

Сравнительный анализ нуклеотидного состава экзонно-интронных последовательностей генов человека

Д. Р. Дуплий, В. В. Калашников¹, Н. А. Чащин, М. Е. Толсторуков²

Институт молекулярной биологии и генетики НАН Украины
Ул. Академика Заболотного, 150, Киев, 03680, Украина

¹Харьковский национальный экономический университет
Пр. Ленина, 9А, Харьков, 61001, Украина

²Харьковский национальный университет им. В. Н. Каразина
Пл. Свободы, 4, Харьков, 61077, Украина

duplijd@gmail.com; tolstorukov@gmail.com

Исследование асимметрий нуклеотидного состава (АНС), связанных с транскрипцией, проведено в различных функциональных участках генов человека. Охарактеризованы четыре типа АНС: аденин-тиминовая, цитозин-гуаниновая, пурин-пиримидиновая и кето-аминовая. Обнаружены существенные различия в типах АНС в экзонах и интронах, выражающиеся системами соотношений: $A > T$; $G < C$; $A + G > C + T$; $A + C < G + T$ для экзонов и $A < T$; $C > G$; $A + G < C + T$; $A + C < G + T$ для интронов. В нетранскрибируемых областях генома величины АНС близки к нулю, а наибольших значений достигают в интронах генов «домашнего хозяйства». В среднем более высокие значения АНС наблюдаются в группе генов «домашнего хозяйства» по сравнению с тканеспецифическими. Полученные результаты демонстрируют, что подробная информация об АНС может быть использована для дальнейшего понимания генной структуры и организации генома в целом.

Ключевые слова: экзоны, интроны, нуклеотиды, асимметрии.

Введение. Понимание закономерностей организации генома человека невозможно без изучения роли эволюционных процессов в формировании первичной последовательности ДНК. Замечено, что частоты встречаемости комплементарных оснований в одной цепи ДНК неравны: $N_G \neq N_C$; $N_A \neq N_T$ [1], т. е. наблюдается асимметрия нуклеотидного состава. В геномах прокариотов обнаружено преобладание частот встречаемости гуанина над теми цитозина ($N_G > N_C$), а тимина – над аденином ($N_T > N_A$) в лидирующей цепи [2, 3]. Этот факт связывают с направлением репликации [2, 4] и применяют для определения точек инициации репликации в бактериальных и вирусных геномах [5, 6].

Наблюдаемые асимметрии обусловлены различием в мутационных и репарационных механизмах, действующих в кодирующей и матричной цепях [1, 7].

Известно, что причинами нуклеотидных замен, кроме повреждающего действия известных мутагенов, являются процессы репликации и транскрипции [8]. Во время репликации неправильно спаренные с точки зрения комплементарности нуклеотиды репарируются ДНК-полимеразной системой [8]. Если ошибка все же сохраняется до наступления транскрипции, то она может быть устранена другим репарирующим ферментным комплексом, специфичным для транскрипции [9]. Нуклеотидные замены также могут происходить во время транскрипции. В частности, движение РНК-полимеразы

вдоль ДНК провоцирует гидролитическое дезаминирование $m^5\text{Cyt}$ (цитозина) [10] с последующей транзицией его в тимин. Замена $\text{C} \rightarrow \text{T}$ в геноме человека наблюдается почти в три раза чаще, чем обратная ей транзиция $\text{T} \rightarrow \text{C}$, а переход $\text{G} \rightarrow \text{A}$ отмечают в два раза чаще, чем $\text{A} \rightarrow \text{G}$ [11]. Сравнение последовательностей ортологичных генов некоторых млекопитающих показало, что абсолютные значения частот пуриновых замен выше по сравнению с пиримидиновыми [12]. Вместе с тем трансверсии $\text{A} \rightarrow \text{T}$ и $\text{C} \rightarrow \text{G}$ происходят примерно с одинаковыми частотами [11].

Активные кодирующие области репарируются значительно быстрее, чем межгенные промежутки [9]. Таким образом, кодирующие части генов находятся под большим контролем естественного отбора [8] и репарационных механизмов. Если допустить, что нуклеотидные замены являются побочным продуктом процессов репликации и транскрипции, то такие замены, происходящие в линии половых клеток, будут сохраняться в последующих поколениях. Поэтому представляется важным исследовать асимметрии нуклеотидного состава в различных функциональных областях генома человека.

Публикации в этой области неполностью освещают данную проблему. Так, в ряде работ анализировали нуклеотидный состав лишь экзонных частей [12, 13] или только интронных [11, 14], вблизи точек репликации [4–6], транскрипции [15] или в целом вдоль генов человека [16]. Целью данной работы является нахождение взаимосвязи асимметрий нуклеотидного состава, ассоциированных с транскрипцией, с экзонно-интронной структурой и характером экспрессии генов.

Материалы и методы. *Общую выборку генов формировали* на основе данных завершённой 35-й версии генома человека [17] из специализированного архива Национального центра биотехнологической информации США (ftp://ftp.ncbi.nih.gov/genbank/genomes/H_sapiens). Для исследования взята модель женского генома, включающая все аутосомы и одну X-хромосому.

