замене серина (TCG) на лейцин (TTG), пролина (CCG) тоже на лейцин (CTG), треонина (ACG) на метионин (ATG) и аланина (GCG) на валин (GTG). Интересно отметить, что транзиция C → T при нахождении в 3-ей позиции не приводит (в силу вырожденности кода) ни к каким заменам

аминокислот в белке, хотя таких кодонов, оканчивающихся на цитозин, насчитывается 16. Это означает, что далеко не все вызванные метилированием замены C → T в экзонах приводят к мутациям.

Результаты теоретического анализа, представленные в таблице 2, показывают, что метилирование цитозиновых остатков в кодирующей цепи ДНК повышает вероятность шести типов аминокислотных замен. Из них только замена аланина на валин может рассматриваться как замещение однотипных аминокислот, мало влияющее на свойства белка, остальные замены могут существенно повлиять на структурно-функциональное состояние белка. Но наиболее круциальное изменение – это образование преждевременного стоп-кодона TGA на месте аргинина (нонсенс-мутация), что должно приводить к синтезу укороченного и, вероятнее всего, неактивного белка.

Метилирование цитозина с последующим дезаминированием возникшего m⁵C и его трансформация в тимин могут происходить не только в кодирующей, но и в комплементарной матричной цепи ДНК. Транзиция C → T в матричной цепи

Таблица 2. Замены аминокислот в белках вследствие эпигенетического мутагенеза – транзиции CpG → TrG в каждой из цепей ДНК

CpG → TrG переход в кодирующей цепи			CpG → TrG переход в матричной цепи		
исходная аминок-та	конечная аминок-та	число кодонов	исходная аминок-та	конечная аминок-та	число кодонов
Arg (R)	Cys (C)	2	Arg (R)	His (H)	2
Arg (R)	Trp (W)	1	Arg (R)	Gln (Q)	2
Ser (S)	Leu (L)	1	Val (V)	Ile (I)	3
Pro (P)	Leu (L)	1	Val (V)	Met (M)	1
Thr (T)	Met (M)	1	Ala (A)	Thr (T)	4
Ala (A)	Val (V)	1	Asp (D)	Asn (N)	2
Arg (R)	Стоп-кодон	1	Glu (E)	Lys (K)	2
			Gly (G)	Ser (S)	2
			Gly (G)	Arg (R)	2
Двойной переход в каждой из цепей ДНК:			Ala (A)	Met (M)	1

приводит к замене G → A в том же месте кодирующей цепи. В этом случае характер мутационных замен аминокислот также строго детерминирован и зависит от положения остатка аденина в соответствующем кодоне.

Результаты влияния на кодируемые белки транзиций C → T в матричной цепи ДНК представлены в таблице 2. В этом случае возможно 9 типов замен аминокислот, стоп-кодоны не образуются. Среди всех возможных мутаций только замены аргинина на гистидин и валина на изолейцин можно считать относительно мало влияющими на свойства белка. Напротив, замещения простейшей аминокислоты глицина (боковая цепь – 1 атом водорода) на объемный, заряженный аргинин (сложная боковая цепь состава C₄N₃H₁₁) или глутаминовой кислоты на основной лизин скорее всего приведут к серьезной структурно-функциональной деформации белка.

Таким образом, вследствие CpG-метилирования обеих цепей ДНК CpG-сайты мутируют двояко: в них либо C замещается T, либо G замещается A. Совместная транзиция обоих нуклеотидов в CpG в результате метилирования-дезаминирования маловероятна, т.к. первая транзиция на любой из цепей ДНК превращает сайт в неметилируемый. Однако кодон GCG для аланина уникален тем, что в нем возможны сразу 2 транзиции в случае его соседства с C предыдущего кодона (т.е. здесь могут соседствовать два CpG-динуклеотида в одной цепи). Одна из комбинаций этих двух транзиций приведет к замене аланина на метионин (табл. 1, 2). В общей сложности метилирова-

ние остатков цитозина теоретически может повысить мутабельность 23 кодонов из имеющихся 61, т.е. почти 40% всех табличных кодонов. Эпигенетические мутации затрагивают кодоны 9 аминокислот, которые в результате 16 вариантов превращений замещаются другими, общим числом 13 (плюс одна нонсенс-мутация). Часть таких превращений была предсказана ранее (Cooper, Yuossoufian, 1988). Из 20 белковых аминокислот только 2 ароматические (фенилаланин и тирозин) никак не связаны (не исчезают или возникают) с эпигенетическим мутагенезом. При этом потенциальные эпигенетически-обусловленные мутации не случайны, а предопределены нуклеотидной последовательностью метилируемых кодонов и различаются по степени опасности для функции белка. Наиболее опасными надо признать замену аргинина на стоп-кодон, а также замены глицина и глутаминовой кислоты на кардинально отличные от них аминокислоты, соответственно на аргинин и лизин.

