

И. Б. Рогозин, Н. А. Колчанов, В. В. Соловьев, Н. Е. Среднева

## КОМПЬЮТЕРНАЯ СИСТЕМА ИССЛЕДОВАНИЯ РОЛИ ПОЛИНУКЛЕОТИДНОГО КОНТЕКСТА В ВОЗНИКНОВЕНИИ ТОЧКОВЫХ МУТАЦИЙ

*В настоящее время имеется много свидетельств в пользу важной роли полинуклеотидного контекста (конкретных вариантов полинуклеотидных последовательностей) в возникновении мутаций. Однако молекулярные механизмы, посредством которых полинуклеотидный контекст влияет на возникновение мутаций, до настоящего времени остаются малоисследованными. Появление эффективных методов секвенирования привело к быстрому накоплению данных о локализации мутаций в последовательностях ДНК. Нами разработана компьютерная система для анализа особенностей полинуклеотидного контекста, влияющих на возникновение мутаций. В настоящее время она состоит из двух частей: банка данных мутационных событий и комплекса методов анализа связи особенностей контекста с возникновением одиночных и множественных замен.*

**Введение.** Представление о пассивной роли ДНК как мишени для возникновения мутаций сменяется пониманием большого значения полинуклеотидной последовательности в определении типа мутации, их локализации и частоты. Экспериментальные данные показывают, что в определенных участках ДНК — «горячих точках» — мутации возникают существенно чаще, чем в других участках. Это свидетельствует о важной роли полинуклеотидного контекста (конкретных вариантов полинуклеотидных последовательностей) в возникновении мутаций [1—7].

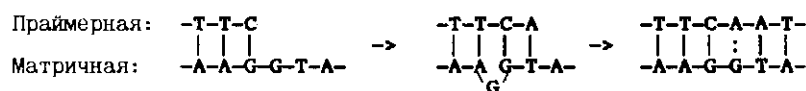


Рис. 1. Схема возникновения предмутационного состояния (отмечено двоеточием) в соответствии с моделью дислокации матричной цепи [5]

Fig. 1. Scheme of the origin of the state before mutational beginning (marked by the colon) according to the model of the matrix chain dislocation [5].

Появление эффективных методов секвенирования привело к быстрому накоплению данных о локализации мутаций в последовательностях ДНК. Однако механизмы влияния полинуклеотидного контекста на возникновение мутаций остаются малоисследованными. В связи с этим важное значение приобретает разработка компьютерных методов для анализа особенностей полинуклеотидного контекста, влияющих на возникновение мутаций. Ниже дано описание разрабатываемой нами компьютерной системы для исследования роли полинуклеотидного контекста в возникновении одиночных и множественных нуклеотидных замен. Рассмотрим возможные механизмы их возникновения.

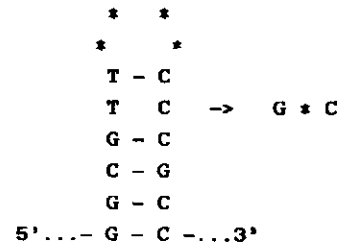
Таутомеризация нуклеотидов [8]. Нуклеотиды помимо наиболее вероятной конформации (в виде комплементарных пар А—Т и G—C) способны к таутомеризации и переходу в такие конформации, при которых они могут формировать неканонические комплементарные пары. Если в момент репликации в родительской цепи ДНК произойдет переход нуклеотида в таутомерную конформацию, то в дочерней цепи ДНК возникнет некомплементарная пара, т. е. предмутационное состояние, способное в дальнейшем вызывать замену нуклеотида в данной позиции.

Дислокация матричной цепи [5]. При репликации ДНК в результате сдвига матричной цепи относительно праймерной и присоединения следующего нуклеотида к праймерной цепи в ДНК образуется некомплементарная пара, которая может привести к возникновению мутации в последовательности (рис. 1) [5].

Генная конверсия [9]. Это нереципирующий перенос информации от одной нити ДНК к другой. Генная конверсия может появляться при формировании несовершенного гетеродуплекса из нитей ДНК, принадлежащих двум гомологичным удаленным участкам генома. Последующая репарация корректирует одну нить ДНК по другой, обеспечивая перенос генетической информации от одного участка ДНК к другому, проявляющийся в возникновении группы мутаций.

Рис. 2. Предполагаемый механизм возникновения мутации Т→G в 188-й позиции гена *LACBSAC*, описанного на рис. 3.  $P(W_{\text{реал}} < W_{\text{случ}}) = 0,13$

Fig. 2. Supposed mechanism of the origin of the mutation T→G in the position 188 of the gene *LACBSAC* described in the Fig. 3.



