

Структурный анализ группы возможных ДНК-мишеней белков RAG1/2, обнаруженных в геноме мыши *in silico*, и их идентификация в известных типах повторяющихся элементов

А. Ю. Губский, В. Г. Зиньковский¹

Одесский национальный университет им. И. И. Мечникова Министерства образования и науки Украины
Ул. Дворянская, 2, Одесса, 65082, Украина

¹Опольский Университет
Ополе, Польша

gubsky2003@mail.ru

С использованием математических методов анализа, а также специально разработанных алгоритмов установлено, что количество ранее обнаруженных в геноме мыши возможных сайтов – мишеней белков RAG1/2 (cRSS) в 5,4 раза превышает теоретически ожидаемое число. В 71 % случаев cRSS являются структурными элементами 390 типов повторов. В структуре около 5 % мотивов обнаружены нуклеотиды, типичные для большинства сигнальных последовательностей рекомбинации функциональных V, D, J-сегментов Ig- и Tcr-генов мыши (fRSS). Существование 25 % из них в ДНК анализируемого вида можно рассматривать как следствие случайных комбинаций нуклеотидов. Структуры спейсерных участков исследуемых 12cRSS и 23cRSS, как правило, имеют 58–67 и 30–47 % гомологии с аналогичными структурами fRSS соответственно.

Ключевые слова: cRSS, V(D)J-рекомбинация, RAG1, RAG2.

Введение. В норме система V(D)J-рекомбинации осуществляет перестройки генов иммуноглобулинов (Ig) и Т-клеточных рецепторов (TCR) в клетках – предшественниках В- и Т-лимфоцитов [1]. В таких случаях белки RAG1 и RAG2 (продукты одноименных генов, активирующих V(D)J-рекомбинацию, далее белки RAG1/2) в комплексе с белками HMG1, HMG2, Ku70, Ku80, TdT, XRCC4 и т. д. образуют двухцепочечные разрывы ДНК, взаимодействуя с 28±1 и 39±1 нуклеотидными сигнальными последовательностями рекомбинации (RSS), расположенными на границах внутригенных V, D, J-сегментов [2, 3]. Однако при нарушении механизма контроля белки RAG1/2 могут опосредовать разрывы ДНК в других участках генома. Возникающие при этом транслокации и делеции являются

причиной повреждения некоторых генов человека (HPRT, SCL и т. д.) и мыши (*Notch1* и др.) [4–6]. Анализ участков разрывов показывает, что в таких случаях в качестве ДНК-мишеней системы V(D)J-рекомбинации выступают так называемые скрытые мотивы рекомбинации (cRSS), которые структурно могут сильно отличаться от RSS. Эти данные наряду с вариабельностью самих RSS [7, 8] указывают на то, что рассматриваемый белковый комплекс может взаимодействовать с широким спектром последовательностей.

В настоящее время cRSS рассматривают в качестве факторов, потенциально способных опосредовать нестабильность генома млекопитающих при нарушении регуляции системы V(D)J-рекомбинации. Поэтому существует необходимость в количественной и качественной характеристике участков ДНК, которые теоретически могут выступать

в качестве сайтов – мишеней белков RAG1/2, а также в идентификации мест их геномной локализации. Так, впервые используя данные, полученные при анализе эффективности эндонуклеазного расщепления белками RAG1/2 плазмидных субстратов *in vivo*, высказано предположение о том, что в ДНК млекопитающих (человека и мыши) размером около 6 млрд нуклеотидов (диплоидный набор хромосом) может существовать порядка 10^7 cRSS (частота встречаемости $1,7 \cdot 10^{-3}$) [9].

Позднее, исследуя математическими методами последовательности кДНК и геномную ДНК этих видов (общим размером порядка 11 млн н.), Ковелл с соавт. обнаружили около 21 тыс. cRSS. Результаты их исследования предполагают, что возможные сайты – мишени белков RAG1/2 в геномах этих видов должны встречаться с меньшей частотой ($5 \cdot 10^{-4}$) [10].

В свою очередь, проведенное нами исследование аннотированной первичной структуры ДНК 21 хромосомы мыши (Build 35.1) *in silico* показывает, что вне локусов *Ig*- и *Tcr*-генов существуют 6724 cRSS с теоретически высоким рекомбинационным потенциалом [11]. В качестве искомым мотивов выступали 28- и 39-нуклеотидные участки, в которых последовательность пары гептамер/наномер полностью соответствует структуре CACAGTG/ACA AAAACC или имеет с ней 14–15 общих нуклеотидов из 16 возможных. Спейсерные участки имели произвольный нуклеотидный состав. Во всех случаях у гептамеров cRSS структура первых трех функционально важных позиций представлена нуклеотидами CAC. У наномеров в пятой, шестой и седьмой позиции находится тринуклеотид AAA.

Нами установлено, что 2887, 7 и 20 cRSS являются соответственно структурными элементами 2373 белоккодирующих генов, 7 псевдогенов и 20 известных дубликаций сегментов *Ig*- и *Tcr*-генов мыши. Некоторые из cRSS обнаружены в таких повторяющихся элементах, как B1_Mus1, B1_Mus2, MLT1A1, L1_Mus3 и т. д. В белоккодирующих генах около 97 % cRSS расположены в интронах. В 87 и 100 белоккодирующих генах cRSS, у которых спейсерные участки состоят из 12 и 23 нуклеотидов (12cRSS и 23cRSS соответственно), теоретически могут участвовать в образовании делеций и инвер-

сий целых экзонов (гены *Large*, *Rpl23*, *Trhde*, *Suz12*, *Ptprn2*, *Bach2* и т. д.).

Мы считаем, что при нарушении регуляции системы $V(D)J$ -рекомбинации обнаруженные в белоккодирующих генах cRSS теоретически способны опосредовать их повреждения. Поэтому следующий этап нашего исследования найденных мотивов (цель настоящей работы) состоял в том, чтобы рассмотреть возможные механизмы их появления в геноме мыши, а также детально охарактеризовать их нуклеотидный состав. В этой связи нами поставлены следующие задачи: выяснить, существует или нет в геноме мыши эволюционное накопление cRSS; установить общее количество мотивов, являющихся структурными элементами повторяющихся последовательностей (далее – повторов); провести сравнительный анализ структуры cRSS с нуклеотидным составом RSS функциональных V , D , J -сегментов *Ig*- и *Tcr*-генов мыши (fRSS).

Материалы и методы. Количество нуклеотидов А, Т, G, С в ДНК каждой хромосомы мыши определяли при помощи специально разработанных программных алгоритмов. Исходным материалом служила первичная структура ДНК 21 хромосомы (Build 35.1), взятая с сервера NCBI (ftp://ftp.ncbi.nih.gov/genomes/M_musculus/ARCHIVE/BUILD.35.1/) в виде 21 файла формата gbk.