Основная функция ДНК заключается в хранении и передаче генетической информации, поэтому стабильность ДНК является важнейшим условием функционирования генома. Однако горячие точки мутаций, вызванных метилированием ДНК, очевидно, представляет реальную опасность для стабильности генетической матрицы и жизнедеятельности клетки и организма. Естественно ожидать, что геномы видов, у которых происходит CpG-метилирование ДНК, эволюционировали таким образом, чтобы уменьшить вероятность нежелательных мутаций. Действительно,

доля динуклеотидов CpG у млекопитающих резко снижена, до 20% от статистически ожидаемой (Романов, Ванюшин, 1980; Romanov, Vanyushin, 1981; Mazin, Vanyushin, 1987; Mazzio, Soliman, 2012). Столь резкое отклонение встречаемости от наиболее вероятной характерно исключительно для динуклеотида CpG. Кроме того, опасность эпигенетически-обусловленных мутаций существенно уменьшается благодаря вырожденности генетического кода. Как следует из теоретического анализа (табл. 2), в экзонной области гена наиболее подвержены мутациям кодоны аргинина из-за присутствия 4 мутабельных CpG-содержащих кодонов как в кодирующей, так и в матричной цепях ДНК. При этом нонсенс-мутация кодона CGA может в большинстве случаев приводить к инактивации белка. Однако аргинин относится к аминокислотам, которым соответствует (в универсальном генетическом коде) наибольшее число кодонов – 6. Помимо CpG-содержащих, у аргинина имеются и безопасные в отношении метилирования кодоны AGA и AGG. Замена метилируемых кодонов аргинина на неметилируемые, для чего требуется лишь одна трансверсия C → A в кодирующей цепи, исключила бы возможность опасных нонсенс-мутаций. Из двух других аминокислот, чьи замены могут блокировать функциональность белка, глицин имеет 4 кодона, 2 из которых (GGU и GGC) в случае метилирования-дезаминирования переходят в кодоны серина, а не аргинина, причем замещение глицина небольшим нейтральным серином представляется существенно менее опасным для активности белка, чем замещение аргинином. У глутаминовой кислоты синонимичных кодоном GARPu безопасных кодонов нет, но по крайней мере для аргинина и глицина эпигенетически-“опасные” кодоны не уникальны и могли бы быть заменены в ходе эволюции на синонимичные безопасные или малоопасные кодоны. Но, главное, кодоны Gly (GGPu) и Glu (GAPu) “опасны” лишь тогда, когда предшествующий кодон оканчивается на C, который в этой позиции всегда может быть заменен на T без какого-либо эффекта на структуру белка.

Таким образом, уже на уровне организации генома с учетом вырожденности генетического кода и редкой встречаемости метилируемых CpG-динуклеотидов можно было бы ожидать эволюционно обусловленную элиминацию “опасных” кодонов, по крайней мере из генов, определяющих здоровье и долголетие индивидуума. Присутствие же “опасных” CpG-содержащих кодонов в гене может резко повысить степень вероятности возрастной дисфункции кодируемого белка. Для выяснения наличия, типа и позиции потенциальных эпигенетически-обусловленных мутаций в

генах высших организмов, нами была разработана специальная компьютерная программа, позволяющая анализировать переходы CpG → TpG в обеих цепях кодирующих последовательностей ДНК и возникающие при этом аминокислотные замены в белках. Иными словами, программа дает возможность предсказать все наиболее вероятные мутации в данном гене. Результаты проведенного анализа на примере гена каталазы человека (ген CAT, GenBank accession NM_001752) показаны в таблице 3.

Каталаза является универсальным консервативным гем-содержащим ферментом, выполняющим важную функцию в клетке – разрушение вредной для макромолекул перекиси водорода (Chelikani et al., 2004). Активность каталазы важна для задержки старения и смерти; у трансгенных мышей с суперэкспрессией гена каталазы продолжительность жизни возросла почти на 20% (Schriner et al., 2005; Wu et al., 2007; Ren et al., 2007). Ген каталазы человека имеет общую длину 33136 нуклеотидов и содержит 12 интронов; экзонная часть гена состоит из 1584 нуклеотидов и кодирует белок, состоящий из 527 аминокислот. Кодирующая структура гена содержит 45 динуклеотидов CpG, что составляет примерно 47% от статистически ожидаемого (96) (табл. 3). Такое уменьшение числа CpG примерно вдвое снижает вероятность нежелательных эпигенетически-обусловленных мутаций. Количество кодонов, способных мутировать в несинонимичные в результате метилирования ДНК, в гене каталазы человека равно 43 (29 CpG на кодирующей цепи, 36 на матричной цепи, из них 22 кодона аргинина общие, плюс одна двойная замена), что составляет всего 8.1% от общего числа кодонов. Таким образом, доля мутагенных кодонов (8.1%) в гене каталазы снижена в 4.7 раз по сравнению с теоретически ожидаемой (37.8%), при этом среди оставшихся CpG-содержащие кодоны аргинина составляют более половины.