Репарационная коррекция несовершенных гетеродуплексов ДНК [1, 2]. Эти гетеродуплексы возникают при неправильном спаривании близкорасположенных взаимокплементарных нитей коротких прямых повторов или комплементарных палиндромов в ДНК (рис. 2).

Индукцированный мутагенез. В основе мишенного мутагенеза лежат прямые взаимодействия мутагенных факторов с ДНК, приводящие к химической модификации оснований. При немшенном мутагенезе происходит химическая модификация пула нуклеотидов или модификация ферментов репликации и репарации ДНК.

Мутации как результату некоего процесса предшествует предмутацонное состояние, т. е. повреждение структуры ДНК, которое с определенной вероятностью может привести к образованию мутации. Становление мутаций — сложный многостадийный процесс. Это затрудняет идентификацию механизмов влияния полинуклеотидного контекста на возникновение мутаций.

**Описание базы данных и комплекс программ анализа полинуклеотидного контекста.** База данных мутационных событий содержит описание мутаций в молекулах ДНК с известными полинуклеотидными последовательностями. Используемый формат данных (рис. 3) включает ряд информационных полей, содержащих библиографическую информацию, нуклеотидную последовательность, локализацию мутаций, вид нуклеотидных замен, качественную характеристику мутаций (спонтанные, индуцированные и т. д.); для индуцированных мутаций — название мутагенного фактора.

Комплекс программ анализа полинуклеотидного контекста предназначен для выявления особенностей полинуклеотидного контекста, связанных с возникновением мутаций. Входной информацией для программ являются полинуклеотидные последовательности с указанием локализации и вида наблюдавшихся в них мутаций. При анализе контекста используется расширенное описание нуклеотидов в 15-буквенном алфавите: A, T, G, C, R=(G, A); Y=(T, C); M=(A, C); K=(G, T); S=(G, C); W=(A, T); H=(A, C, T); B=(G, T, C); V=(G, C, A); D=(G, A, T); N=(A, T, G, C).

Анализ распределения мутаций по типам нуклеотидных позиций. Исследуется вектор  $F(4)$  (где  $F(1)$  — частота мутирования нуклеотида A;  $F(2)$  — T;  $F(3)$  — G;  $F(4)$  — C). Для анализа используется критерий, основанный на биномиальном распределении. Пусть общее число мутаций равно  $n$ , число мутаций нуклеотида типа  $i$  равно  $k_i$ . Причем частота мутирования нуклеотида  $F(i) = k_i/n$  больше 0,25 (ожидаемой при равномерном распределении мутаций по типам нуклеотидов). Вероятность такого события по случайным

причинам равна

$$P(k) = \sum_{j=k}^n P(n, j, 0,25) = \sum_{j=k}^n C_n^j \cdot (0,25)^j (1 - (0,25))^{n-j}. \quad (1)$$

Малое значение ( $P < 0,01$ ) свидетельствует о том, что для исследуемого мутационного процесса характерно неслучайно частое мутирова-

Идентификатор последовательности: LACBSAC			
Тип последовательности: ДНК			
Авторы: Т.А. Kunkel			
Название: The mutational specificity of DNA polymerases- $\alpha$ and $\gamma$ during in vitro DNA synthesis			
Ссылка: The J. of Biol. Chemistry			
Классификация объекта: in vitro			
Ключевые слова: eucariotic DNA polymerase, point mutation, mispair			
Тип мутаций: single base substitution, deletion			
Предполагаемый механизм возникновения мутаций: misinsertion of nucleotides, model of dislocation mutagenesis			
Способ возникновения мутаций: spontaneous mutations			
Описание особенностей эксперимента: The base substitutions errors to be analysed were produced during in vitro DNA synthesis of the lacZa gene in M13mp2 DNA by purified DNA polymerases $\alpha$ and $\gamma$ , in reactions containing all four dNTP substrates at equimolar concentrations. The assay measures errors which result in a decrease in $\alpha$ -complementation, identified as M13mp2 plaques having light blue color.			
Разметка функциональных зон:			
репт	127.-	0.	lacZa gene
Локализация мутаций:			
substitution		number of	
position	transition	mutations	
19	G $\rightarrow$ T	1	
28	C $\rightarrow$ T	7	
28	C $\rightarrow$ G	2	и т.д.
последовательность:			
GCGCAACGCAATTAATGTGAGTTAG... и т.д.			

Рис. 3. Пример описания последовательности [10] в банке мутационных событий. В дальнейшем на примере этой последовательности будут иллюстрироваться методы, представленные в настоящей работе

Fig. 3. Example of the sequence description in the bank of the mutational events. Later, methods described in the present work will be illustrated on the basis of this example.