Теоретически ожидаемое количество 28- и 39-нуклеотидных участков (значения M), которые в ДНК хромосом мыши могут соответствовать критериям искомым с определенным типом последовательности гептамер/наномер 12cRSS и 23cRSS (435 уникальных типов последовательностей), определяли по формуле

$$M = (N - 1) P_A^{n_A} P_T^{n_T} P_G^{n_G} P_C^{n_C}, \quad (1)$$

где N – количество нуклеотидов в последовательности ДНК хромосомы; v – общее количество нуклеотидов в последовательности 12cRSS или 23cRSS; P_A, P_T, P_G, P_C – доля нуклеотидов А, Т, G, С в ДНК хромосомы; n_A, n_T, n_G, n_C – общее количество нуклеотидов А, Т, G, С в рассматриваемом типе последовательности гептамер/наномер.

Также по этой формуле находили теоретически ожидаемое количество аналогичных участков, расположенных в комплементарной цепи ДНК хромосом. Сумма полученных значений M для всех 435 уникальных типов 12cRSS и 23cRSS служила величиной, оценивающей общее количество cRSS, существование которых в геноме мыши следует рассматривать как продукты случайных комбинаций нуклеотидов.

В качестве альтернативного подхода в решении этой же задачи нами предложен метод, принцип которого состоял в непосредственном поиске 28- и 39-нуклеотидных аналогов cRSS (tRSS) в наборе последовательностей, имеющих случайные комбинации нуклеотидов (ПЧН). Все пять используемых в работе наборов ПЧН состояли из 21 уникальной последовательности (по размеру и количеству нуклеотидов А, Т, G, С), каждая из которых являлась аналогом нативной ДНК строго определенной хромосомы мыши и создана нами при помощи генератора случайных чисел. Общий размер последовательностей в каждом наборе ПЧН соответствовал размеру генома мыши и составлял 2,57 млрд нуклеотидов (749228240, 750000080, 536396870, 536258140 нуклеотидов А, Т, G, С соответственно). Поиск tRSS в ПЧН проводили теми же программными алгоритмами, что и ранее cRSS в ДНК исследуемого вида.

Среднее арифметическое значение количества tRSS, обнаруженных в пяти наборах ПЧН, принято как величина, оценивающая общее теоретически ожидаемое количество cRSS в геноме мыши. Оба вышеописанных метода являются собственной разработкой.

Для выяснения нуклеотидного состава cRSS применен метод относительных частотных матриц (стандартный при решении таких задач) [12, 13]. Применяемые частотные матрицы получены нами после анализа 199 и 105 последовательностей 12fRSS и 23fRSS *Ig*- и *Tcr*-генов мыши (12- и 23-спейсерные fRSS соответственно). Эти последовательности являются частью fRSS, использованных Ковелл с соавт. в указанной выше работе, и взяты нами с соответствующего интернет-ресурса (<http://www.duke.edu/~Igcowell>). Ячейки матриц содержат информацию об относительной частоте

встречаемости каждого из четырех нуклеотидов А, Т, G, С в конкретно взятой позиции упомянутой группы 12fRSS или 23fRSS. Весовой коэффициент (W) каждой из 6724 исследуемой последовательности cRSS находили по формуле [14]

$$W = \frac{1}{m} \prod_{i=1}^m \left(\frac{n_i}{a_i} \right), \quad (2)$$

где n_i – частота нуклеотида в позиции i , взятая из частотной матрицы; a_i – частота наиболее распространенного в этой позиции нуклеотида; m – общее количество нуклеотидов в анализируемой последовательности. Максимальное значение W , равное 1, подразумевает, что нуклеотидный состав 12cRSS или 23cRSS полностью соответствует консенсусной структуре 12fRSS или 23fRSS. В свою очередь W , равное 0, показывает, что среди последовательностей fRSS нет ни одной, имеющей хотя бы один общий нуклеотид с анализируемой последовательностью cRSS. Те же частотные матрицы применены нами для характеристики нуклеотидного состава последовательностей 12tRSS и 23tRSS, а также самих 12fRSS и 23fRSS.

Чтобы проанализировать 100-нуклеотидные участки, фланкирующие гептамеры cRSS, мы использовали программу BLASTN (<http://www.ncbi.nlm.nih.gov/igblast/>) [15], позволяющую сравнивать нуклеотидный состав исследуемых участков с более чем 68 тыс. различных фрагментов *V*-сегментов *Ig*- и *Tcr*-генов мыши и их известными дубликциями (база данных igSeqNt).

Нами разработаны программные алгоритмы, сопоставляющие координаты cRSS с таковыми повторяющихся элементов, чтобы выяснить, какие именно cRSS в геноме мыши являются частью повторов. Координатами начала и конца cRSS служили установленные ранее координаты 1-го нуклеотида гептамера и 9-го нуклеотида наномера. Координаты 9262721 известных в геноме мыши повторов взяты с сервера NCBI (ftp://ftp.ncbi.nih.gov/gnomes/M_musculus/ARCHIVE/BUILLD.35.1/masking_coordinates.gz).

Найденные типы повторов классифицированы нами, основываясь на данных, взятых из биологической базы данных Repbase [16] и аннотации повторов, обнаруженных при помощи программы

RepeatMasker (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) в некоторых фрагментах геномной ДНК мыши.

Результаты и обсуждение. В качестве искомым ранее в геноме мыши cRSS рассматривали участки ДНК, в которых гептамеры и наномеры, полностью соответствующие последовательности CACAGTG/ACAAAAACC или имеющие с ней 14–15 общих нуклеотидов из 16 возможных, разделены спейсерными участками произвольного состава размером 12 или 23 нуклеотида. Согласно этому, структура 12cRSS, 23cRSS, а также их аналогов, расположенных в комплементарной цепи ДНК, может быть представлена одним из 435 уникальных типов последовательностей гептамер/наномер. Сделав соответствующие расчеты («Материалы и методы», формула (1)), мы установили теоретически ожидаемое количество 12cRSS и 23cRSS, имеющих соответствующие типы последовательностей гептамер/наномер в ДНК каждой из 21 хромосомы мыши. Суммировав полученные данные, определили, что в геноме мыши (размер 2,57 млрд нуклеотидов) в обеих цепях ДНК должно существовать 1238 ± 35 cRSS (по 619 ± 25 как 12cRSS, так и 23cRSS). Аналогичные данные получены нами при анализе пяти ПЧН. В среднем в последовательностях произвольных сочетаний нуклеотидов обнаружено 1241 tRSS (606 12tRSS и 635 23tRSS). Как показывает сравнение, теоретически ожидаемое количество cRSS оказывается в 5,4 раза меньше количества cRSS, реально обнаруженного нами в геноме мыши *in silico*. Это дает основание полагать, что только 18 % cRSS, выявленных ранее в ДНК исследуемого организма, можно рассматривать как продукты случайных комбинаций нуклеотидов. В свою очередь, существование 82 % сайтов-мишеней является следствием эволюционных процессов. К ним нужно отнести 17 12cRSS и 3 23cRSS, обнаруженных нами в участках ДНК, являющихся дубликациями известных V-сегментов генов *Igk* (ген, кодирующий κ -цепь иммуноглобулина) и *Tra* (ген, кодирующий α -цепь T-клеточного рецептора). В этих случаях подобные структуры следует рассматривать как нативные RSS *Ig*- и *Tcr*-генов (далее такие мотивы обозначены как dRSS).