Из текстовых документов с последовательностями ДНК и аннотациями о структуре локусов (контигов) для хромосом 1–22 и X сформирован

массив, включающий 25538 наименований генов. Информация об экзонно-интронной структуре мРНК имеется для 22777 генов, а оставшийся 2761 псевдоген исключен из анализа. Несмотря на то, что для каждого гена описано от 1 до 23 вариантов мРНК, отбирали самый протяженный вариант мРНК, 5'-конец которого расположен наиболее близко к 5'-концу гена; таким образом, выделены 22777 последовательностей.

Чтобы избежать искажения результатов за счет повторения одних и тех же последовательностей, из анализа исключены также высокомолекулярные гены иммуноглобулинов и рецепторов лимфоцитов. Во избежание разброса значений на коротких последовательностях исключены и те гены, у которых длина хоть одного экзона или интрона составляла меньше 10 п. н. Несмотря на то, что подавляющее большинство генов человека имеет прерывистую структуру, описаны 1169 безынтронных генов, которые также изъяты из общей выборки. Таким образом, получен массив из 18566 мРНК, на котором изучали распределение величин асимметрий.

Известно, что на экзонно-интронных границах присутствуют регуляторные последовательности. Поэтому для исключения влияния присутствия этих мотивов при расчете средних величин асимметрий удаляли по 10 нуклеотидов с каждой стороны каждого функционального фрагмента. Так как вследствие этого длины всех последовательностей уменьшались на 20 нуклеотидов, а гены с короткими фрагментами исключали (см. выше), то массив уменьшился до 17830 генов.

Для оценки вклада величин каждой из трех позиций кодона в среднее значение асимметрии экзонов выделены последовательности кодонных позиций. Таким образом, массив составил 10839 генов, для которых рассчитывали средние величины асимметрий.

Последовательности Y-хромосомы и митохондриальной ДНК не рассматривали, перекрывающиеся гены не удаляли, а области повторов не маскировали.

Выделение нуклеотидных последовательностей интронно-экзонных фрагментов генов осуществляли согласно границам, данным в описаниях контигов. Последовательности значащей цепи ана-

лизировали следующим образом: для каждого гена выделяли первый, транскрибируемый экзон, начинающийся с первого 5'-нуклеотида мРНК, первый интрон, внутренние экзоны, внутренние интроны, последний экзон. Кроме того, анализировали последовательности прилегающих к гену областей, которые формировали соответственно из 1000 п. н., отсчитанных влево от 5'-конца и вправо от последнего нуклеотида 3'-конца мРНК. Экзон, не являющийся ни первым, ни последним, классифицировали как внутренний экзон. Все интроны, кроме первого, считали внутренними. В предварительном анализе рассматривали отдельно первые 10 экзонов и интронов (данные не показаны), но из-за однотипности характеристик они были объединены с другими внутренними экзонами (интроны – аналогично). Координаты транскрибирующихся, но нетранслируемых областей – НТО (5'-UTR и 3'-UTR) определяли по разнице границ первичной мРНК (mRNA) и зрелой мРНК (CDS).

Выборки групп генов формировали на основе списков генов «домашнего хозяйства» [18] и тканеспецифических [19] генов. Перекрывающиеся гены получали с помощью анализа генных координат в описаниях контигов. Выделены три группы генов – неперекрывающиеся, частично перекрывающиеся и гены, чьи границы находятся в пределах другого гена.

Асимметрии нуклеотидного состава (АНС) – коэффициенты выборочных скошенностей [7, 11, 20] – рассчитывали для прилегающих к гену областей и каждого экзонно-интронного фрагмента гена. Вычисляли четыре типа асимметрий:

аденин-тиминовая – $AT_{асим} = (N_A - N_T)/(N_A + N_T)$;
 гуанин-цитозиновая – $GC_{асим} = (N_G - N_C)/(N_G + N_C)$;
 пурип-пиримидиновая – $R_{асим} = (N_A + N_G - N_T - N_C)/L$;
 кето-аминовая – $K_{асим} = (N_G + N_T - N_A - N_C)/L$,
 где N_A, N_G, N_T, N_C – абсолютные количества соответствующих нуклеотидов на фрагменте длиной L ($L = N_A + N_G + N_T + N_C$).

Внутренние экзоны (интроны) объединяли для каждого гена и рассчитывали АНС на суммарную длину.

Для изучения изменений нуклеотидного состава вдоль гена последовательности структурных фрагментов выравнивали по левой границе и рас-

считывали асимметрии в каждой позиции, двигаясь слева направо (шагом в 1 нуклеотид), а затем эти же фрагменты генов выравнивали по правой границе и двигались справа налево с тем же шагом.

Статистическую обработку проводили, рассчитывая средние арифметические величины асимметрий, а также средние квадратичные ошибки [12], отображенные на рисунках в виде Y -погрешностей.