ние нуклеотида типа  $i$ . При анализе распределения мутаций по типам нуклеотидов может использоваться также расширенный алфавит нуклеотидов. Пример использования этого подхода приведен в табл. 1.

Анализ мутационных переходов. Полезную информацию о молекулярных механизмах мутагенеза может дать анализ мутационных переходов  $i \rightarrow j$  (где  $i$  — исходный тип нуклеотида;  $j$  — тип нуклеотида, возникшего в результате мутации). Анализируется матрица частот мутационных переходов  $T(4 \times 4)$  (где  $T(i, j)$  — частота мутационных переходов из нуклеотидов  $i$ -го в нуклеотиды  $j$ -го типа). Пусть  $k_{ij}$  — число переходов из  $i$ -го в  $j$ -й нуклеотид,  $n$  — общее число мутаций. Пусть частота мутационного перехода  $T(i, j) = k_{ij}/n$  превышает

1/12, ожидаемую при равномерном распределении мутаций по возможным типам переходов. Для оценки неслучайности превышения наблюдаемого числа переходов над ожидаемым используется критерий (1) биномиального распределения  $P(k_{ij})$ .

Малое значение  $P(k_{ij}) < 0,01$  показывает, что для исследуемых мутаций характерен неслучайно частый переход из  $i$ -го в  $j$ -й тип нуклеотида. Так, анализ мутаций в гене *LACBSAC* выявил следующее количество неслучайных типов переходов:  $C \rightarrow T$  — 57, ожидаем 9,  $P(k_{ct}) < 10^{-5}$ ;  $G \rightarrow Y$  — 40, ожидаем 19,  $P(k_{gy}) < 10^{-5}$ . При анализе мутационных переходов может использоваться расширенный алфавит нуклеотидов.

Таблица 1

Частота возникновения мутаций в нуклеотидах определенного типа в последовательности *LACBSAC*

Frequency of the mutational origin in the nucleotides of the definite type in the sequences *LACBSAC*

Тип нуклеотида	A	T	G	C	S
Частота мутирования	0,05	0,06	0,37	0,52	0,88
$P(k)$	—	—	$4 \cdot 10^{-3}$	$10^{-5}$	$10^{-5}$

Выявление «горячих точек» мутирования. Использован подход, приведенный в работе [11]. Распределение мутаций по нуклеотидным позициям оценивалось исходя из распределения Пуассона. Пусть среднее количество мутаций в позициях определенного типа равно  $m$ . Для группы позиций, в которых произошло  $i$  мутаций, рассчитывается ожидаемое число таких позиций в последовательности по формуле

$$R(i) = S \cdot (m^i / i!) \cdot e^{-m}, \quad (2)$$

где  $S$  — общее число позиций определенного типа в исследуемой последовательности. Если наблюдаемое число таких сайтов превышает их ожидаемое число, то указанная группа является «горячими точками» мутирования.

Оценка кластеризации мутаций. Для оценки ожидаемого расстояния между случайно разбросанными мутациями мы использовали предложенную в работе [12] оценку. Доля мутаций, находящихся на расстоянии не более  $r$  оснований друг от друга, вычисляется по формуле

$$K(r) = \sum_{i=1}^r f \cdot (1 - f)^{i-1}, \quad (3)$$

где  $f$  — средняя частота сайтов мутирования в пересчете на один нуклеотид (вычисленная для определенной последовательности). Если наблюдаемые расстояния между сайтами возникновения мутаций отличаются от ожидаемых (в соответствии с критерием  $\chi^2$ ), то считается, что в последовательности наблюдается кластеризация сайтов мутирования.

Анализ модели репарационной коррекции. Для проверки различных гипотез о механизмах возникновения мутаций в молекулах ДНК нами разработан метод статистических весов. Этот метод является эффективным «инструментом» для изучения роли полинуклеотидного контекста в процессах возникновения мутаций. На основе выбранной модели механизма возникновения мутаций строится функция  $W(x_1, x_2, \dots, x_N)$ , отражающая зависимость локализации мутационных событий от факторов ( $x_i$ ) полинуклеотидного контекста, ис-

пользующихся в модели. Такими факторами могут быть выявленные в последовательности структуры (повторы различных типов, сайты, политракты и т. д.), которые, как предполагается в рассматриваемой модели, обеспечивают процессы возникновения мутаций или косвенно влияют на них. Для  $i$ -й мутации в исследуемой последовательности вычисляется значение построенной функции  $W^i_{\text{реал}} = W(x_1, x_2, \dots, x_N)$ . Полученные веса усредняют по всем мутациям, в результате чего получается статвес группы мутаций  $W_{\text{реал}}$ . Затем на ЭВМ с помощью генератора случайных чисел моделируются случайные мутации в исследуемой последовательности, расположение которых *a priori* не зависит от рассматриваемых факторов контекста. Для исследуемой последовательности строится значение  $W_{\text{случ}}$  при случайном «разбрасывании» 1000 групп мутаций.