Как упоминалось выше, некоторые из возможных сайтов – мишеней белков RAG1/2 выявлены нами ранее в составе повторяющихся элементов. Поэтому мы предположили, что cRSS рассматриваемой группы в целом могут быть специфичными для этих структур. В этой связи, чтобы подтвердить или опровергнуть выдвинутое предположение, нам необходимо было установить, какое именно количество cRSS является частью повторов. Используя для этой цели специально разработанные программные алгоритмы, мы установили, что 4772 (71 %) cRSS (2077 12cRSS и 2695 23cRSS) являются структурными элементами 4752 повторов. 2779 (58 %), 1975 (41 %), 18 (<1 %) таких мотивов расположены в межгенном пространстве, интронах и экзонах 1698 белоккодирующих генов. В 140 из них (*Adk*, *Ankrd12*, *Abcc4*, *Cacna2d1*, *Ches1*, *Dock1*, *Dpp6*, *Dnahc8*, *Eif2c3*, *Fut8*, *Hk1*, *Il1rap11*, *Katnall1*, *Map3k5*, *Mapk8ip3*, *Pkhd11l1*, *Prkce*, *Prkg1*, *Ptprg*, *Ptprk*, *Tasp1*, *Tcf12*, *Ulk2*, *Zfp262* и т. д.) 12cRSS и 23cRSS теоретически (по правилу «12/23») могут опосредовать образование делеций и инверсий целых экзонов. Обычно в повторах можно обнаружить только одну последовательность 12cRSS или 23cRSS. Исключение составляют 20 повторов (три повтора *B1_Mus1*, по два повтора *B1_Mur2*, *B1_Mus2*, *B4* и по одному повтору *B1_Mur1*, *B1_Mur4*, *PB1D10*, *L1_Rod*, *L1Mca*, *L1VL4*, *L1MEb*, *ORR1A2*, *ORR1D1-i*, *MTD*, *A-rich*), в которых 23cRSS и 12cRSS перекрываются друг с другом. Здесь первые 28 нуклеотидов последовательности 23cRSS выступают в качестве последовательности 12cRSS. Только в 2199 (46 %) случаях нуклеотидная последовательность cRSS полностью расположена внутри повторов (далее – cRSS группы «А»). 1561 (71 %), 582 (26 %), 25 (1 %), 15 (0,7 %) и 16 (0,7 %) из них обнаружены соответственно в неавтономных ретротранспозонах (НРТ), эндогенных ретровирусах и LTR-ретротранспозонах (ЭР/LTR-РТ), ДНК-транспозонах (ДТП), простых повторах (ПП) и не классифицированных нами типах повторов (НТП). В свою очередь 2573 (54 %) cRSS расположены в повторах только частично (далее – cRSS группы «Б») и одна часть cRSS является фрагментом 3'- или 5'-конца повтора, а другая – участком ДНК, фланкирующим этот эле-

Таблица 1

Количественная оценка характера локализации 12cRSS и 23cRSS группы «Б» в структуре разных классов повторов, а также семейств B1, L1, MaLR, ERVK

12cRSS группы «Б»		Количество нуклеотидов 12cRSS, расположенных в повторе						
Локализация	Количество	1–5	6–10	11–15	16–20	21–23	24–25	26–27
НРТ	392	16 ₍₄₎	12 ₍₃₎	20 ₍₅₎	31 ₍₈₎	41 ₍₁₁₎	35 ₍₉₎	237 ₍₆₀₎
ЭР/LTR-РТ	57	18 ₍₃₁₎	2 ₍₄₎	2 ₍₄₎	5 ₍₉₎	8 ₍₁₄₎	7 ₍₁₂₎	15 ₍₂₆₎
ПП	43	17 ₍₃₉₎	5 ₍₁₂₎	11 ₍₂₆₎	6 ₍₁₄₎	3 ₍₇₎	1 ₍₂₎	–
ДТР	3	1 ₍₃₃₎	–	–	–	–	–	2 ₍₆₇₎
НТП	9	1 ₍₁₁₎	3 ₍₃₄₎	–	2 ₍₂₂₎	2 ₍₂₂₎	–	1 ₍₁₁₎
B1	322	2 _(<1)	3 ₍₁₎	5 ₍₂₎	20 ₍₆₎	35 ₍₁₁₎	30 ₍₉₎	227 ₍₇₀₎
L1	27	1 ₍₄₎	6 ₍₂₂₎	9 ₍₃₃₎	5 ₍₁₉₎	3 ₍₁₁₎	1 ₍₄₎	2 ₍₇₎
MaLR	42	11 ₍₂₆₎	1 ₍₂₎	1 ₍₂₎	4 ₍₁₀₎	7 ₍₁₇₎	7 ₍₁₇₎	11 ₍₂₆₎
ERVK	7	3 ₍₄₃₎	–	–	1 ₍₁₄₎	1 ₍₁₄₎	–	2 ₍₂₉₎

23cRSS группы «Б»		Количество нуклеотидов 23cRSS, расположенных в повторе						
Локализация	Количество	1–5	6–10	11–20	21–25	26–29	30–31	32–38
НРТ	1887	21 ₍₁₎	18 ₍₁₎	40 ₍₂₎	75 ₍₄₎	105 ₍₆₎	1436 ₍₇₆₎	192 ₍₁₀₎
ЭР/LTR-РТ	94	43 ₍₄₅₎	1 ₍₁₎	5 ₍₅₎	8 ₍₉₎	8 ₍₉₎	15 ₍₁₆₎	14 ₍₁₅₎
ПП	66	16 ₍₂₄₎	2 ₍₃₎	25 ₍₃₈₎	10 ₍₁₅₎	7 ₍₁₁₎	2 ₍₃₎	4 ₍₆₎
ДТР	6	–	2 ₍₃₃₎	1 ₍₁₇₎	–	2 ₍₃₃₎	–	1 ₍₁₇₎
НТП	16	–	3 ₍₁₉₎	4 ₍₂₆₎	2 ₍₁₂₎	2 ₍₁₂₎	3 ₍₁₉₎	2 ₍₁₂₎
B1	1761	4 _(<1)	2 _(<1)	22 ₍₁₎	58 ₍₃₎	97 ₍₆₎	1416 ₍₈₀₎	162 ₍₉₎
L1	35	6 ₍₁₇₎	10 ₍₂₈₎	8 ₍₂₃₎	5 ₍₁₄₎	3 ₍₉₎	–	3 ₍₉₎
MaLR	64	38 ₍₅₉₎	1 ₍₂₎	5 ₍₈₎	5 ₍₈₎	3 ₍₅₎	4 ₍₆₎	8 ₍₁₂₎
ERVK	17	4 ₍₂₄₎	–	–	1 ₍₆₎	4 ₍₂₄₎	7 ₍₄₁₎	1 ₍₆₎

Примечание. НРТ – неавтономные ретротранспозоны; ЭР/LTR-РТ – эндогенные ретровирусы и LTR-ретротранспозоны; ПП – простые повторы; ДТР – ДНК-транспозоны; НТП – неклассифицированные типы повторов; в индексе цифры в скобках означают проценты.