Всего кодонов аргинина в гене каталазы 31, среди них неметилируемых (безопасных) типа AGA оказалось всего 5, а остальные имеют в своем составе динуклеотиды CpG. Среди последних присутствуют 2 наиболее “опасных” кодона типа CGA, способных мутировать в преждевременные стоп-кодона. Анализ метилома человека (http://neomorph.salk.edu/ips_methylomes/browser.html) подтвердил реальность такого перехода, так как все найденные CpG-сайты гена каталазы практически полностью метилированы *in vivo*. Потенциальные стоп-кодона имеют порядковые номера 66 и 382, т.е. при их образовании белок каталазы будет иметь длину не 527 аминокислот, как в норме, а 65 или 381. Очевидно, что в обоих случаях

Таблица 3. Результаты анализа гена и белка каталазы человека, полученные с помощью программы выявления потенциальных эпигенетически-обусловленных мутаций

<i>Переходы CpG → TrG в кодирующей последовательности ДНК^a</i>		
Число нуклеотидов в последовательности		1584
Число кодонов		528
Общее число CpG		45
Ожидаемое статистически число CpG		96–97
Число CG: С в первой позиции кодона		22
Число CG: С во второй позиции кодона		9
Число CG: С в третьей позиции кодона		14
Число новообразуемых стоп-кодонов		2
Позиции стоп-кодонов в последовательности ДНК		66 и 382
<i>Вызванные CpG → TrG переходами замены аминокислот в белке</i>		<i>(число замен)^a</i>
Кодирующая цепь ДНК	Arg(R) → Cys(C)	10
	Arg(R) → Trp(W)	10
	Ser(S) → Leu(L)	1
	Pro(P) → Leu(L)	3
	Thr(T) → Met(M)	0
	Ala(A) → Val(V)	5
	Число замен аминокислот в последовательности	29
Матричная цепь ДНК	Arg(R) → His(H)	10
	Arg(R) → Gln(Q)	12
	Val(V) → Ile(I)	2
	Val(V) → Met(M)	1
	Ala(A) → Thr(T)	4
	Asp(D) → Asn(N)	3
	Glu(E) → Lys(K)	0
	Gly(G) → Ser(S)	2
	Gly(G) → Arg(R)	2
	Число замен аминокислот в последовательности	36
Двойной переход в каждой из цепей ДНК	Ala(A) → Met(M)	1
Количество возможных типов замен аминокислот		15

^a Замены, наиболее опасные для функциональности белка, выделены жирным шрифтом.

укороченный белок должен потерять свою энзиматическую активность, с учетом того, что основная функциональная область каталазы человека находится в пределах от 68-го до 497-го аминокислотных остатков (Zamosky et al., 2008). Метилирование матричной цепи ДНК может вызвать также “опасный” переход глицин → аргинин в кодонах №№ 490 и 516. Однако эти остатки гли-

цина расположены на периферии белка и их замены могут не сильно отразиться на его функциональных свойствах. Всего же в гене насчитывается 70 кодонов глицина (GGN), теоретически 17–18 из них должны находиться после остатка цитозина, однако на практике таких кодонов всего 4 (вероятность случайности такого отклонения $P < 0.0001$): два из них могут трансформироваться

Таблица 4. Консервативность эпигенетически “опасных” кодонов на примере генов каталаз представителей разных родов млекопитающих. Мутабельные кодоны и нуклеотиды, сохраняющиеся в данной позиции белка, выделены жирным шрифтом

Вид	Продолжительность жизни, лет ^a	Кодоны в генах каталаз:				
		Thr29	Gly30	Arg47	Arg203	Arg382
Мышь	4	acc	gga	cga	cga	cga
Крыса	3	acc	gga	cga	cga	cgc
Волк	14	acc	ggg	cgg	cga	cga
Бык	20–25	act	gga	cga	cga	cga
Кабан	20–30	acc	ggc	cga	cga	agg
Человек	100–120	act	gga	cgt	cgg	cga

^a По данным Флиндт (1992).