Результаты и обсуждение. На начальном этапе исследования значения АНС рассчитывали в каждой позиции функциональных фрагментов генов. Обнаружено, что распределение величин асимметрий вдоль функциональных фрагментов довольно однородно (рис. 1). Отклонения от среднего по фрагменту значения отмечены лишь в непосредственной близости интронно-экзонных границ. Такие изменения величин асимметрий, возможно, обусловлены наличием на границах регуляторных элементов (например, сайтов сплайсинга AG), которые при расчете средних величин асимметрий удаляли (см. «Материалы и методы»). Наиболее существенные отклонения проявляются по величине $GC_{асим}$, которая положительна во всех генных фрагментах, причем в экзонах выше, чем в интронах, а также по $AT_{асим}$ и $R_{асим}$. Дальнейший анализ был сфокусирован на средних значениях АНС для каждого функционального фрагмента гена.

Исследование средних величин АНС выявило устойчивые и существенные различия в нуклеотидном составе (рис. 2). Нетранскрибируемые области, прилегающие к границам гена, имели наименьшие величины асимметрий. Примечательным является то, что значения асимметрий внутренних экзонов и интронов имеют противоположные знаки. Внутренние интроны характеризуются отрицательными $AT_{асим}$ ($-0,0446 \pm 0,0007$) и $R_{асим}$ ($-0,0153 \pm 0,0005$), а экзоны – положительным $AT_{асим}$ ($0,0705 \pm 0,0012$) и $R_{асим}$ ($0,0397 \pm 0,0009$). Причем величины $AT_{асим}$ и $R_{асим}$ экзонов и интронов отличаются почти на 200 %. Существенные различия, включая перемену знака, наблюдаются по $K_{асим}$, равной в экзонах $0,0321 \pm 0,0007$, а в интронах – $0,0331 \pm 0,0004$. Показатель $GC_{асим}$ положителен во всех генных фрагментах, но в интронах ($0,0198 \pm 0,0006$) он несколько выше, чем в экзонах ($0,0145 \pm 0,0010$). Таким образом, максимальных