мент. 2279 (89 %), 151 (6 %), 109 (4 %), 9 (0,3 %) и 25 (1 %) cRSS этой группы были обнаружены нами в НРТ, ЭР/LTR-РТ, ПП, ДТР и НТП соответственно. Наиболее часто представители группы «Б», а это – 322 (64 %) 12cRSS и 1761 (85 %) 23cRSS – являются структурными элементами повторов семейства B1. В табл. 1 подробно представлены данные о том, какое количество нуклеотидов 12cRSS и 23cRSS можно обнаружить в составе рассматриваемых элементов. Так, наиболее часто, в структуре ЭР/LTR-РТ расположены всего 1–5 или 26–27 нук-

леотидов последовательности 12cRSS. В 51 % случаев 3'- и 5'-концы повторов семейства L1 представлены 6–20-нуклеотидными фрагментами 23cRSS.

Исследуемые cRSS обнаружены нами в структуре 390 уникальных типов повторов. Из них 345 являются членами таких 18 суперсемейств, как AcHobo, B1, B2, B4, CR1, ERV1, ERVK, ERVL, ID, L1, L2, MaLR, Mariner, MER1, MER2, MIR, MuDR и Tip100. Остальные 32 типа относятся к простым повторам ((A)_n, (CA)_n, (CAA)_n, (T)_n, (TG)_n,

(TTTG)*n*, (CATG)*n*, (CTGTG)*n*, (TCTG)*n*, (GAA)*n* и т. д.), а 13 – к группе НТП (типы A-rich, T-rich, GA-rich, RMER1B, MMSAT4, YREP_Mm и др.). Количественная оценка показывает, что наиболее часто cRSS являются частью повторов B1_Mus1 и B1_Mus2 семейства B1. На их долю приходится 22 и 21 % всех cRSS. В целом в структуре 30 типов повторов этого семейства (типы B1_Mus1, B1_Mus2, B1_Mur1, B1_Mur1, B1_Mur1, B1_Mur1, B1F, PB1, PB1D1, PB1D10, PB1D10B, PB1D10I, PB1D10M, PB1D9 и т. д.) нами обнаружено 3111 cRSS (65 % общего количества мотивов, найденных в повторах). В повторах следующих трех семейств – MaLR (74 типа повторов: MTD, MTC, MTEa, ORR1A1, ORR1A2, ORR1B1, ORR1B2, MLT1C, MLT1B, MLT1A1 и т. д.), L1 (102 типа повторов: Lx9, Lx8, Lx7, Lx6, Lx5, L1VL4, L1Md_F, L1_Rod, L1_Mur3, L1_Mur2, L1_Mur1 и т. д.) и ERVK (46 типов повторов: ETnERV, ETnERV2, RMER6B, RMER6A, RMER17B, RLTR8, RLTR27, RLTR25B, RLTR25A, IAP-d и т. д.) выявлено 12, 9 и 2 % cRSS соответственно.

Полученные в результате исследования данные показывают, что 33 и 38 % cRSS полностью или частично являются структурными элементами повторов. Поэтому мы считаем справедливым, если такие мотивы будут рассматриваться как структуры, в большей степени специфичные для последовательностей именно этих элементов генома. В 97 % случаях предполагаемые разрывы ДНК, опосредуемые белками RAG1/2 при взаимодействии с cRSS, будут происходить внутри повторов, а в 3 % – в примыкающих к ним участках ДНК, так как именно в них у 165 cRSS группы «Б» расположен гептамер или как минимум его первый нуклеотид. В настоящее время нельзя однозначно ответить на вопрос, в каких именно типах повторов появление cRSS вызвано эволюционными процессами. Однако с полной уверенностью можно утверждать, что существование большинства cRSS в повторах не является случайным. Это основано на том, что общее количество исследуемых мотивов, являющихся частью подобных элементов, в 3,8 раза превышает свое теоретически ожидаемое общегеномное количество. Что именно служит этому причиной? Мы считаем, что распространение в ге-

номе мыши некоторых типов повторов может сопровождаться как возникновением полноразмерных cRSS, являющихся их структурными элементами, так и появлением cRSS группы «Б» *de novo* непосредственно в участках транспозиции. Таким образом, определенные типы повторяющихся элементов могут обеспечивать эволюционное накопление в геноме мыши возможных сайтов – мишеней белков RAG1/2.

Используя метод частотных матриц, нами установлено, что нуклеотидный состав исследуемых мотивов, являющихся и не являющихся частью повторов, в целом очень отличается от сигнальных последовательностей рекомбинации функциональных сегментов *Ig*- и *Tcr*-генов мыши (подробные результаты представлены в табл. 2). Диапазоны значений весовых коэффициентов у 12cRSS и 23cRSS равны 0,59–0,95 и 0,57–0,88 соответственно. Наблюдаемое отличие, как и следовало ожидать, главным образом состоит в особенностях нуклеотидного состава их спейсерных участков. Так, нами установлено, что 0,6; 0,2; 0,5; 7; 41; 31; 16; 4; 0,2 % спейсеров 12cRSS являются 100; 92; 83; 75; 67; 58; 50; 42; 33 %-ми гомологами спейсеров 12fRSS. В свою очередь, 0,1; 0,03; 0,3; 1; 3; 8; 19; 31; 33; 5; 0,03 % спейсеров 23cRSS имеют 100, 91, 65, 61, 57, 52, 48, 43, 39, 35, 30 %-ю гомологию со спейсерами 23fRSS. В общем структура найденных мотивов очень вариабельна. Она представлена 57, 146, 392, 1878 и 2623 уникальными типами гептамеров, наномеров, сочетаний гептамер/наномер, 12- и 23-нуклеотидными типами спейсеров соответственно.

