

ОБ ОСОБЕННОСТЯХ МЕХАНИЗМА ТОЧЕЧНЫХ СПОНТАННЫХ МУТАЦИЙ В СТРУКТУРЕ Y-ХРОМОСОМЫ ЧЕЛОВЕКА

Самченко А.А., Киселев С.С., Кондратьев М.С., Комаров В.М.

Институт биофизики клетки РАН

ул. Институтская, 3а, г. Пущино, 142290, РФ; e-mail: komarov_vm@mail.ru

Поступила в редакцию: 05.07.2018.

Аннотация. В работе с использованием методов сравнительной геномики проанализирована специфика распределения мононуклеотидных и смешанных треков W- и S-типа в структуре ДНК половой Y-хромосомы и в близкой к ней по размерам 22 хромосомы человека. Обнаружено, что у Y-хромосомы в ее регуляторной области наблюдается аномально высокая плотность S (G/C)-треков, состоящих из относительно коротких (до 6 пар нуклеотидов) последовательностей из G/C- (и C/G-) пар. Для 22 хромосомы небольшая избыточность таких S-треков наблюдается лишь в структуре ее кодирующих областей (экзонов). Ранее в своих исследованиях мы указывали на не тривиальную, повышенную способность G/C-пар по сравнению с A/T-парами к спонтанному изменению геометрии своего комплементарного H-связывания (из-за разных размерностей скрытого структурного полиморфизма уотсон-криковских G/C- и A/T-пар). В данном случае, выявленная большая плотность S(G/C)-треков в межгенной части Y-хромосомы указывает на возможное увеличение числа событий спонтанного нарушения комплементарности в этой области по сравнению с аналогичными событиями в аутосомной 22 хромосоме. То есть, таким способом может инициироваться аномально высокая частота точечных спонтанных мутаций Y-хромосомы по сравнению с частотой аналогичных мутаций в других хромосомах человека. Что и наблюдается экспериментально.

Ключевые слова: ДНК структура хромосом, доминирование S-треков, структурный полиморфизм уотсон-криковских пар, частота спонтанных мутаций.

Как известно, биологическая функция крохотной Y-хромосомы заключается в определении мужского пола организма у подавляющего большинства млекопитающих. Размеры этой хромосомы невелики. У человека она содержит 59 миллионов пар нуклеотидов (59 Mb) и занимает лишь 1,6 % от размера всего генома (3 000 Mb). Y-хромосома состоит примерно из 76 генов, кодирующих белки, главный из которых – ген *SRY* (sexdetermining region Y). В нем отсутствуют интроны. Его единственный экзон имеет 897 п.н. В целом однако считается, что почти 95 % генов Y-хромосомы связаны с развитием мужских признаков и фертильности [1].

Как и другие хромосомы, данная хромосома подвержена определенным мутационным воздействиям и в первую очередь действию точечных спонтанных мутаций. Они оказывают существенное влияние на эволюционный характер развития организма, а также являются первопричиной возникновения некоторых патологий. В то же время в связи с отсутствием гомологов в мужском хромосомном наборе, в Y-хромосоме нет межхромосомного обмена ДНК, так характерного для мутационных процессов остальной части генома. Как следствие, из-за отсутствия рекомбинационных событий консерватизм (изолированность) является отличительной характеристикой этой небольшой хромосомы.

Тем не менее, весьма неожиданным обстоятельством оказывается наблюдаемая здесь довольно высокая скорость точечных спонтанных мутаций, которая почти в 5 раз выше таковой во всех других хромосомах [2-8]. Происхождение данного, очень важного явления, до сих пор не выяснено. Поэтому установление специфики нуклеотидной организации Y-хромосомы (вариативности её структуры по сравнению с другими хромосомами) представляется нам крайне актуальной биофизической задачей.

ЦЕЛЬ РАБОТЫ

Прежде, чем перейти к формулировке конкретной цели нашей работы, необходимо сделать два предварительных замечания, касающихся общего фона структурной организации геномов эукариот.

Во-первых, установлено, что основу структурно-функциональной организации хромосом различных видов организмов составляют уникальные и повторяющиеся нуклеотидные последовательности (треки) [9-12]. Доля таких последовательностей может достигать иногда 98 % всей ДНК. Хотя функции большинства из них до сих пор не определены, некоторые закономерности в статистике их встречаемости как в GC-дефицитных, так и в GC-богатых геномах про- и эукариот уже замечены [13-22]. Например, нами с использованием методов сравнительной геномики было выявлено [22], что частота встречаемости в ДНК эукариот мононуклеотидных poly(A)_n- и poly(T)_n-треков, а также смешанных олигонуклеотидных W-треков (т.е. треков из A/T- и T/A-пар) практически всегда доминирует над встречаемостью мононуклеотидных poly(G)_n-, poly(C)_n-треков и смешанных S-треков (треков из G/C- и C/G-пар) равной протяженности. Физической первопричиной такой необычной асимметрии распределения A/T- и G/C-пар в составе геномов, как было показано в этой же работе, может выступать исходно разная размерность скрытого структурного полиморфизма геометрии комплементарного спаривания A/T- и G/C-пар. Одиночные G/C-пары по сравнению с A/T-парами по причине своей высокой,