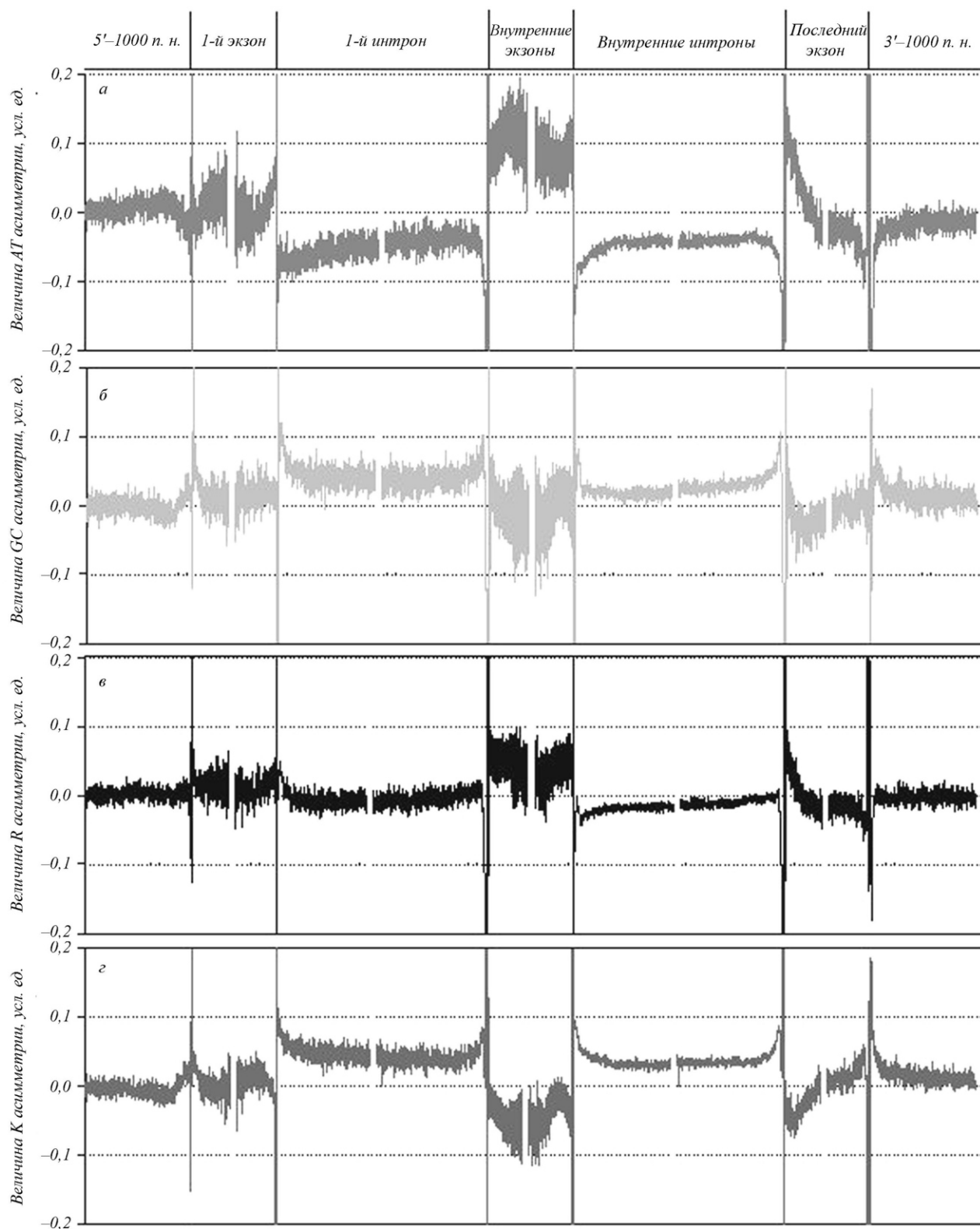


Рис. 1. Распределение величин асимметрий вдоль гена (шаг равен 1 нуклеотиду): *а* – аминок-тиминовая (АТ); *б* – гуанин-цитозиновая (GC); *в* – пурин-пиримидиновая (R); *г* – кето-аминовая (К). По оси абсцисс – функциональный фрагмент ДНК. В экзонах показано по 400 п. н., в интронах – по 1000 п. н. от соответствующих границ

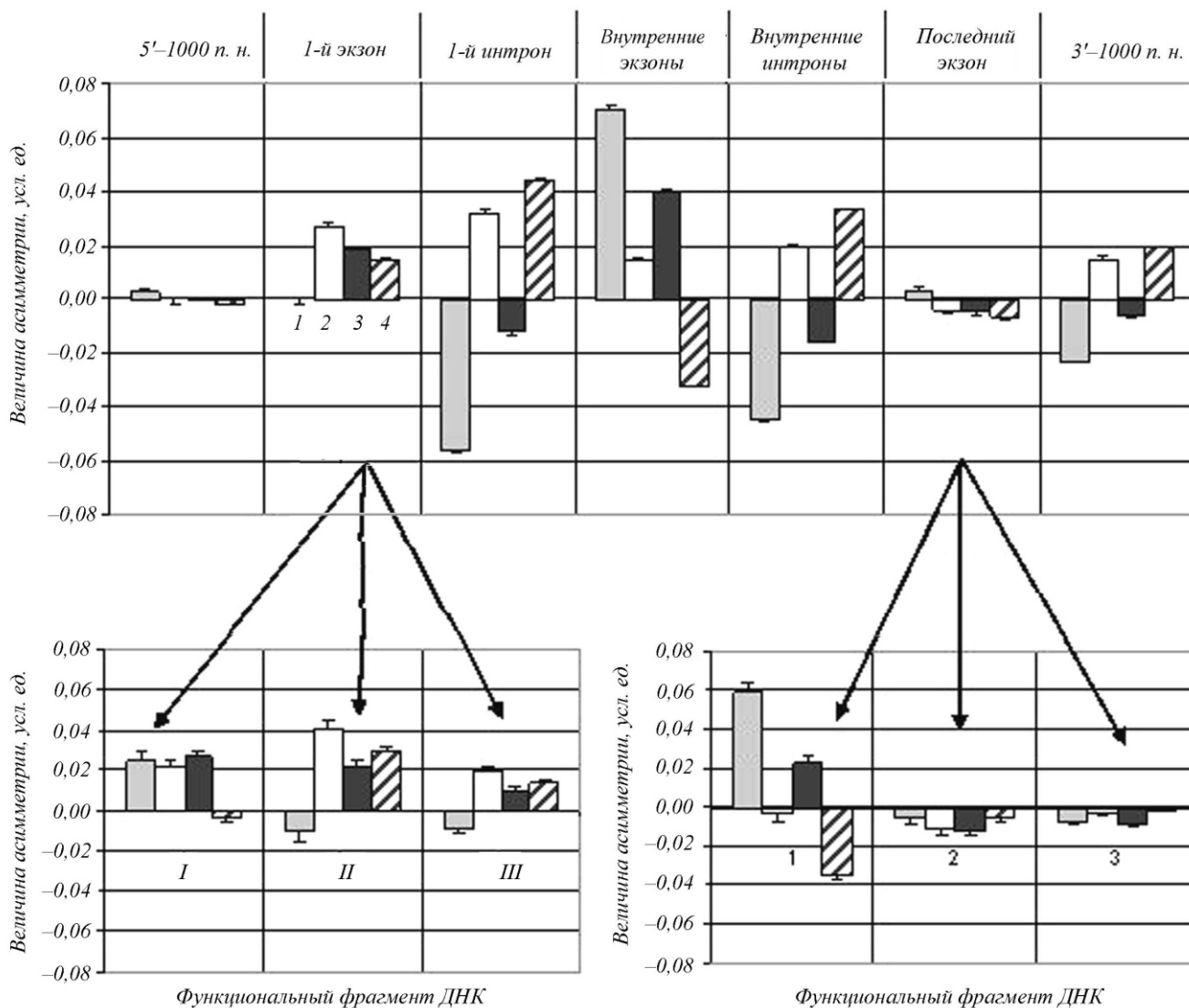


Рис. 2. Средние величины асимметрии нуклеотидного состава экзонов, интронов и прилегающих областей генов человека – *a* (1 – АТ_{асим}; 2 – GC_{асим}; 3 – R_{асим}; 4 – К_{асим}). Стрелками показано разделение первых (*a*) и последних (*б*) экзонов на I – транскрируемые; II – нетранскрируемые; III – смешанные