в кодоны аргинина, что отмечалось выше, а еще два (№№ 390 и 465) – к более “щадящему” переходу в кодоны серина. Знакомство со структурой каталазы показывает, что в последнем случае такие мутации не затронут функционально важных аминокислот. В гене каталазы человека имеется также 25 кодонов GArU для глутаминовой кислоты, которые могли бы трансформироваться в кодоны для нежелательного лизина при их нахождении после цитозина предыдущего кодона. Однако ни один из этих кодонов Glu после C не находится, хотя среднестатистически таких кодонов должно быть порядка 6 ($P < 0.001$). Очевидно, что первичная структура гена каталазы эволюционировала таким образом, чтобы многократно уменьшить вероятность нежелательных мутаций, вызванных метилированием цитозина в динуклеотиде CpG. Но важно отметить, что структурные основы для таких мутаций (особенно аргинин → нонсенс переход) все же сохраняются, хотя, казалось бы, нонсенс-мутации могли бы быть полностью исключены заменой опасных кодонов на синонимичные неметируемые. При этом в гене каталазы сохранены структурные предпосылки и для других деструктивных эпигенетически-обусловленных мутаций. Помимо нонсенс-мутаций, в той же каталазе человека потенциальные мутационные переходы, вызванные метилированием ДНК, затрагивают аргинин-354, важный для связывания гема (возможны переходы аргинин → цистеин или гистидин); аргинин-365 (возможны переходы аргинин → цистеин или гистидин) и пролин-391 (возможен переход пролин → лейцин), важные для формирования тетрамера каталазы. Кроме того, мутации метилируемых сайтов ДНК, соответствующих кодоном №№ 380

(Arg), 382 (Arg), 388 (Arg), 390 (Asp/Gly) и 391 (Pro), могут блокировать предполагаемые сайты фосфорилирования тирозина в белке.

Таким образом, на примере гена *CAT* человека можно заключить, что первичная структура гена организована таким образом, чтобы исключить абсолютное большинство, хотя и не все потенциально “опасные” метилируемые сайты в экзонах. Эти последние в ходе онтогенеза с высокой вероятностью вызовут конкретную предсказуемую мутацию, приводящую к полной или частичной инактивации белка. Такая особенность каталазы человека далеко не уникальна: в генах каталаз других млекопитающих также сохраняется некоторое количество метилируемых кодонов CGA – предшественников нонсенс-мутаций. У свиней таких кодонов 2, у собак и коров – 3, а у мышей и крыс – целых 5 (табл. 4). На примере генов каталаз прослеживается обратная зависимость видовой продолжительности жизни от числа CGA-кодонов. Отметим, что положение CGA-кодонов в генах инвертаз не хаотично, а устойчиво воспроизводится в одних и тех же позициях генов (табл. 4), несмотря на дивергенцию этих видов от общего предка более 65 миллионов лет назад. Этот факт служит прямым доказательством неслучайности присутствия “опасных” кодонов в геноме: если бы эти кодоны возникали в результате некоего стохастического мутационного равновесия, то позиции “опасных” кодонов в генах давно дивергировавших видов были бы различными. При этом факт превращения отдельных “опасных” кодонов в менее опасные или неметируемые с сохранением белковой последовательности подтверждает отсутствие каких-либо физических ограничений для таких переходов. В генах каталаз

Таблица 5. Прогнозируемые эпигенетически-обусловленные мутации в значимых для продолжительности жизни генах человека

Гены	Ассоциированные болезни ¹	Общее число кодонов	Нонсенс-мутации		Транзиции Glu→Lys		Транзиции Glu→Arg	
			Число	Позиция	Число	Позиция	Число	Позиция
<i>SAT</i>	Рак, альфа-талассемия	528	2	66, 382	0		2	490, 516
<i>APOСIII</i>	Сахарный диабет	100	1	19	1	24	0	
<i>APOE</i>	БА, ССЗ	318	0		12	31, 37, 84, 139, 149, 186, 189, 197, 230, 249, 273, 284	2	183, 296
<i>ATM</i>	Атаксия-телеангиэктазия	3057	21	23, 35, 248, 250, 447, 457, 805, 1466, 1618, 1730, 1875, 1882, 1898, 2034, 2443, 2486, 2598, 2763, 2849, 2993, 3047	4	871, 1169, 2039, 2052	3	138, 204, 2508
<i>CSA</i>	Синдром Коккейна	397	1	268	0		0	
<i>CYP1A1</i>	Рак, лейкозы	513	2	77, 377	0		1	225
<i>CYP1B1</i>	Рак, горм. нарушения	544	2	355, 444	9	100, 167, 173, 176, 220, 229, 260, 262, 473	3	232, 286, 306
<i>ERCC2_1</i>	Синдром Коккейна	761	3	88, 156, 196	12	20, 100, 235, 278, 284, 313, 327, 419, 576, 589, 606, 712	4	7, 47, 593, 615
<i>ERCC2_2</i>	Синдром Коккейна	406	3	64, 132, 172	6	76, 211, 254, 260, 289, 303	1	23
<i>ERCC6</i>	Синдром Коккейна	1494	15	77, 176, 453, 612, 629, 652, 683, 735, 857, 947, 975, 1087, 1221, 1288, 1467	4	272, 387, 495, 647	5	53, 411, 798, 1382, 1410
<i>HSP70-A1A</i>	Снижение стресс-резистентности	642	2	258, 311	19	48, 117, 132, 175, 218, 268, 303, 367, 386, 444, 446, 460, 475, 516, 530, 534, 564, 588, 591	12	52, 98, 202, 224, 338, 356, 372, 407, 437, 609, 618, 623
<i>LMNA</i>	Синдром Хатчинсона	703	3	225, 321, 596	9	65, 73, 82, 84, 290, 347, 358, 383, 554	4	13, 86, 438, 523
<i>PON1</i>	ССЗ	356	4	27, 32, 214, 306	1	49	0	
<i>WRN</i>	Синдром Вернера	1433	9	225, 369, 732, 741, 889, 987, 1305, 1321, 1406	1	975	1	1362