Рис. 4. Возможный способ возникновения мутации Т→С (позиция 49) в гене *LACBSAC* на основе дислокации матричной цепи в ходе репликации;  $P(W < W) = 0,08$   
 Fig. 4. Possible way of the origin of the mutation T→C (position number 49) at the gene *LACBSAC* on the basis of the dislocation of matrix chain during replication.

Каждая группа содержит число случайных мутаций, равное наблюдаемому. На основе распределения  $W_{\text{случ}}$  оценивается вероятность  $P(W_{\text{реал}} < W_{\text{случ}})$  для группы случайных мутаций иметь значение  $W_{\text{реал}}$  или большее. Если вероятность  $P(W_{\text{реал}} < W_{\text{случ}}) < Q$  (где  $Q$  — малое число), то выявляется достоверная связь данных мутаций с рассматриваемыми факторами в последовательности. Аналогично исследуются конкретные (индивидуальные) мутации, когда вместо группы мутаций рассматривается только одна.

Модель репарационной коррекции предполагает наличие в районе мутирования прямого повтора или комплементарного палиндрома. Функция веса имеет вид:  $W = (L - K) / R$ , где  $L$  — длина повтора, репарационная коррекция которого могла бы вызвать возникновение данной мутации;  $K$  — число несовпадений между участками повтора;  $R$  — расстояние между участками повтора. Для каждой мутации учитывали повтор с максимальным статвесом. Пример результатов анализа приведен на рис. 2.

Анализ модели дислокации матричной цепи в ходе репликации. Эта модель предполагает наличие в районе мутирования политракта (рис. 1). Связь политрактов с позициями мутаций оценивается с помощью метода статвесов (с использованием статвеса  $W = L$ , где  $L$  — длина политракта). Пример анализа приведен на рис. 4 для мутации Т→С в 49-й позиции последовательности гена *LACBSAC*, описанного в реферативной карте. В данном случае выявлена связь положения этой мутации с политрактом ТТТ, указывающая на возможность ее возникновения на основе дислокации матричной цепи в ходе репликации, как это показано на рис. 4.

Анализ модели генной конверсии. В настоящее время эта схема возникновения мутаций исследуется на основе двух программ. С помощью первой из них анализируется характер распределения кластера мутаций при сравнении предполагаемых донорских и акцепторных последовательностей с использованием метода статвесов. Статвес для группы мутаций (рис. 5) вычисляется следующим образом:  $W = L_2 - L_1$ , где  $L_2$  и  $L_1$  — правая и левая границы расположения кластера в полинуклеотидной последовательности. Результаты, полученные на основе этой программы, показаны на рис. 5, где приведен пример выявления генной конверсии между *V*-геном и псевдогеном цыпленка. Анализ показал, что вероятность наблюдать такой кластер мутаций по случайным причинам —  $P(W_{\text{реал}} < W_{\text{случ}}) < 0,01$ . Это явно свидетельствует в пользу генной конверсии как возможного механизма

возникновения мутаций в такой последовательности. Вторая программа способствует выявлению консервативных районов, характерных для геной конверсии, при сравнении группы последовательностей, в которых возникли мутации.

	<b>L1</b>	<b>L2</b>
<b>1</b>	<b>AAATCGGGCTCCACAGCCACATTAACCATCACTGGGGTCCGAGCCGACGACAAT</b>	
<b>2</b>	<b>CTCtccgTATTcaCaCAcaccattaaccatcactggggtccAagccgaGgaTGaG</b>	
<b>3</b>	<b>CTC    TA T    CA</b>	<b>A    G TG G</b>

Рис. 5. Пример выявления геной конверсии между  $V_{\lambda}$ -геном (1) и псевдогеном (2) [13]; внизу приведены нуклеотидные замены (3);  $P(W_{\text{случ}} < W_{\text{реал}}) < 0,01$ . Позиции предполагаемого конверсионного тракта  $L1=65$ ,  $L2=122$

Fig. 5. An example of revealing gene conversion between  $V_{\lambda}$  gene (1) and pseudogene (2) [13]; nucleotide substitutions are shown below. Position of proposed conversion tract is 65 for L1 and 122 for L2.