величин асимметрии нуклеотидного состава достигают в интронах, при этом значения в первом интроне идентичны значениям внутренних интронов.

В отличие от интронов профили АНС первого и внутренних экзонов заметно отличаются (рис. 2, *a*). Поскольку известно, что у многих генов на 5'- и 3'-концах присутствуют НТО, рассмотрен их вклад в величины асимметрий. Все первые и последние экзоны разделили на три группы в зависимости от длины НТО: транскрируемые экзоны (не имеющие НТО); нетранскрируемые экзоны (полностью находящиеся в пределах НТО), а также смешанные (со-

держащие и транскрируемую, и нетранскрируемую части). Из рис. 2, *б, в*, следует, что профиль АНС нетранскрируемых экзонов сходен с таковым внутренних интронов, в то же время профили АНС транскрируемых последних и внутренних экзонов практически совпадают. Это свидетельствует о связи кодирующей функции последовательности с величинами нуклеотидных асимметрий.

На рис. 3 изображены величины асимметрий для каждой позиции кодонов. Заметно, что основной вклад в средние АНС экзонов (рис. 3) вносит состав первых позиций кодонов. Это свиде-

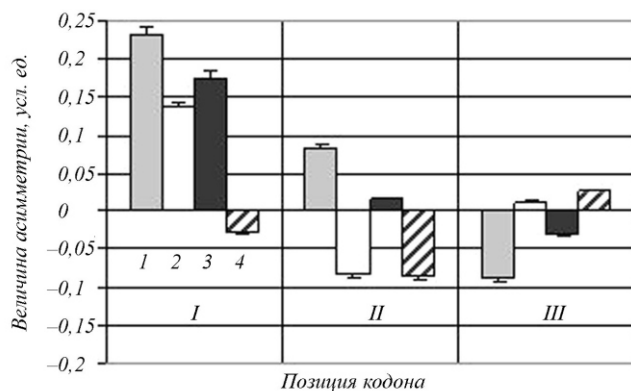


Рис. 3. Средние величины асимметрии нуклеотидного состава в первой (I), второй (II), третьей (III) позициях кодонов транслируемых частей генов, где 1 – AT_{асим}; 2 – GC_{асим}; 3 – RC_{асим}; 4 – KC_{асим}

тельствует о том, что средние величины асимметрий в экзонах определяются, главным образом, функциональными требованиями, а именно – репертуаром аминокислот, кодирующихся данными экзонами.

Таким образом, обнаружена зависимость величин асимметрий нуклеотидного состава от функции генного фрагмента.

Для того чтобы проверить, каким образом выявленные закономерности проявляются в генах с разным характером экспрессии, на следующем этапе исследования сравнивали величины асимметрий в конститутивных и тканеспецифических, а также перекрывающихся и неперекрывающихся генах.

Поскольку известно о существовании генов, способных перекрываться, включая в свою структуру одну и ту же область ДНК [21], интересно было проследить взаимосвязь такой особенности строения генов с нуклеотидными асимметриями.

Выявлено, что полностью перекрываются 2,61 % генов (из 10839), в то время как 91,06 % генов не перекрываются. Примерно такое же соотношение сохраняется и в разных группах генов: из 475 генов «домашнего хозяйства» перекрываются 8 (1,68 %), а из 281 тканеспецифических – 7 (2,49 %). Как можно видеть из рис. 4, а, б профили асимметрий перекрывающихся и неперекрывающихся генов практически не отличаются. Возможно, факт перекрывания интронных или экзонных частей не коррелирует с высокой функциональной активностью. Так, оказалось, что в линии половых клеток

средние величины экспрессии (в условных единицах) генов «домашнего хозяйства» значительно выше (1147 ± 81), чем перекрывающихся (384 ± 44), неперекрывающихся (450 ± 10) и тканеспецифических генов (314 ± 31) [Gene Expression Omnibus (GEO) www.ncbi.nlm.nih.gov/geo/]. Взаимосвязь нуклеотидных асимметрий с уровнем экспрессии представляет самостоятельный интерес и является предметом дальнейших исследований. Тем не менее, уже результаты данной работы демонстрируют, что наибольшие значения асимметрий наблюдаются в группе генов «домашнего хозяйства» (рис. 4). Максимальные значения асимметрий ($AT_{асим} > K_{асим} > GC_{асим}$) обнаружены во внутренних интронах генов «домашнего хозяйства» (рис. 4, в), тогда как величины асимметрий экзонов этих генов несущественно отличаются от соответствующих величин экзонов тканеспецифических (рис. 4, г) и общей выборки генов (рис. 2).