¹ Сокращения: ССЗ – сердечно-сосудистые заболевания; БА – болезнь Альцгеймера.

рассмотренных видов млекопитающих присутствуют также кодоны, определяющие возможные транзиции глутаминовая кислота → лизин (Glu → Lys) и глицин → аргинин (Gly → Arg); наибольшее число (4) “опасных” кодонов Glu найдено в гене мыши, а “опасных” кодонов Gly (3) – в гене крысы, оба вида с относительно короткой (3–4 года) продолжительностью жизни.

С помощью созданной программы мы проанализировали также ряд генов человека, имеющих отношение к процессам старения, на присутствие кодонов “опасных” эпигенетически-обусловленных мутаций в кодируемых белках (табл. 5). Взятые для анализа гены характеризуются самыми разными свойствами и функциями, необходимыми для жизнедеятельности клетки (Анисимов, 2008; Wolters, Schumacher, 2013): репарация ДНК (гены *ATM*, *CSA*, *ERCC*), раскручивание ДНК (*WRN*), стабильность ядра и транскрипции (*Lamin A*), антиоксидантная защита (*CAT*, *PON1*), восстановление структуры белков (*HSPA1A*), метаболизм ксенобиотиков (*CYP1A1*, *CYP1B1*) и липидов (*APOCIII*, *APOE*). Тем не менее основные черты, важные для эпигенетического мутагенеза, в этих генах сохраняются. В подавляющем большинстве генов присутствует небольшое число (1–3) кодонов CGA – потенциальных стоп-кодонов, способных инактивировать белок судя по их положению в гене. В генах крупных белков (*ATM*, *ERCC6*) число таких “опасных” кодонов может превысить 1–2 десятка, что указывает на вероятную недолговечность активности таких белков в онтогенезе. Редкий случай полного отсутствия “опасных” CGA-кодонов в небольшом гене *APOE* компенсируется присутствием необязательных “опасных” кодонов GGRu глицина и многочисленных (12) – “опасных” кодонов GARu глутаминовой кислоты. В абсолютном большинстве изученных генов (кроме гена *CSA*), в дополнение к потенциальным стоп-кодонам, также имеются другие “опасные” кодоны, эпигенетически-обусловленные мутации которых могут с большой вероятностью вызвать инактивацию белка. В целом проведенный анализ показал большое разнообразие генов по набору и соотношению “опасных” кодонов: в одних генах (*ATM*, *ERCC6*, *PON1*, *WRN*) явно преобладают кодоны CGA, в других (*APOE*, *CYP1B1*, *ERCC2*, *HSPA1A*, *Lamin A*) – кодоны, грозящие заменами Glu → Lys и Gly → Arg, в остальных (*APOCIII*, *CAT*, *CSA*, *CYP1A1*) содержание “опасных” кодонов всех видов крайне мало. Хотя обнаружены гены с полным отсутствием тех или иных типов этих кодонов (гены *APOCIII*, *CAT*, *CSA*, *CYP1A1*, *PON1*), тем не менее общим правилом является наличие хотя бы одного или нескольких эпигенетически “опасных” кодонов в

каждом гене (табл. 5). При этом все “опасные” кодоны не уникальны и могли бы быть легко переведены путем той или иной одиночной нуклеотидной замены в менее опасные или безопасные синонимы.

ОБСУЖДЕНИЕ

В результате теоретического анализа мы описали все 17 типов возможных мутаций белков при метилировании и дезаминировании цитозина в последовательности CpG в составе экзона. Среди этих эпигенетически-обусловленных мутаций отмечены те, которые чреваты наихудшими последствиями для структуры и функции белка, это переходы Arg → стоп-кодон, а также Glu → Lys и Gly → Arg. Варианты аминокислотных замен определяются строением кодона, содержащего CpG, что позволяет охарактеризовать гены с точки зрения их подверженности эпигенетическому мутагенезу, исходя из экзонных последовательностей ДНК.