4-кратной неоднозначности геометрии уотсон-криковского спаривания оснований оказались повышенным источником спонтанного нарушения регулярной геометрии Н-связывания комплементарных нуклеотидных цепей. Вполне вероятно, что именно по этой причине содержание G/C-пар в составе геномов живых организмов оказывается важнейшим критическим фактором. GC-состав ДНК, по-видимому, сильно «минимизирован» природой для сохранения стабильности воспроизведения уникальной геометрии комплементарного спаривания азотистых оснований, для обеспечения большей надежности протекания генетических процессов и, в особенности, такого важного процесса, как репликация ДНК.

Во-вторых, как показывают многочисленные генетические и биохимические исследования, возникающие в хромосомах мутации неравномерно локализируются по всему геному. В некоторых сайтах происходит гораздо больше мутаций, чем следует ожидать при их случайном распределении. Такие сайты называют *горячими точками*. Природа этих горячих точек в настоящее время не совсем ясна. Но если рассматривать мутационные сайты (*под сайтом в данном случае рассматривается одна пара оснований*) внутри определенной последовательности ДНК, то возникает естественный вопрос: все ли пары оснований в одинаковой степени чувствительны к изменениям в ее структуре или некоторые из них просто мутируют с большей частотой, чем другие?

В этом отношении показательны найденные закономерности. Так, плотность появления горячих точек оказалась хорошо коррелирующей с GC-составом и числом повторов в гене. В случае с наиболее распространенным спонтанным нарушением ДНК в ходе репликации и репарации – к потере оснований и их дезаминированию – особенно чувствительными оказываются почему-то цитозиновые остатки. Кроме того, обнаружилось, что спонтанно мутируют по типу транзиций главным образом гены, имеющие в своей структуре большой процент динуклеотидов CG [4, 23-26].

Все эти обстоятельства указывают, таким образом, на особую роль нуклеотидных последовательностей и повторов из GC-пар (в нашей терминологии – треков S(G/C)-природы) в инициации процесса спонтанного нарушения комплементарных спариваний в структуре двойной спирали ДНК.

Цель данной работы – на основе сравнения частот встречаемости треков разной природы, т.е. W(A/T)- и S(G/C)-треков, в первичной структуре ДНК определить специфику локализации «активных» в точечных заменах последовательностей из G/C-пар в двух близких по размерам хромосомах – Y-хромосоме и 22 хромосоме *Homo sapiens*. Связать найденные отличия распределения треков S(G/C) природы с наблюдаемыми характеристиками точечных спонтанных мутаций этих хромосом.

РЕЗУЛЬТАТЫ

Анализ частот встречаемости мононуклеотидных (A_n , T_n , G_n , C_n) и смешанных W- и S-треков в Y-хромосоме и 22 хромосоме человека проводился по данным секвенирования геномов эукариот, взятых из базы **GenBank** (www.ncbi.nlm.nih.gov/genbank; ftp://ncbi.nlm.nih.gov/genomes/H_sapiens). Оценивались частоты появления треков постепенно увеличивающейся длины n от 1 до n_{max} , где n_{max} – максимальная длина нуклеотидного трека, обнаруживаемого в хромосомах. Все расчеты выполнялись на основе разработанной собственной компьютерной программы, особенности алгоритма которой описаны в [22].

Обнаружено, что в обеих хромосомах (как в целых хромосомах, так и внутри генов) частоты встречаемости мононуклеотидных (A_n и T_n) и смешанных W-треков оказались доминирующими над встречаемостью мононуклеотидных (G_n и C_n) и смешанных S-треков (рис. 1, 2, 4, 5, 7, 8, 10, 11), причём аналогичные результаты были получены ранее на полноразмерных геномах эукариот [22], (На представленных графиках из-за очень большого диапазона изменений значений частот встречаемости величины f приведены в логарифмическом масштабе, $\log f$). В тоже время в экзонах обнаружилось уже свои специфические частотные особенности. Так, согласно рисункам 3, 9 распределения всех 4 типов мононуклеотидных треков оказываются сближенными по частоте встречаемости в обеих хромосомах. При этом если у половой, Y-хромосомы наблюдается сближение частот встречаемости еще и W- и S-треков т.е. треков смешанной природы (рис. 6), то у аутосомной 22 хромосомы человека (рис. 12) мы фиксируем значительное преобладание встречаемости S-треков над W-треками. В свете сделанных выше замечаний об отличительных особенностях внутреннего структурного полиморфизма G/C-пар обнаруженное обстоятельство указывает на повышенную склонность кодирующих областей этой и, по-видимому, других аутосом к такому механизму иницирования спонтанных точечных генных мутаций.