Следует отметить, что количество рассчитанных в работе типов асимметрий несколько избыточно, другие авторы для получения представления о нуклеотидных отклонениях обычно используют $AT_{асим}$ и $GC_{асим}$ [11] или только $GC_{асим}$ [15]. Нами рассчитаны четыре типа асимметрий для того, чтобы выяснить, какие именно нуклеотидные основания создают наблюдаемые отклонения. Кроме того, полученные результаты позволяют представить различия в нуклеотидном составе интронов и экзонов генов человека в виде системы неравенств:

Экзоны	Интроны
$A > T;$	$A < T;$
$G < C;$	$C > G;$
$A + G > C + T;$	$A + G < C + T;$
$A + C < G + T.$	$A + C > G + T.$

Одним из основных различий в составе экзонов и интронов генов человека является то, что экзонные фрагменты существенно обогащены аденином, а интронные – тиминном. Об этом свидетельствует разница величин $AT_{асим}$ (почти в 200 %) между экзонами и интронами (рис. 2). Благодаря $R_{асим}$ обнаружено, что преобладание аденина в экзонах сдвигает равновесие в сторону пуринов ($A + G > C + T$).

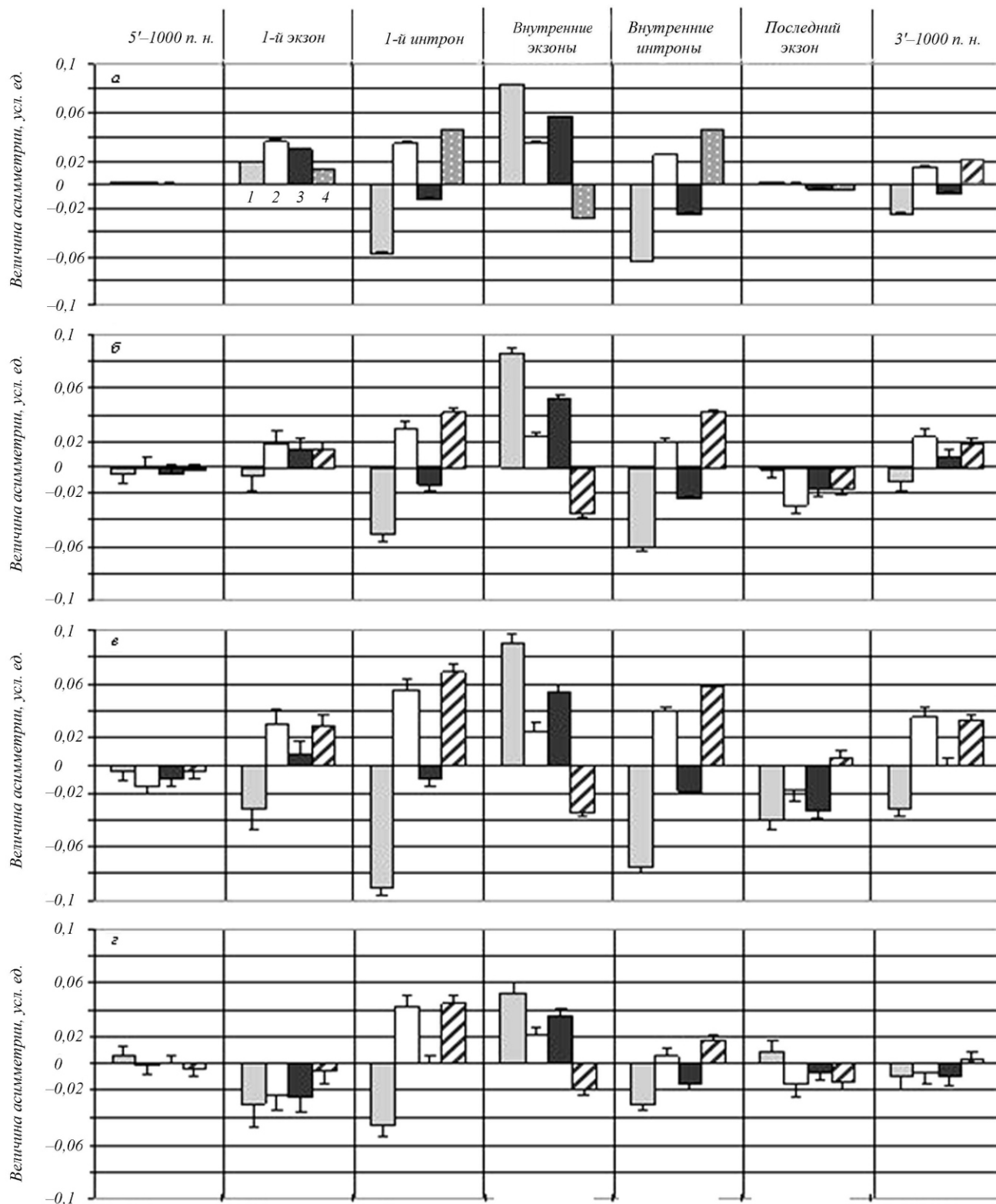


Рис. 4. Средние величины асимметрий нуклеотидного состава в различных группах генов: *а* – неперекрывающиеся; *б* – полностью перекрывающиеся; *в* – гены «домашнего хозяйства»; *г* – тканеспецифические, где 1 – АТ_{асим}; 2 – GC_{асим}; 3 – R_{асим}; 4 – K_{асим}. По оси абсцисс – функциональный фрагмент ДНК

Для состава интронов характерно неравенство $A + C < G + T$, следующее из рассчитанной кето-аминовой асимметрии и характеризующее наиболее частые типы нуклеотидных замен. Действительно, пуриновая транзигия $A \rightarrow G$ и пиримидиновая $C \rightarrow T$ являются наиболее частыми типами нуклеотидных замен [7, 11]. Поэтому $K_{асим}$ позволяет заметить в интронах накопление замен $A \rightarrow G$ и $C \rightarrow T$, чего в экзонах не наблюдается.

Из полученных результатов следует, что величины асимметрий нуклеотидного состава достаточно полно характеризуют различия между интронами и экзонами и демонстрируют связь с функцией кодирования.

Характер экспрессии генов также влияет на значения асимметрий нуклеотидного состава. Наибольшие величины асимметрии нуклеотидного состава наблюдаются в интронах генов «домашнего хозяйства», а наименьшие – в прилегающих к гену нетранскрибируемых областях. Выявленные асимметрии нуклеотидного состава, вероятнее всего, связаны с интенсивностью транскрипции и отражают факт накопления нуклеотидных замен $A \rightarrow G$ и $C \rightarrow T$ в интронах, находящихся под менее сильным давлением отбора. Преобладание именно этих видов замен следует из значений $AT_{асим}$ и $K_{асим}$ в экзонах и интронах.

Связь величин асимметрий с уровнем транскрипции подтверждает и то, что в нетранскрибируемых областях, прилегающих к 5'-концу, значения АНС близки нулю, как и в среднем по геному. В отличие от этого профиль областей АНС, прилегающих к 3'-концу гена, сопоставим с интронными, но меньше по величине. Последнее можно объяснить тем, что сигнал терминации транскрипции не является единственным, он необязательно приводит к остановке РНК-полимеразного комплекса, поэтому фактически транскрипция может продолжаться и за пределами 3'-конца данной мРНК [8].