Нами был проанализирован ряд генов человека и млекопитающих с помощью специально созданной программы, позволяющей прогнозировать все возможные эпигенетически-обусловленные мутации в гене, в первую очередь наиболее опасные для структуры и функции белков. Можно было ожидать, что в результате длительной эволюции геномов число возможных эпигенетически “опасных” сайтов будет сведено к минимуму, т.е. практически к нулю. Действительно, в генах человека и млекопитающих число кодонов, подверженных эпигенетическому мутагенезу, оказалось в целом во много раз меньше статистически ожидаемого. Тем не менее, определенное, хотя и очень небольшое, число таких кодонов присутствует в подавляющем большинстве проанализированных генов, причем позиции этих кодонов в гене консервативны даже у давно дивергировавших видов, что указывает на неслучайность их наличия в геноме. Впрочем, нельзя исключить, что консервативность расположения части “опасных” кодонов может объясняться консервативными особенностями структуры мРНК, а именно наличием функциональных элементов гена типа энхансеров сплайсинга, примерно 10% которых включают CpG-динуклеотиды (Caceres, Hurst, 2013). Таким образом, отдельные CpG-содержащие кодоны могут выполнять множественные функции в геноме: с одной стороны, поддерживать жизнедеятельность клетки, участвуя в кодировании белка и в сплайсинге пре-мРНК, с другой стороны, ограничивать жизнь клетки посредством эпигенетического мутагенеза. Стоит отметить обнаружение генов для неко-

торых особо крупных белков, где насчитывается более 1–2 десятков этих “кодонов смерти”.

Важно подчеркнуть, что все “опасные” метилируемые кодоны не являются необходимыми и могли бы без особых сложностей быть заменены в ходе эволюции синонимичными кодонами, малоопасными или безопасными в отношении эпигенетического мутагенеза. Сохранение же “кодонов смерти” в экзонах предполагает их участие в генетически запрограммированных процессах возрастной инактивации гена и кодируемого белка. Эта функция в равной мере может выполняться как “консервативными” CpG, так и возникающими в результате спонтанных замен оснований в ДНК. При этом количество, положение и степень опасности кодонов, ответственных за эпигенетический мутагенез, позволяют оценить степень подверженности белка, кодируемого данным геном, возрастной дисфункции.

Рассмотренный нами эпигенетический мутагенез ответственен хотя и не за все, но за значительную часть мутаций генома, ведущих к дисфункции белков. Известно, что более трети (до 70%) всех мутаций в геноме человека связаны с CpG динуклеотидами, частота $C \rightarrow T$ или $G \rightarrow A$ транзаций в которых может до 40 раз превышать фоновую частоту случайных мутаций (Cooper, Youssoufian, 1988; Mazin, 2009). Таким образом, эпигенетический мутагенез объединяет обе основные гипотезы старения: с одной стороны, он запрограммирован, так как мутирующие кодоны и направленность их изменения предопределены исходной структурой гена; с другой стороны, он стохастичен, так как сроки и очередность отдельных мутаций неопределенны и могут варьировать в широких пределах. Это свидетельствует о том, что механизмы апоптоза и старения действительно закодированы в геноме, однако они далеко не жесткие и носят в значительной степени случайный характер. Правда, в связи с открытием у позвоночных животных деаминаз ядерной ДНК, способных энзиматически превращать C в U , а m^5C в T (Conticello 2008; Petit et al., 2009; Fritz et al., 2010), случайность эпигенетически-обусловленных мутаций может быть и достаточно условной. К примеру, вызванная теми или иными факторами активация ДНК-деаминаз может вызвать массовые не репарируемые транзаций $m^5C \rightarrow T$ в геноме с потерей функциональности у множества белков, что неизбежно приведет к летальному исходу. В отличие от ядерного генома, эпигенетический мутагенез не может прямо влиять на функционирование генома органелл (митохондрий), т.к. метилирование CpG-сайтов в их ДНК не происходит. Тем не менее митохондриальная ДНК отличается высокой мутабельностью (Данилен-

ко, Давыденко, 2003). По-видимому, запрограммированное старение генома митохондрий, если и происходит, использует иные механизмы. Интересно отметить, что триплет TGA, необходимый для наиболее деструктивных эпигенетически-обусловленных мутаций, в митохондриях кодирует триптофан, а не терминирует трансляцию. Если у TGA была бы такая же функция и в ядерном геноме, степень опасности эпигенетического мутагенеза была бы гораздо ниже.