Анализ консенсусов сайтов мутирования. Участки полинуклеотидного контекста, окружающие сайты мутирования, могут быть представлены в виде набора выравненных последовательностей. Для этого набора последовательностей может быть осуществлен поиск

Позиция	Последовательность сайта	Число мутаций
<b>28</b>	<b>GCT <u>C</u> ACT</b>	<b>9</b>
<b>48</b>	<b>AGG <u>C</u> TTT</b>	<b>4</b>
<b>53</b>	<b>TTA <u>C</u> ACT</b>	<b>6</b>
<b>159</b>	<b>TTA <u>C</u> ACC</b>	<b>6</b>
<b>165</b>	<b>CGT <u>C</u> GTG</b>	<b>4</b>
<b>192</b>	<b>ACC <u>C</u> AAC</b>	<b>5</b>
<b>213</b>	<b>GCA <u>C</u> ATC</b>	<b>4</b>
<b>217</b>	<b>ATC <u>C</u> CCC</b>	<b>4</b>
<b>225</b>	<b>TTC <u>G</u> CCA</b>	<b>6</b>
<b>232</b>	<b>GCT <u>G</u> GCG</b>	<b>4</b>
<b>233</b>	<b>CTG <u>G</u> CGT</b>	<b>9</b>
<b>234</b>	<b>TGG <u>C</u> GTA</b>	<b>4</b>
<b>243</b>	<b>AGC <u>G</u> AAG</b>	<b>9</b>
<b>250</b>	<b>AGG <u>C</u> CCG</b>	<b>4</b>

Динуклеотид - **WK C R**  
 $P(k) =$         **0,0015 0,0003**

Рис. 6. Анализ динуклеотидов в выборке горячих точек мутирования в гене *LACBSAC*. Выявленные консервативные динуклеотиды подчеркнуты в последовательностях выборки  
 Fig. 6. Analysis of the dinucleotides in the set of hot mutational points in the gene *LACBSAC*. Revealed conservative dinucleotides are underlined in the set consequences.

консенсуса (в расширенном 15-буквенном алфавите). Например, при построении статистического консенсуса для набора «горячих точек» в каждой позиции набора выравненных последовательностей поочередно для всех 15 типов нуклеотидов оценивается их встречаемость.

Пусть в исследуемом наборе из  $n$  последовательностей (на рис. 6  $n=14$ ) в  $j$ -й позиции рассматривается основание типа  $i$ . Оно встретилось в этой позиции  $k_i$  раз.

Вероятность такого события оценивается из биномиального распределения  $P(k_i) = \sum_{l=k_i}^n P(n, l, F(i))$ , где  $F(i)$  — частота основания  $i$  в исследуемой последовательности (в которой произошли мутации). Для

позиции  $j$  выбирается такой тип основания, для которого характерно минимальное значение  $P(k_i)$ , при условии, что  $P(k_i) < 0,01$ .

Статистический консенсус может быть построен также для динуклеотидов. Рассматриваются две соседние позиции  $j$  и  $j+1$  (нуклеотиды рассматриваются в 15-буквенном алфавите). Для определенного нуклеотида  $(i1, i2)$  подсчитывается число  $k_{i1,i2}$  его встречаемости в  $j$ -й и  $(j+1)$ -й позициях. Частота встречаемости динуклеотида  $(i1, i2)$  в исследуемой последовательности  $F(i1, i2)$ . Из соотношения  $P(k_{i1,i2}) = \sum_{l=k_{i1,i2}} P(n, l, F(i1,$

$i2))$  вычисляется вероятность наблюдения  $k_{i1,i2}$  пар типа  $(i1, i2)$  по случайным причинам (см. формулу (1)). Как и выше, отбирается пара с минимальным  $P(k_{i1,i2})$ , если  $P(k_{i1,i2}) < 0,01$  (см. рис. 6). Аналогично статистический консенсус строится для тринуклеотидов, а также для комбинаций нуклеотидов в позициях, разделенных пространственно. Полученная информация в дальнейшем может подвергаться дополнительной статистической обработке.

Анализ связи мутаций с консенсусом. С помощью метода статтесов нами оценивается неслучайность связи мутаций в исследуемой последовательности с консенсусом (полученным ранее для мутаций в этой или другой последовательности). Если для определенной позиции  $j$  характерно наличие данного консенсуса, то ей присваивается статтес, равный количеству мутаций в этой позиции:

$$W = m. \quad (4)$$

**Исследование роли полинуклеотидного контекста в возникновении соматических мутаций в иммуноглобулиновых генах.** Рассмотрим возможности описанного выше комплекса программ на примере исследования соматических мутаций в иммуноглобулиновых генах. Соматические мутации в  $V$ -генах в большинстве представлены нуклеотидными заменами. Они возникают с высокой частотой в  $V$ -районах ( $V$ -сегменты генов иммуноглобулинов и прилегающие к ним фланкирующие районы) и не наблюдаются в других районах. Молекулярные механизмы этого явления интенсивно исследуются, но многие особенности остаются неясными.