Необходимо отметить, что величины асимметрий являются усредненными и не отражают индивидуальной вариабельности состава генов. Тем не менее, оценка величин и типов асимметрий показывает статистическую значимость наблюдаемых различий. Таким образом, результаты, полученные в данной работе, могут использоваться как допол-

нительная характеристика структурной организации генов.

Авторы выражают благодарность С. А. Дуплию за многочисленные и плодотворные обсуждения.

D. R. Duplij, V. V. Kalashnikov, M. O. Chaschyn, M. E. Tolstorukov

Comparative analysis of nucleotide composition skew in exons and introns of human genes

Summary

An analysis of transcription-related bias of base-pair composition was performed in different functional regions of human genes. Four types of the bias (skew indexes) were considered: AT-skew, GC-skew, Purine-skew, and Keto-skew. Our results show the essential differences between base-pair composition of exons and introns. On average, exons are characterized by the following rules: $A > T$; $G < C$; $A + G > C + T$; $A + C > G + T$, while the rules for introns are: $A < T$; $C > G$; $A + G < C + T$; $A + C < G + T$. The indexes reach the highest values in the internal introns and are close to zero in nontranscribed regions of the genome. We also observed that the bias is pronouncedly stronger in the housekeeping genes than in the analyzed groups of tissue-specific genes. Our results suggest that the detailed knowledge of the base-pair compositional bias may help to further our understanding of the overall gene organization.

Keywords: exons, introns, nucleotides, skew.

Д. Р. Дуплій, В. В. Калашиников, М. О. Чащин, М. Е. Толсторуков

Порівняльний аналіз нуклеотидного складу екзонно-інтронних послідовностей генів людини

Резюме

Дослідження асиметрій нуклеотидного складу (АНС), пов'язаних з транскрипцією, здійснено в різних функціональних ділянках генів людини. Охарактеризовано чотири типи АНС: аденін-тимінова, цитозин-гуанінова, пурін-піримідинова і кето-амінова. Виявлено істотні відмінності в типах АНС в екзонах та інтронах. Екзони можна охарактеризувати такою системою співвідношень: $A > T$; $G < C$; $A + G > C + T$; $A + C > G + T$, а інтрони – $A < T$; $C > G$; $A + G < C + T$; $A + C < G + T$. У ділянках геному, що не транскрибуються, величини АНС близькі до нуля, а найбільших значень вони досягають в інтронах генів «домашнього господарства». В середньому вищі величини АНС спостерігаються у групі генів «домашнього господарства» порівняно з тканиноспецифічними. Одержані результати демонструють, що докладна інформація щодо асиметрії нуклеотидного складу може бути використана для подальшого розуміння генної структури і організації геному в цілому.