Имеющиеся в литературе данные в целом согласуются с результатами проведенного нами анализа потенциальных эпигенетически-обусловленных мутаций в генах животных и человека (табл. 4, 5). Сравнительный глобальный анализ геномов приматов выявил наличие более 10^6 экзонных CpG в 14348 генах человека, т.е. более 70 CpG в расчете на 1 ген. При этом 14% этих динуклеотидов (в среднем ~ 10 CpG в расчете на ген) повышали, по мнению авторов, риск вредных мутаций гена, а выборка генов с увеличенным содержанием “опасных” CpG включала и повышенную долю генов, связанных с заболеваниями (Ying, Huttley, 2011). Разумеется, судьба отдельного гена в каждой из клеток строго индивидуальна, т.к. сроки эпигенетически-обусловленных мутаций могут варьировать в широких пределах. К тому же мутировать может либо динуклеотид CpG кодирующей цепи, либо противолежащий CpG матричной цепи ДНК, а это вызовет совершенно разные последствия для белка. Однако вследствие большого числа потенциально “опасных” кодонов (более 10^5 в геноме человека) они в конце концов неизбежно приведут к дисфункции значительного числа генов и белков. При этом индивидуальные различия скорости дезинтеграции генома, по закону больших чисел, должны нивелироваться и, как следствие, потенциальная продолжительность жизни у представителей данного вида должна усредняться, что и наблюдается в действительности. Хотя потеря дифференцированной мутантной клетки для организма и малосущественна (за исключением ее перерождения в раковую), но неизбежное накопление мутаций в клетках-предшественниках и стволовых клетках приведет к необратимым последствиям для всего их потомства. Конечно, в каких-то особо редких случаях деструктивные эпигенетически-обусловленные мутации у данного индивидуума по счастливой случайности могут длительное время отсутствовать, что может способствовать уникальному долгожительству. И, наоборот, при обратном варианте быстрого накопления вредных мутаций неудачливому индивидууму будет отмерен более короткий срок жизни. Что касается видовых различий, то можно предположить, что чем меньше

сайтов эпигенетического мутагенеза в геноме, особенно в важнейших для жизнедеятельности генах, тем на более долгий срок рассчитано существование данного генома, и, соответственно, данного вида. По всей видимости, наследуемая программа эпигенетического мутагенеза выполняет в онтогенезе ограничительную функцию, не допуская чрезмерно долгого существования индивидуума после завершения репродуктивного периода. Это позволяет своевременно освободить популяцию от организмов, уже “бесперспективных” с позиции размножения, но продолжающих претендовать на ограниченные ресурсы. Эпигенетическая программа возрастной дисфункции отличается простотой и эффективностью, не требуя для своего поддержания и осуществления особых материальных и/или энергетических ресурсов. Важно отметить, что понимание пусковых механизмов апоптоза и старения, запрограммированных непосредственно в кодирующей структуре генов и связанных с эпигенетической модификацией (метилированием) ДНК, дает основу для направленного вмешательства в геном с целью изменения закрепленной в эволюции видовой продолжительности жизни.

Работа поддержана программой Президиума РАН “Молекулярная и клеточная биология”. Выражаем благодарность д.б.н. Л.И. Патрушеву за полезные рекомендации при подготовке материала к публикации.