Уместно указать на имеющиеся наблюдения по направленности спонтанных процессов нуклеотидных транзиций и трансверсий в ядерных геномах эукариот. Как ни странно, именно G/C-пары обладают повышенной (иногда в несколько раз) частотой точечных спонтанных мутаций $G:C \rightarrow A:T$ ($G:C \rightarrow T:A$) по сравнению с A/T-парами, с их заменами $A:T \rightarrow G:C$ или $A:T \rightarrow C:G$ [7,8].

Поразительным оказался результат сравнительного анализа распространенности смешанных W- и S-треков для межгенных (регуляторных) областей этих хромосом. На общем фоне доминирования повторяющихся нуклеотидных последовательностей из A/T-пар Y-хромосома демонстрирует здесь неожиданно высокую плотность встречаемости $\rho(n)$ S-треков (где n – длина трека) довольно короткой протяженности, до $n \leq 6$ нуклеотидов (рис. 13). (Плотность $\rho(n)$ определялась как отношение числа встречаемости трека длины n к общему числу нуклеотидных пар в хромосоме). По сравнению с аналогичной областью 22 хромосомы, как видно из графиков, отличия значительные!

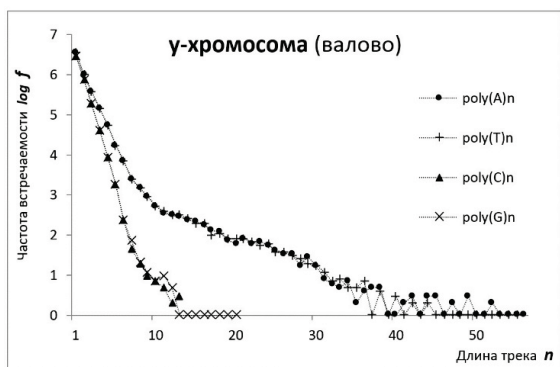


Рисунок 1. Частота встречаемости монуклеотидных треков в целой структуре Y-хромосомы

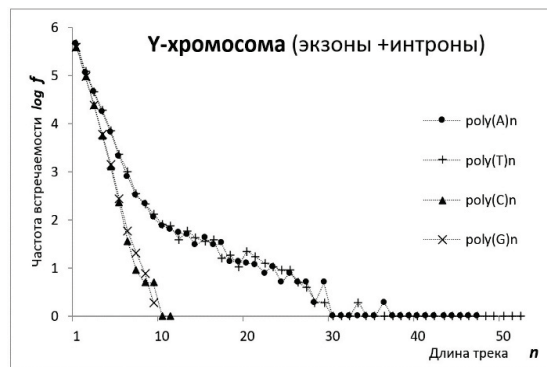


Рисунок 2. Частота встречаемости монуклеотидных треков в генах Y-хромосомы

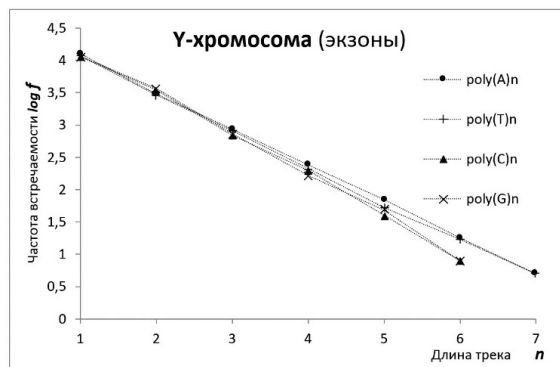


Рисунок 3. Частота встречаемости монуклеотидных треков в кодирующей части Y-хромосомы

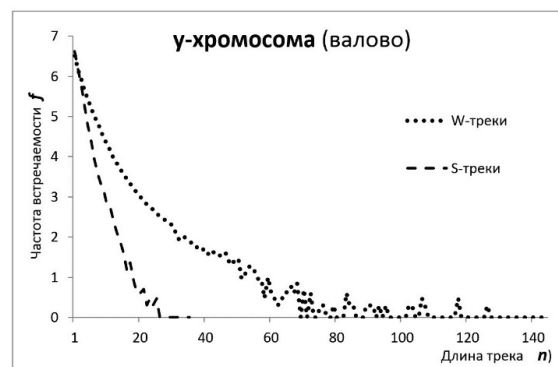


Рисунок 4. Частота встречаемости смешанных W- и S-треков в целой структуре Y-хромосомы