Ключові слова: екзони, інтрони, нуклеотиди, асиметрії.

СПИСОК ЛИТЕРАТУРЫ

1. Sueoka N. Directional mutation pressure, selective constraints, and genetic equilibria // J. Mol. Evol.–1992.– 34.–P. 95–114.

2. Grigoriev A. Analyzing genomes with cumulative skew diagrams // Nucl. Acids Res.–1998.–**26**.–P. 2286–2290.
3. Mrazek J., Karlin S. Strand compositional asymmetry in bacterial and large viral genomes // Proc. Nat. Acad. Sci. USA.–1998.–**95**.–P. 3720–3725.
4. Mackiewicz P., Mackiewicz D., Kowalczyk M. Flip-flop around the origin and terminus of replication in prokaryotic genomes // Genome Biol.–2001.–**15**.–P. 1001–1004.
5. Zawilak A., Cebrat S., Mackiewicz P. Identification of a pupative chromosomal replication origin from *Helicobacter pylori* and its interaction with the initiator protein DNA // Nucl. Acids Res.–2001.–**29**.–P. 2251–2259.
6. Ma X. R., Xiao S. B., Guo A. Z., Lv J. Q., Chen H. C. DNAskew: statistical analysis of base compositional asymmetry and prediction of replication boundaries in the genome sequences // Acta Biochim. et Biophys. Sin. (Shanghai).–2004.–**36**.–P. 16–20.
7. Francino M. P., Ochman H. Strand asymmetries in DNA evolution // Trends Genet.–1997.–**13**.–P. 240–245.
8. Мушкамбаров Н. Н., Кузнецов С. Л. Молекулярная биология.–М.: МИА, 2003.–529 с.
9. Svejstrup J. Q. Transcription repair coupling factor a very pushy enzyme // Mol. Cell.–2002.–**9**.–P. 1151–1152.
10. Beletskii A., Bhagwat A. S. Transcription-induced cytosine-to-thymine mutations are not dependent on sequence context of the target cytosine // J. Bacteriol.–2001.–**183**.–P. 6491–6493.
11. Touchon M., Nicolay S., Arneodo A., d'Aubenton-Cafara Y., Thermes C. Transcription-coupled TA and GC strand asymmetries in the human genome // FEBS Lett.–2003.–**555**.–P. 579–582.
12. Green P., Ewing B., Miller W., Thomas P. J., NISC Comparative Sequencing Program, Green E. D. Transcription-associated mutational asymmetry in mammalian evolution // Nat. Genet.–2003.–**33**.–P. 514–517.
13. Zhang M. Q. Statistical features of human exons and their flanking regions // Hum. Mol. Genet.–1998.–**7**.–P. 919–932.
14. Castillo-Davis C. I., Mekhedov S. L., Hartl D. L., Koonin E. V., Kondrashov F. A. Selection for short introns in highly expressed genes // Nat. Genet.–2002.–**31**.–P. 415–418.
15. Fujimori S., Washio T., Tomita M. GC-compositional strand bias around transcription start sites in plants and fungi // BMC Genomics.–2005.–**6**.–P. 26.
16. Louie E., Ott J., Majewski J. Nucleotide frequency variation across human genes // Genome Res.–2003.–**13**.–P. 2594–2601.
17. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome // Nature.–2004.–**431**.–P. 931–945.
18. Eisenberg E., Levanon E. Y. Human housekeeping genes are compact // Trends Genet.–2003.–**19**.–P. 362–365.
19. Hsiao Li-Li, Dangond F., Yoshida T., Hong R., Jensen R. V., Misra J., Dillon W., Lee K. F., Clark K. E., Haverty P., Weng Z., Mutter G. L., Frosch M. F., MacDonald M. E., Milford E. L., Crum C. P., Bueno R., Pratt E., Mahadevappa M., Warrington J. A., Stephanopoulos Gr., Stephanopoulos G., Gullans S. R. A compendium of gene expression in normal human tissue // Physiol. Genomics.–2001.–**7**.–P. 97–104.
20. Гайдюшев И. Анализ и обработка данных: специальный справочник.–СПб: Питер, 2001.–752 с.
21. Gerstein M. B., Bruce C., Rozowsky J. S., Zheng D., Du J., Korbel J. O., Emanuelsson O., Zhang Z. D., Weissman S., Snyder M. What is a gene, post-ENCODE? History and updated definition // Genome Res.–2007.–**17**.–P. 669–681.

УДК 577.3

Надійшла до редакції 24.04.08