СПИСОК ЛИТЕРАТУРЫ

- Анисимов В.Н.* Молекулярные и физиологические механизмы старения. Т. 1. С.-П.: Наука, 2008. 482 с.
- Ванюшин Б.Ф.* Эпигенетическое метилирование ДНК – эпигенетический контроль за генетическими функциями клетки // *Биохимия*. 2005. Т. 70. С. 598–611.
- Даниленко Н.Г., Давыденко О.Г.* Миры геномов организмов. Минск: Техналоҗя, 2003. 495 с.
- Мазин А.Л., Ванюшин Б.Ф.* Потеря динуклеотидов CpG из ДНК. I. Метилированный и неметилированный компартменты генома у эукариот с различным содержанием 5-метилцитозина в ДНК // *Молекулярная биология*. 1987. Т. 21. С. 543–551.
- Романов Г.А., Ванюшин Б.Ф.* Метилирование ДНК у эукариот I. Метилируемые последовательности и ДНК-метилазы // *Науч. докл. высш. школы, Биол. науки*. 1980. № 11. С. 5–20.
- Флиндт Р.* Биология в цифрах. М.: Мир, 1992. 304 с. (*Flindt R.* *Biologie in Zahlen*. Stuttgart, New York: Gustav Fischer Verlag, 1988).
- Bernstein C., Bernstein H.* *Aging, Sex, and DNA Repair*. San Diego: Acad. Press, Inc., 1991.
- Cáceres E.F., Hurst L.D.* The evolution, impact and properties of exonic splice enhancers // *Genome Biol*. 2013. V. 14: R143.
- Calvanese V., Lara E., Kahn A., Fraga M.F.* The role of epigenetics in aging and age-related diseases // *Ageing Res. Rev*. 2009. V. 8. P. 268–276.
- Chelikani P., Fita I., Loewen P.C.* Diversity of structures and properties among catalases // *Cell. Mol. Life Sci*. 2004. V. 61. P. 192–208.
- Conticello S.G.* The AID/APOBEC family of nucleic acid mutators // *Genome Biol*. 2008. V. 9: 229.
- Cooper D.N., Youssoufian H.* The CpG dinucleotide and human genetic disease // *Hum. Gen*. 1988. V. 78. P. 151–155.
- Fritz E.L., Papavasiliou F.N.* Cytidine deaminases: AIDing DNA demethylation? // *Genes Dev*. 2010. V. 24. P. 2107–2114.
- Longo V.D., Mitteldorf J., Skulachev V.P.* Programmed and altruistic ageing // *Nature Reviews Genetics*. 2005. V. 6. P. 866–872.
- Mazin A.L.* Suicidal function of DNA methylation in age-related genome disintegration // *Ageing Res. Rev*. 2009. V. 8. P. 314–327.
- Mazzio E.A., Soliman K.F.* Basic concepts of epigenetics: impact of environmental signals on gene expression // *Epigenetics*. 2012. V. 7. P. 119–130.
- Petit V., Vartanian J.-P., Wain-Hobson S.* Powerful mutators lurking in the genome // *Phil. Trans. R. Soc. B*. 2009. V. 364. P. 705–715.
- Ren J., Li Q., Wu S. et al.* Cardiac overexpression of antioxidant catalase attenuates aging-induced cardiomyocyte relaxation dysfunction // *Mech. Ageing Dev*. 2007. V. 128. P. 276–285.
- Romanov G.A., Vanyushin B.F.* Methylation of reiterated sequences in mammalian DNAs. Effects of the tissue type, age, malignancy and hormonal induction // *Biochim. Biophys. Acta*. 1981. V. 653. P. 204–218.
- Schriner S.E., Linford N.J., Martin G.M. et al.* Extension of murine life span by overexpression of catalase targeted to mitochondria // *Science*. 2005. V. 308. P. 1909–1911.
- Trindade L.S., Aigaki T., Peixoto A.A., Balduino A., Mânica da Cruz I.B., Heddle J.G.* A novel classification system for evolutionary aging theories // *Front. Genetics*. 2013. V. 4. Art. 25.
- Wu S., Li Q., Du M. et al.* Cardiac-specific overexpression of catalase prolongs life-span and attenuates ageing-induced cardiomyocyte contractile dysfunction and protein damage // *Clin. Exp. Pharmacol. Physiol*. 2007. V. 34. P. 81–87.
- Wolters S., Schumacher B.* Genome maintenance and transcription integrity in aging and disease // *Front. Genetics*. 2013. V. 4. Art. 19.
- Ying H., Huttley G.* Exploiting CpG hypermutability to identify phenotypically significant variation within human protein-coding genes // *Genome Biol. Evol*. 2011. V. 3. P. 938–949.
- Zamocky M., Furtmüller P.G., Obinger C.* Evolution of catalases from bacteria to humans // *Antioxid. Redox Signal*. 2008. V. 10. P. 1527–1548.

Epigenetic Mutagenesis as Program of Age-Related Protein Dysfunction and Aging

G. A. Romanov^{a, b}, V. S. Sukhoverov^c, and B. F. Vanyushin^a

^a *Belozersky Institute of Physicochemical Biology, Moscow State University, Moscow, 119991 Russia*

^b *Timiryazev Institute of Plant Physiology, Russian Academy of Sciences,
ul. Botanicheskaya 35, Moscow, 127276 Russia*

^c *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences,
ul. Profsoyuznaya 65, Moscow, 117997 Russia*

e-mail: gromanov@yahoo.com, vanyush@genebee.msu.ru, suhoverv@ipu.ru

Received September 4, 2014; in final form, September 29, 2014

Abstract—DNA methylation plays an important polyfunctional role in ontogenesis of human and mammals. A steep rise in probability of mutational substitution of CpG dinucleotide on TpG dinucleotide in the genome is one of the consequences of DNA methylation. All spectrum (17) of possible DNA and protein mutations caused by CpG-dinucleotide methylation in DNA were characterized, and the three most dangerous mutations (able to result in protein inactivation) were isolated. The computer program that allows one to predict all most probable mutations in the analyzed gene and encoded protein was created. On the example of genes from humans and various mammals, it was demonstrated that the amount of potentially dangerous sites of epigenetic mutagenesis in exons was drastically decreased as a result of genome evolution. But, at the same time, unforced preservation of such sites and their persistence were established, indicating the occurrence of age-related protein dysfunction built into the genome epigenetic program, resulting in apoptosis and aging; this program is based on the set and position of methylated codons in exonic gene regions. It is assumed that the program of epigenetic mutagenesis limits the lifetime of an individual, accelerating the deliverance of the population from long-lived individuals that completed the reproductive period.

Keywords: gene, protein, epigenetic mutagenesis, methylation, desamination, genome, mutation, CpG-site