Только 121 (4 %) 12cRSS и 182 (5 %) 23cRSS имеют весовые коэффициенты, равные 0,80–0,95 и 0,75–0,88 (всего 303 cRSS). В структуре именно этих мотивов можно обнаружить нуклеотиды типичные для последовательностей 140 (70 %) 12fRSS и 80 (76 %) 23fRSS. К этой группе мотивов относятся уже упомянутые выше последовательности 20 dRSS. Остальные 163, 3, 113 и 4 cRSS локализованы в межгенном пространстве, псевдогенах *Kif22-ps*, *LOC50777*, *Olfr1452-ps1*, интронах 78 белоккодирующих генов (*Astn1*, *Grik3*, *Jak1*, *Kras*, *Mapk4*, *Pcsk5*, *Rb1cc1* и т. д.) и экзонах генов *Xkr5*, *LOC545698*, *Slc30a4*, *Neb* соответственно. В генах

Таблица 2

Результаты анализа нуклеотидного состава cRSS, обнаруженных в разных участках генома мыши, а также последовательностей fRSS и tRSS, методом частотных матриц

Структура			Количество структур, имеющих определенные значения весовых коэффициентов						
Локализация	Тип	Количество	0,95–0,90	0,89–0,85	0,84–0,80	0,79–0,75	0,74–0,70	0,69–0,65	0,64–0,57
Ig- и Tcr-гены	12fRSS	199	54 ₍₂₇₎	45 ₍₂₃₎	41 ₍₂₁₎	20 ₍₁₀₎	20 ₍₁₀₎	10 ₍₅₎	9 ₍₄₎
	23fRSS	105	1 ₍₁₎	34 ₍₃₂₎	27 ₍₂₆₎	18 ₍₁₇₎	16 ₍₁₅₎	4 ₍₄₎	5 ₍₅₎
ПСН	12tRSS	606	–	1 _(<1)	18 ₍₃₎	127 ₍₂₁₎	258 ₍₄₃₎	175 ₍₂₉₎	27 ₍₄₎
	23tRSS	635	–	–	3 _(<1)	55 ₍₉₎	235 ₍₃₇₎	280 ₍₄₄₎	62 ₍₉₎
ДНК ¹	12cRSS	3034	8 _(<1)	14 _(<1)	99 ₍₃₎	474 ₍₁₆₎	1797 ₍₅₉₎	583 ₍₁₉₎	59 ₍₂₎
	23cRSS	3690	–	1 _(<1)	10 _(<1)	171 ₍₅₎	1591 ₍₄₃₎	1660 ₍₄₅₎	257 ₍₇₎
ДНК ²	12cRSS	2077	1 _(<1)	4 _(<1)	52 ₍₂₎	286 ₍₁₄₎	1372 ₍₆₆₎	343 ₍₁₇₎	19 ₍₁₎
	23cRSS	2695	–	–	2 _(<1)	99 ₍₄₎	1230 ₍₄₅₎	1230 ₍₄₅₎	134 ₍₅₎
ДНК ³	12cRSS	932	–	–	39 ₍₄₎	188 ₍₂₀₎	424 ₍₄₆₎	240 ₍₂₆₎	41 ₍₄₎
	23cRSS	988	–	–	4 _(<1)	72 ₍₇₎	358 ₍₃₆₎	432 ₍₄₄₎	122 ₍₁₂₎
МП ^A	12cRSS	498	–	3 _(<1)	24 ₍₅₎	108 ₍₂₂₎	223 ₍₄₅₎	121 ₍₂₄₎	19 ₍₄₎
	23cRSS	533	–	–	1 _(<1)	35 ₍₇₎	197 ₍₃₇₎	233 ₍₄₄₎	67 ₍₁₂₎
МП ^B	12cRSS	1170	1 _(<1)	2 _(<1)	32 ₍₃₎	169 ₍₁₄₎	743 ₍₆₄₎	214 ₍₁₈₎	9 ₍₁₎
	23cRSS	1609	–	–	2 _(<1)	63 ₍₄₎	693 ₍₄₃₎	768 ₍₄₈₎	83 ₍₅₎
ИБГ ^A	12cRSS	394	–	–	17 ₍₄₎	75 ₍₁₉₎	174 ₍₄₄₎	109 ₍₂₈₎	19 ₍₅₎
	23cRSS	430	–	–	4 ₍₁₎	35 ₍₈₎	151 ₍₃₅₎	189 ₍₄₄₎	51 ₍₁₁₎
ИБГ ^B	12cRSS	904	–	2 _(<1)	20 ₍₂₎	116 ₍₁₃₎	628 ₍₆₉₎	129 ₍₁₄₎	9 ₍₁₎
	23cRSS	1071	–	–	–	35 ₍₃₎	535 ₍₅₀₎	451 ₍₄₂₎	50 ₍₅₎
ЭБГ ^A	12cRSS	44	–	–	1 ₍₂₎	5 ₍₁₁₎	24 ₍₅₅₎	12 ₍₂₇₎	2 ₍₅₎
	23cRSS	26	–	–	–	2 ₍₈₎	10 ₍₃₈₎	10 ₍₃₈₎	4 ₍₁₆₎
ЭБГ ^B	12cRSS	4	–	–	–	1 ₍₂₅₎	3 ₍₇₅₎	–	–
	23cRSS	14	–	–	–	1 ₍₇₎	3 ₍₂₂₎	8 ₍₅₇₎	2 ₍₁₄₎
ПГ	12cRSS	3	–	–	2 ₍₆₇₎	–	–	–	1 ₍₃₃₎
	23cRSS	4	–	–	1 ₍₂₅₎	–	1 ₍₂₅₎	2 ₍₅₀₎	–
ДС ¹	12cRSS	17	7 ₍₄₁₎	7 ₍₄₁₎	3 ₍₁₈₎	–	–	–	–
	23cRSS	3	–	1 ₍₃₃₎	2 ₍₆₇₎	–	–	–	–
ДС ²	12cRSS	2	–	1 ₍₅₀₎	1 ₍₅₀₎	–	–	–	–
	23cRSS	2	–	–	2 ₍₁₀₀₎	–	–	–	–
Del	12cRSS	91	–	1 ₍₁₎	1 ₍₁₎	11 ₍₁₂₎	59 ₍₆₅₎	17 ₍₁₉₎	2 ₍₂₎
	23cRSS	91	–	–	–	2 ₍₂₎	52 ₍₅₇₎	30 ₍₃₃₎	7 ₍₈₎
Inv	12cRSS	104	–	–	3 ₍₃₎	13 ₍₁₃₎	75 ₍₇₂₎	12 ₍₁₁₎	1 ₍₁₎
	23cRSS	104	–	–	–	2 ₍₂₎	60 ₍₅₈₎	35 ₍₃₃₎	7 ₍₇₎

Примечание. ПСН – последовательности случайных комбинаций нуклеотидов; ДНК¹ – вся анализируемая ДНК мыши; ДНК² – участки генома, включающие только повторы; ДНК³ – участки генома, не являющиеся повторами и дупликациями V сегментов Ig- и Tcr-генов мыши; МП – межгенное пространство; ИБГ – интроны белоккодирующих генов; ЭБГ – экзоны белоккодирующих генов; ПГ – псевдогены; ДС¹ – дупликации, содержащие dRSS; ДС² – дупликации, содержащие d’RSS. Del и Inv – cRSS, теоретически способные опосредовать делеции и инверсии в белоккодирующих генах; Степени ^A и ^B означают, что в данных участках генома мыши анализируемые структуры соответственно не являются и являются частью повторов; в индексе цифры в скобках означают проценты.