Для анализа была взята выборка из 14  $V$ -генов, где выявлены 164 соматические мутации (табл. 2) (см. обзор [14]). При исследовании модели репарационной коррекции показано, что в 11 из 14  $V$ -генов зна-

Таблица 2

Результаты анализа  $V$ -генов иммуноглобулинов  
The results of analysis of immunoglobulin  $V$ -genes

Краткое наименование	Число мутаций	$P(W < W_{сл})$ для модели репарационной коррекции	$P(W < W_{сл})$ для встречаемости консенсуса RSY	$P(W < W_{сл})$ для модели дислокации матричной цепи
VM1	3	0,11	0,05	—
VM2	6	0,13	0,10	0,11
VM3	7	0,07	0,16	—
VM4	4	0,12	0,34	—
VM5	13	0,07	0,35	0,29
VM6	3	0,18	0,04	—
VM7	10	0,35	0,13	0,52
VM8	13	0,45	0,30	0,36
VM9	10	0,13	0,44	0,49
VM10	10	0,01	0,01	0,27
VM11	9	0,08	0,01	0,47
VM12	10	0,63	0,11	0,17
VM13	42	0,24	0,01	0,29
VM14	24	0,21	0,01	0,87

Примечание. Черточка обозначает отсутствие мутаций, возникновение которых могло объясняться исследуемым механизмом.

чение  $P(W_{\text{реал}} < W_{\text{случ}})$  меньше 0,25. Вероятность данного события по случайным причинам в соответствии с критерием биномиального распределения равна  $3,7 \cdot 10^{-5}$ . Столь низкое ее значение является весо-мым аргументом в пользу подобного механизма возникновения соматических мутаций. Однако анализ индивидуальных мутаций показал, что этот механизм не объясняет возникновения всех наблюдаемых в таких генах соматических мутаций [14].

Анализ контекста, окружающего точки, для которых наблюдалось наибольшее число соматических мутаций в последовательностях VM13 и VM14, выявил новую закономерность. Подавляющая часть «горячих точек» мутирования имела консенсус RSY, при этом сайт мутирования характеризовался наличием нуклеотида S. Результаты анализа значимости консенсуса RSY в 14 V-генах приведены в табл. 2 с использованием статвеса, вычисленного из соотношения (4) (рис. 7). Можно ви-

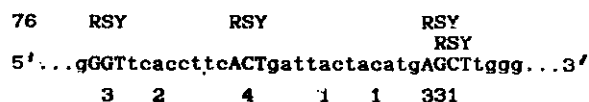


Рис. 7. Часть последовательности V-гена, в котором происходили соматические мутации (их число указано цифрами под позицией, в которой они наблюдались) [15]. Заглавными буквами выделены сайты мутирования, в которых обнаружен консенсус RSY

Fig. 7. A part of a sequence of V-gene, in which the somatic mutations were observed (their number is indicated under position in which they were observed). Mutational sites in which consensus RSY is found are marked by the capital letters.

деть, что для 10 из 14 V-генов значение  $P(W_{\text{реал}} < W_{\text{случ}})$  меньше 0,25. Значение вероятности такого события, вычисленное по биномиальному распределению и равное  $3,4 \cdot 10^{-4}$ , показывает, что контекст RSY связан с возникновением соматических мутаций в исследованных генах. Интересно, что анализ RSY-сайта для мутаций в последовательности гена *c-myc* [16] (встроенного в кластер иммуноглобулиновых генов) также выявил явную связь мутаций с этим контекстом ( $P(W_{\text{реал}} < W_{\text{случ}}) < 0,01$ ).

Влияние соседних оснований на частоту возникновения мутаций отмечалось и ранее для прокариотических ДНК-полимераз. Авторы [17] предполагают, что соседние основания слева и справа могут иметь большое значение для стабилизации неканонической пары, возникающей в ходе репликации. Похожие закономерности выявлены при анализе спонтанных мутаций, возникающих при трансфекции вектора *pZ189* в клетки обезьяны. Для всех мутантных позиций оказалось характерным наличие слева вполне определенных нуклеотидов. Консенсус имеет вид YC, где нуклеотид C — позиция мутирования. Предполагается, что эти мутации возникли в ходе репарации трансфицированной ДНК, поврежденной клеточными нуклеазами [17].