Таблица 3

Результаты анализа нуклеотидного состава 12cRSS и 23cRSS, расположенных в разных классах повторов, методом частотных матриц

Класс повторов	Группа	cRSS		Количество cRSS, имеющих определенные значения весовых коэффициентов						
		Тип		0,95–0,90	0,89–0,85	0,84–0,80	0,79–0,75	0,74–0,70	0,69–0,65	0,64–0,57
		12/23	Количество							
НРТ	А	12cRSS	1181	–	–	22 ₍₂₎	142 ₍₁₂₎	865 ₍₇₃₎	141 ₍₁₂₎	11 ₍₁₎
		23cRSS	380	–	–	1 _(<1)	35 ₍₉₎	143 ₍₃₈₎	167 ₍₄₄₎	34 ₍₉₎
	Б	12cRSS	392	–	–	7 ₍₂₎	56 ₍₁₄₎	277 ₍₇₁₎	49 ₍₁₂₎	3 ₍₁₎
		23cRSS	1887	–	–	–	32 ₍₂₎	935 ₍₅₀₎	858 ₍₄₅₎	62 ₍₃₎
ЭР/LTR-PT	А	12cRSS	360	–	4 ₍₁₎	14 ₍₄₎	45 ₍₁₂₎	157 ₍₄₄₎	135 ₍₃₈₎	5 ₍₁₎
		23cRSS	222	–	–	1 _(<1)	21 ₍₉₎	91 ₍₄₁₎	89 ₍₄₀₎	20 ₍₉₎
	Б	12cRSS	57	1 ₍₂₎	–	6 ₍₁₀₎	14 ₍₂₄₎	34 ₍₆₀₎	2 ₍₄₎	–
		23cRSS	94	–	–	–	9 ₍₁₀₎	24 ₍₂₆₎	54 ₍₅₇₎	7 ₍₇₎
ДТР	А	12cRSS	9	–	–	–	2 ₍₂₂₎	3 ₍₃₃₎	4 ₍₄₅₎	–
		23cRSS	16	–	–	–	–	4 ₍₂₅₎	10 ₍₆₂₎	2 ₍₁₃₎
	Б	12cRSS	3	–	–	2 ₍₆₇₎	–	1 ₍₃₃₎	–	–
		23cRSS	6	–	–	–	–	5 ₍₈₃₎	1 ₍₁₇₎	–
ПП	А	12cRSS	9	–	–	–	6 ₍₆₇₎	1 ₍₁₁₎	2 ₍₂₂₎	–
		23cRSS	6	–	–	–	–	4 ₍₆₇₎	2 ₍₃₃₎	–
	Б	12cRSS	43	–	–	–	15 ₍₃₅₎	22 ₍₅₁₎	6 ₍₁₄₎	–
		23cRSS	66	–	–	–	2 ₍₃₎	18 ₍₂₇₎	39 ₍₅₉₎	7 ₍₁₁₎
НТП	А	12cRSS	14	–	–	1 ₍₇₎	4 ₍₂₉₎	6 ₍₄₃₎	3 ₍₂₁₎	–
		23cRSS	2	–	–	–	–	2 ₍₁₀₀₎	–	–
	Б	12cRSS	9	–	–	–	2 ₍₂₂₎	6 ₍₆₇₎	1 ₍₁₁₎	–
		23cRSS	16	–	–	–	–	4 ₍₂₅₎	10 ₍₆₃₎	2 ₍₁₂₎

Примечание. НРТ – неавтономные ретротранспозоны; ЭР/LTR-PT – эндогенные ретровирусы и LTR-ретротранспозоны; ДТР – ДНК-транспозоны; ПП – простые повторы; НТП – неклассифицированные типы повторов; в индексе цифры в скобках означают проценты.

Bai3, *Large*, *LOC629706*, *Plcb4*, *Tcf4*, *Tnpo3* и *Prkg1*, *Ptprn2*, *Slc24a2*, *Trhde* теоретически они могут опосредовать образование делеций и инверсий экзон. Исследовав при помощи программы BLASTN участки, примыкающие к гептамерам описываемой группы мотивов, мы установили, что две 12cRSS и одна 23cRSS, расположенные в межгенном пространстве, а также 23cRSS псевдогена *LOC50777* являются частью ранее не описанных дупликаций *V*-сегментов генов *Igk* и *Tra*. Это утверждение сделано нами на том основании, что 79–90-нуклеотидные фрагменты анализируемых участков имеют 84–92 %-ю гомологию с последовательностями *V*-сегментов из базы данных igSeqNt. Поэтому так же, как и dRSS, такие мотивы (далее обозначены d’RSS) следует рассматривать в качестве нативных сигнальных последовательностей

рекомбинации. В свою очередь, 158 (52 %) cRSS обнаружено нами в структуре 60 типов повторов (B1_Mur2, B1_Mur3, B1_Mur4, B1_Mus1, B1_Mus2, B3, B4, ETnERV, Lx5, Lx6, Lx7, Lx8, Lx9, MTC, MTD, MTEa, MTEb, PB1, PB1D10, PB1D7, RLTR14, RSINE1 и др.). Важно отметить, что такие же весовые коэффициенты в среднем имеют 19 (3 %) 12tRSS и 58 (9 %) 23tRSS, обнаруженных в ПСН (всего 77 tRSS). Как упоминалось выше (см. «Материалы и методы»), количество tRSS использовано нами в качестве величины, оценивающей теоретически ожидаемое количество cRSS в геноме мыши (их появление обусловлено случайными комбинациями нуклеотидов). Поэтому, сделав соответствующие сопоставления, мы можем утверждать, что существование в геноме мыши 16 % (19/121) 12cRSS и 32 % (58/182) 23cRSS с весовыми

Таблица 4
Консенсусные последовательности 12cRSS и 23cRSS, идентифицированные в 18 семействах повторов

СП	ТЛ	12cRSS		23cRSS	
		Количество	Консенсусная последовательность	Количество	Консенсусная последовательность
AcHobo	A	0	–	1	CACCGTGCAAAGTCTAAAAAAAAAAAAAAAAAAAAACC
B1	A	924	CACAGAGAAACCCGTGTCTCAAAAAAAC	104	CACAGTGARTTCCAGGMTAGCCAGGAATACAAAAAAC
	Б	322	CACAGAGAAACCCGTGTCTCAAAAAAAC	1761	CACAGAGAAACCCGTGTCTCGAAAAACAAAAACAAAAACC
B2	A	31	CACAGTGTGAGTGTAGGAACAAAAAAC	12	CACAGTGTACTCAAMAACAAAAACAAAAACAAAAAAC
	Б	9	CACAGTGAAACAAAAAAMAAAAACAAAAAAC	15	CACAGTGAAATAAAAAAAAAAAAAAHAAAAACAAAAAAC
B4	A	48	CACAGTGAAAGAGCSTAACACAAAAAAC	63	CACAGTGCCTGGGTTCCATCCCCAGCACCACAAAAAAC
	Б	24	CACAGTGAAAAAAAAACAAACAAAAAAC	42	CACAGTGAATCATGTAACAACAAAAAAMAAAAACAAAAAAC
CR1	A	1	CACACTGAAATCACAGTCAACTAAAACC	0	–
ERV1	A	5	CACAGTGAGRGWKAARMAACAAAAAAC	7	CACAGTSAGTMTHTATMCAGCAAAACAAAAACAAAAAAC
	Б	4	CACAGTGATTCAWCWGTACASAAAAAGC	2	CACAGTSYTRRYMCMRARMWWRKSMAGMRGAMARAAACM
ERVK	A	22	CACAGTGTCAAGMAATWWWACAAAAAAC	62	CACAGTGGTTGTTGTTGATCTTGTGGAAAAACACAAACC
	Б	4	CACAGTGGRNCANRAMAAAAACAAAAATC	10	CACAGTGATWAATWWCAAAATACATAAGGTACAAAAAAC
ERVL	A	4	CACAGTGNRNACAAAWTNMACMAAAACC	9	CACAGTGAKAMMTWACTGGACCHAAWMGAGACAAAAAAC
	Б	2	CACAGTGTTWAWMWMRRCKWCAGAAACC	4	CACAKTGNAGTAGCCWAGTTTGWAAAGCNRACRAAAACC
ID	A	0	–	2	CACAGWGCMCYRKGCTKKRTCCCYARMAYYWCAAAAAASY
	Б	3	CACAGTGAADAAAAHADGAACAAAAAAC	6	CACAGTGCCCTGGGTTCAATCCCCARCAACAAAAAAC
L1	A	163	CACAGTGAAAAAAAAAAAAAACAAAAAAC	189	CACAGTGAAACATGGATACAACAAAAAAACAAAAAAC
	Б	27	CACAGTGGCTYTAWAAGAAACAAAAAAC	36	CACAGTGWAAATAWAAAAAAAAAAAAAAAAAAAAACAAAAAAC
L2	A	6	CACAGTGRGTGMGAMASAVACAAAAAAC	1	CACAGACTAAAAATCCAGTGGGGAAGACAGACAAAAAAC
	Б	0	–	4	CACAGTGCNAAAGCANNTATYGGNNNAWAMACAAAAAMC
MaLR	A	319	CACAGTGTCTGTTCACAGCACTAAAACC	143	CACAGTGAGGAAATACTGGACCAAAGCAAGACAAAAAAC
	Б	41	CACAGTGATAAAAAACATAACAAAAAAC	64	CACAGTAATGTTAACTTACACCATAACCTACAAAAACA
Mariner	A	0	–	1	CACATTTCTTTCTGCCAATGTCTTTCTGTAACAAAAAAC
MER1	A	6	CACAGTGWYAAVCCCKARACAAAAAAC	13	CACAGTGDWGAAWAAAGTCAWTAAGRAAAAAACAAAAAAC
	Б	3	CACAGBGADACGTVCCDAACAAAAAAC	5	CACAGTGGCAGAAACATWGTWYAAACMATACAAAAAAC
MER2	A	2	CACAGTCYKMAAAWAWRAAMARAAACC	2	CACASYGTRWTWACWKYWKMSWMMKRWSKASAAAAAYC
MIR	A	1	CACAGTGAAGAGTACTTAGACTAAAACA	7	CACAGAGHYTGTATYAHCTAATTTAAWTAYACAAAAAAC
	Б	4	CACAGTGTGNAATATCNTAACAAAAAAC	7	CACAGTGHTGKYHTCTTTCTAMTTCATCTCACAAAAAAC
MuDR	A	0	–	1	CACAGTGATTTTCGAAGAACTAGGCTGCCTACATAAATC
Tip100	A	1	CACAGGTGCCTTTACTTACACAAAAAAC	0	–
ПП	A	9	CACAGAGAKACACATACASACAAAAACA	3	CACAGTGCAHVACACACATATGVATACACACACAAAAACA
	Б	43	CACAGTGAAAAAAAAAAAAAACAAAAAAC	64	CACAGTGMAAAAACAAAAAAACAAAAAACAAAAAAC

Примечание. СП – семейства повторов; ТЛ – типы локализации cRSS в повторах; ПП – простые повторы; Y, R, W, S, K, M, D, V, H, B, N обозначены позиции консенсусной последовательности, в которых с одинаково максимальной частотой встречаются нуклеотиды C/T, A/G, A/T, G/C, T/G, C/A, A/T/G, A/G/C, A/T/C, T/G/C, A/T/G/C соответственно.

коэффициентами, равными 0,80–0,95 и 0,75–0,88, является случайным. Возникает вопрос: в каких именно участках генома (генах, повторах, межгенном пространстве) расположены такие структуры?

Наиболее часто в структуре cRSS доминируют редкие в последовательностях fRSS нуклеотиды. Такие мотивы имеют более низкие значения W (0,65–0,74). На их долю приходится 2380 (78 %) 12cRSS и 3249 (88 %) 23cRSS. Однако подобный нуклеотидный состав имеют 30 (15 %) 12fRSS и 20 (19 %) 23fRSS. Поэтому мы считаем, что cRSS этой группы будут эффективно взаимодействовать с белками RAG1/2. К тому же в их гептамерах и наномерах присутствуют все функционально важные нуклеотиды, тогда как изменения в структуре спейсеров не являются настолько критичными, чтобы исключить саму возможность эндонуклеазного расщепления ДНК [17, 18]. Тем не менее, мы предполагаем, что рассматриваемые взаимодействия все же будут менее эффективными, чем взаимодействия указанных белков с большей частью fRSS и с cRSS, имеющими более высокие значения весовых коэффициентов.

Подробные результаты анализа нуклеотидного состава cRSS, расположенных в разных классах повторов, представлены в табл. 3. Наибольшие значения весовых коэффициентов, равные 0,85–0,88 и 0,90, имеют пять 12cRSS повторов – MTD, MLT1A1, ORR1B2, MTC и ORR1A4 семейства MaLR (LTR-ретротранспозоны).

Следует отметить, что в одних и тех же типах повторов последовательности cRSS могут сильно отличаться друг от друга. Исключения составляют повторы семейства B1: обнаруженные в них 51 % 12cRSS и 11 % 23cRSS имеют структуры CACAGAGAAACCCTGTCTCAAAAAAACC и CACAGAGAAACCCTGTCTCGAAAAACAACAAAACC (весовые коэффициенты равны 0,71 и 0,73 соответственно). В табл. 4 приведены консенсусные последовательности 12cRSS и 23cRSS, идентифицированных нами в разных семействах повторов.