Анализ механизма дислокации праймерной цепи не выявил его связи с возникновением мутаций в исследованных V-генах (см. табл. 2).

Анализ механизма межгенной конверсии проводили для последовательности V-гена цыпленка. Ранее данные в пользу межгенной конверсии были обнаружены при сравнении нуклеотидной последовательности этого гена и 25 его псевдогенов [13]. Проведенное изучение областей выявило следующую картину распределения соматических мутаций в этом V-гене длиной 303 н. п. Предполагаемые области конверсии образовали пять основных групп сгущения соматических мутаций. В составе районов, разделяющих группы сгущения (спейсеры), неслучайно часто встречается тракт TCAS (рис. 8). Этот тракт встретился в последовательности спейсеров 5 раз (суммарная длина спейсеров 103 нуклеотида). Всего таких трактов найдено 6. Вероятность того, что в отрезок  $P=103/303$  попали 5 трактов из 6, в соответствии с биномиальным распределением, 0,02 (формула (1)). Такое малое ее значение по-



казывает, что данный тракт может играть роль определенного сигнала при генной конверсии  $V$ -гена и его псевдогенов.

Можно предположить следующую схему возникновения соматических мутаций в генах иммуноглобулинов (см. [2]). Роль пенцилирующего фактора в возникновении соматических мутаций играет нарушение нормальной структуры хроматина в  $V$ -районах генов иммуноглобулинов, происходящее в ходе их перестройки. Освобождение ДНК от нук-

Позиция начала спейсера	Последовательность спейсера
1	GCGCTGAC TCAG CCG
50	AGA TCAC STGCTCCGGGG
103	CCTGGCAGTGCCSCTG TCAC T
155	ACATCCST TCAC GATTCTCCGGTTCC
198	CACATTAACCA TCAC TGGGGTCC
Консенсус	TCAS

Рис. 8. Результаты анализа спейсерных районов  $V_{\lambda}$ -гена цыпленка [13]. Анализ выявил консервативный тракт TCAS, который, как мы предполагаем, может быть сигналом генной конверсии

Fig. 8. Results of the analysis of the spacer regions of chicken  $V_{\lambda}$  gene [13]. The analysis revealed the conservative tract TCAS, which, as we supposed may be the signal of the gene conversion

леосом приводит к ее локальному плавлению и формированию гетеродуплексов и шпилечных структур, которые могут подвергаться репарационной коррекции. С другой стороны, такой район в значительной степени экспонирован для клеточных нуклеаз, которые могут повреждать ДНК  $V$ -районов. Дальнейшая репарация с ошибками (особенно в RSY-сайте) приводит к появлению точковых замен. Для межгенной конверсии также необходимо освобождение ДНК от нуклеосом. Относительные вклады и возможность взаимосвязи этих трех механизмов требуют дальнейшего исследования.

#### A COMPUTER SYSTEM FOR THE ANALYSIS OF A POLY-NUCLEOTIDE CONTEXT FOR THE ROLE IT PLAYS IN THE APPEARING OF POINT MUTATIONS

*I. B. Rogozin, N. A. Kolchanov, V. V. Solovyov, N. E. Sredneva*

Institute of Cytology and Genetics, Academy of Sciences  
of the USSR, Siberian Department, Novosibirsk

#### Summary

It has become a fact by now that the polynucleotide context of specified variants of the polynucleotide sequences is of profound importance as for the appearing of mutations. However, the molecular mechanisms by means of which the polynucleotide context affects the appearance of mutations are what still hangs in the balance. The efficacious methods of sequencing led to the fast piling-up of data on the mutation localization in the sequences. We have worked out a computer system which proved to promote analysing the features of a polynucleotide context that influences the mutations appearing. At present it consists of two substructures, namely, The Mutational Events Data Bank and The Complex of Methods to Analyse The Correlation Between The Context Features And The Appearing of Singular and Multiple Substitutions.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Pipley L. S.* Model for the participation of quasipalindromic DNA sequences in frame shift mutation // Proc. Nat. Acad. Sci. USA.— 1982.— 79, N 13.— P. 4128—4132.
2. *Kolchanov N. A., Solovyov V. V., Rogozin G. B.* Peculiarities of immunoglobulin gene structures as a basis for somatic mutations emergence // FEBS Lett.— 1987.— 214, N 1.— P. 87—91.

3. Burns P. A., Gordon A. J. E., Glickman B. W. Influence of neighbouring base sequence on N-methyl-N'-nitro-N-nitrosoguanidine mutagenesis in the *lacI* gene *Escherichia coli* // *J. Mol. Biol.*— 1987.— 194, N 4.— P. 385—390
4. Golding G. B., Glickman B. W. Sequence-directed mutagenesis: evidence from a phylogenetic history of human  $\alpha$ -interferon genes // *Proc. Nat. Acad. Sci. USA.*— 1985.— 82, N 12.— P. 8577—8581.
5. Drake J. W., Baltz R. H. The biochemistry of mutagenesis // *Annu. Rev. Biochem.*— 1976.— 45.— P. 14—37.
6. Индукция повторяющихся нуклеотидных последовательностей. Вероятные механизмы эволюции генома и геной конверсии / Р. И. Салганик, А. В. Мазин, Г. Л. Дианов, Л. П. Овчинникова // *Генетика.*— 1984.— 20, № 8.— С. 1244—1252.
7. Льюин Б. Гены.— М.: Мир, 1987.— 398 с.
8. Уотсон Дж. Молекулярная биология гена.— М.: Мир, 1978.— 467 с.
9. Meselson M., Radding C. A general model for genetic recombination // *Proc. Nat. Acad. Sci. USA.*— 1975.— 72, N 5.— P. 358—361.
10. Kunkel T. A. The mutational specificity of DNA polymerases- $\alpha$  and - $\gamma$  during *in vitro* DNA synthesis // *J. Biol. Chem.*— 1985.— 260, N 26.— P. 12866—12879.
11. Topal M. D., Eadie J. S., Conrad M. O<sup>6</sup>-Methylguanine mutation and repair is nonuniform // *Ibid.*— 1986.— 261, N 21.— P. 9879—9885.
12. Gearhart P. J., Bogenhagen D. F. Clusters of point mutations are found exclusively around rearranged antibody variable genes // *Proc. Nat. Acad. Sci. USA.*— 1983.— 80, N 16.— P. 3439—3443.
13. A hyperconversion mechanism generates the chicken light chain preimmune repertoire / C.-A. Reynaud, V. Anquez, H. Grimal, J.-C. Weill // *Cell.*— 1987.— 48, N 2.— P. 379—388.
14. Рогозин И. Б., Соловьев В. В., Колманов Н. А. Контекстная преддетерминированность мутационного процесса (соматические, спонтанные и индуцированные точечные мутации).— Новосибирск, 1988.— 54 с.— (Препринт / Сиб. отд-ние АН СССР, Ин-т цитологии и генетики, № 08031).
15. Berek C., Griffiths G. M., Milstain C. Molecular events during maturation of the immune response to oxazoline // *Nature.*— 1985.— 316.— P. 412—418.
16. Rabbits T. N., Hamlyn P. H., Baer R. Altered nucleotide sequences of a translocated *c-myc* gene in Burkitt lymphoma // *Ibid.*— 1983.— 306.— P. 760—765.
17. Hauser J., Levine A. S., Dixon K. Unique pattern of point mutations arising after gene transfer into mammalian cells // *EMBO J.*— 1987.— 6, N 1.— P. 63—67.

Ин-т цитологии и генетики Сиб. отд-ния АН СССР,  
Новосибирск

Получено 06.04.90

УДК 577.323.435

А. П. Гультяев, Ю. П. Монаков

## МЕТОД ПОСТРОЕНИЯ ВТОРИЧНОЙ СТРУКТУРЫ РНК НА ОСНОВЕ ПРИНЦИПОВ САМООРГАНИЗАЦИИ

*Описан метод компьютерного построения вторичной структуры РНК на основе последовательности нуклеотидов. Алгоритм использует метод Монте-Карло для моделирования самоорганизации РНК при последовательном образовании новых спиральных участков. Метод может быть использован для расчетов структур достаточно длинных РНК (вплоть до 5000 нуклеотидов) на персональном компьютере без использования больших объемов памяти.*

**Введение.** Известно, что молекулы РНК имеют тенденцию к образованию двунитчатых участков с уотсон-криковскими парами нуклеотидов, связывая друг с другом отдельные фрагменты молекулы. Образующаяся структура именуется вторичной структурой РНК (этот термин используется также для описания молекулярной геометрии РНК и ДНК, т. е. А, В, С-структур и т. д., но в нашем случае мы под вторичной структурой будем понимать только наличие определенных двунитчатых участков в данной молекуле РНК). Вторичная структура РНК играет важную роль во многих процессах [1].

Зачастую экспериментальное определение вторичной структуры РНК затруднено, поэтому большое значение имеют теоретические мето-

© А. П. ГУЛЬТЯЕВ, Ю. П. МОНаКОВ, 1990