Выводы. В результате проведенной работы установлено, что в геноме мыши происходит эволюционное накопление возможных сайтов – мишеней белков RAG1/2, у которых гептамеры и нано-

меры полностью соответствуют структуре CACAGTG/ACAAAAACC или имеют с ней 14–15 общих нуклеотидов из 16 возможных. В 71 % случаев cRSS являются структурными элементами 390 типов повторов. В структуре около 5 % мотивов можно обнаружить нуклеотиды, типичные для большинства сигнальных последовательностей рекомбинации функциональных V , D , J -сегментов Ig - и Tcr -генов мыши (fRSS). Существование 25 % из них в ДНК исследуемого вида следует рассматривать как следствие случайных комбинаций нуклеотидов. В большинстве случаев структуры спейсерных участков анализируемых 12cRSS и 23cRSS имеют 58–67 и 30–47 % гомологии с аналогичными структурами fRSS соответственно. Эти результаты рассматриваются нами в качестве основных характеристик изучаемой группы мотивов.

A. Yu. Gubsky, V. G. Zinkovsky

Structural analysis of possible target-sites of RAG1/2 protein, discovered in mouse genome *in silico*, and their identification in repeating elements

Summary

We have established that the quantity of possible target-sites of RAG1/2 (cRSS) is 5.4 times higher than theoretically expected one using mathematical methods and specially developed algorithms. 71% of cases revealed cRSS in the structure of 390 types of repeats. The structure of 5% motives includes nucleotides, typical for the majority of signal sequences of recombination of functional V, D, J segments of Ig, Tcr mouse genes (fRSS). The existence of 25 % of such motives in mouse DNA may be considered as the consequence of random nucleotide combinations. In the majority of cases the structures of spacers 12cRSS and 23cRSS are 58–67 % and 30–47 % homologous to spacers 12fRSS and 23fRSS respectively.

Keywords: cRSS, V(D)J-recombination, RAG1, RAG2.

А. Ю. Губський, В. Г. Зінковський

Структурний аналіз групи можливих ДНК-мішеней білків RAG1/2, виявлених у геномі миші *in silico*, та їхня ідентифікація у відомих типах повторюваних елементів

Резюме

З використанням математичних методів аналізу, а також спеціально розроблених алгоритмів встановлено, що кількість виявлених у геномі миші можливих сайтів – мішеней білків RAG1/2 в 5,4 рази перевищує теоретично очікуване число. У 71 % випадків cRSS є структурними елементами 390 типів повторів. У структурі близько 5 % мотивів виявлено нуклеотиди, типові для більшості сигнальних послідовностей рекомбінації функціональних V, D, J-сегментів Ig- і Tcr-генів миші (fRSS). Існування 25 % з них у ДНК досліджуваного виду потрібно роз-

глядати як наслідок випадкових комбінацій нуклеотидів. Структури спейсерних ділянок аналізованих 12cRSS і 23cRSS, як правило, мають 58–67 і 30–47 % гомології з аналогічними структурами fRSS відповідно.

Ключові слова: cRSS, V(D)J-рекомбінація, RAG1, RAG2.

СПИСОК ЛІТЕРАТУРИ

1. Tonegawa, S. Somatic generation of antibody diversity // Nature.—1983.—**302**.—P. 575–581.
2. Tonegawa S., Brack C., Hozumi N., Pirrotta V. Organization of immunoglobulin genes // Cold Spring Harbor Symp. Quant. Biol.—1978.—**42**.—P. 921–931.
3. Oettinger M., Schatz D., Gorka C., Baltimore D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination // Science.—1990.—**248**.—P. 1517–1523.
4. Aplan P. D., Lombardi D. P., Ginsberg A. M., Cossman J., Bertness V. L., Kirsch I. R. Disruption of the human SCL locus by «illegitimate» V-(D)-J recombinase activity // Science.—1990.—**256**.—P. 1426–1429.
5. Tsuji H., Ishii-Ohba H., Katsube T., Ukai H., Aizawa S., Doi M., Hioki K., Ogiu T. Involvement of illegitimate V(D)J recombination or microhomology-mediated nonhomologous end-joining in the formation of intragenic deletions of the *Notch1* gene in mouse thymic lymphomas // Cancer Res.—2004.—**15**.—P. 8882–8890.
6. Scheerer J. B., Xi L., Knapp G. W., Setzer R. W., Bigbee W. L., Fuscoe J. C. Quantification of illegitimate V(D)J recombinase mediated mutations in lymphocytes of newborns and adults // Mutat. Res.—1999.—**431**.—P. 291–303.
7. Губський А. Ю. Структурний аналіз сигнальних послідовностей рекомбінації трьох типів V-, D-, J-сегментів генів імуноглобулінів і T-клітинних рецепторів людини // Одеськ. мед. журн.—2005.—**5**.—С. 10–12.
8. Matsuda F., Ishi K., Bourvagnet P., Kumma I., Hayashida H., Miyata T., Honjo T. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus // J. Exp. Med.—1998.—**188**.—P. 2151–2162.

9. Lewis S. M., Agard E., Suh S., Czyzyk L. Cryptic signals and the fidelity of V(D)J joining // Mol. and Cell. Biol.—1997.—**17**.—P. 3125–3136.
10. Cowell L. G., Davila M., Yang K., Kepler T. B., Kelsoe G. Prospective estimation of recombination signal efficiency and identification of functional cryptic signals in the genome by statistical modeling // J. Exp. Med.—2003.—**197**.—P. 207–220.
11. Губський А. Ю., Зиньковський В. Г. Пошук і ідентифікація ділянок локалізації в геномі миші послідовностей cRSS, структура яких припускає високий рекомбінаційний потенціал // Одеськ. мед. журн.—2006.—**98**, № 6.—С. 11–14.
12. Hawley D. K., McClure W. R. Compilation and analysis of *Escherichia coli* promoter DNA sequences // Nucl. Acids Res.—1983.—**11**.—P. 2237–2255.
13. Stormo G. D. DNA binding sites: representation and discovery // Bioinformatics.—2000.—**16**.—P. 16–23.
14. Harr R., Haggstrom M., Gustafsson P. Search algorithm for pattern math analysis of nucleic acid sequences // Nucl. Acids Res.—1983.—**11**.—P. 2943–2957.
15. Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // Nucl. Acids Res.—1997.—**25**.—P. 3389–3402.
16. Jurka J. Repbase update: a database and an electronic journal of repetitive elements // Trends Genet.—2000.—**16**.—P. 418–420.
17. Fanning L., Connor A., Baetz K., Ramsden D., Wu G. E. Mouse RSS spacer sequences affect the rate of V(D)J recombination // Immunogenetics.—1996.—**40**.—P. 146–150.
18. Akamatsu Y., Tsurushita N., Nagawa F., Matsuoka M., Okazaki K., Imai M., Sakano H. Essential residues in V(D)J recombination signals // J. Immunol.—1994.—**153**.—P. 4520–4529.

УДК 575.113 + 577 .2 + 599.89
Надійшла до редакції 02.10.07