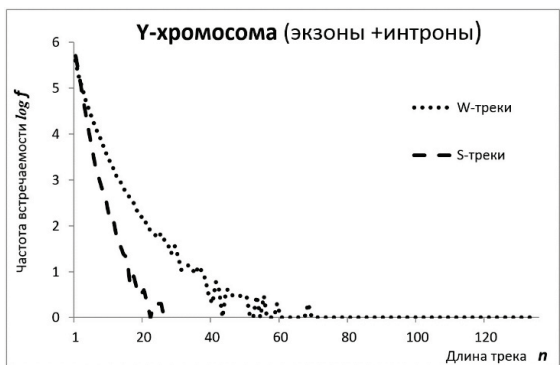


Рисунок 5. Частота встречаемости смешанных W- и S-треков в генах Y-хромосомы

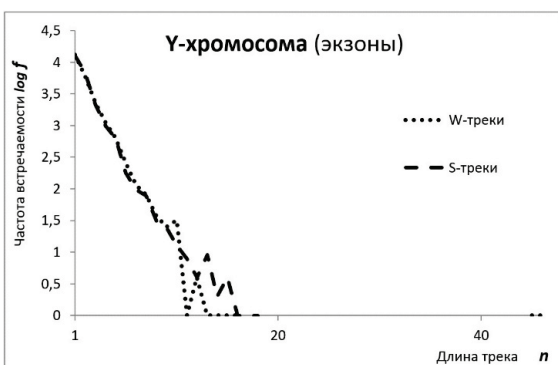


Рисунок 6. Частота встречаемости смешанных W- и S-треков в кодирующей части Y-хромосомы

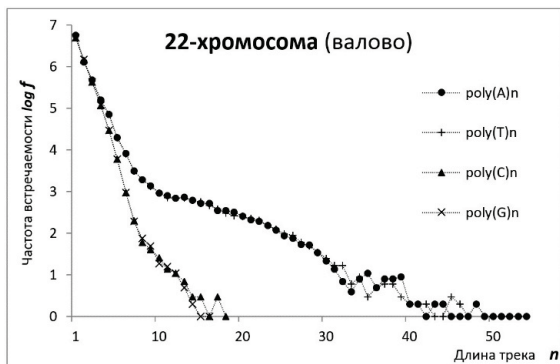


Рисунок 7. Частота встречаемости монуклеотидных треков в целой структуре 22-хромосомы

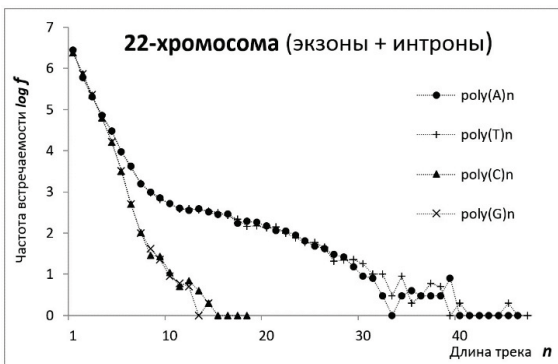


Рисунок 8. Частота встречаемости монуклеотидных треков в генах 22-хромосомы

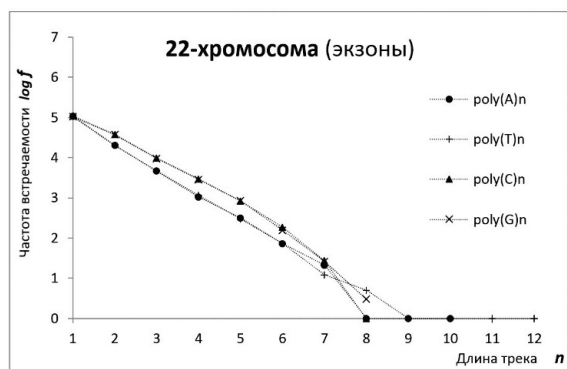


Рисунок 9 - Частота встречаемости монуклеотидных треков в кодирующей части 22-хромосомы

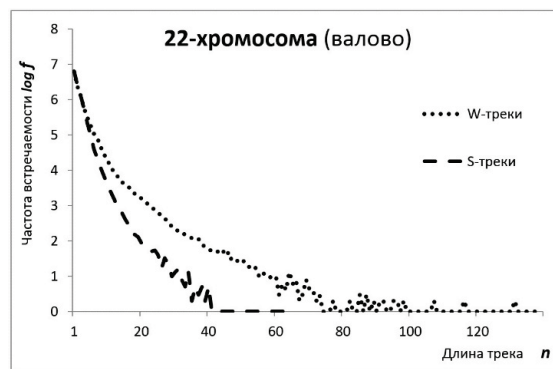


Рисунок 10 - Частота встречаемости смешанных W- и S-треков в целой структуре 22-хромосомы

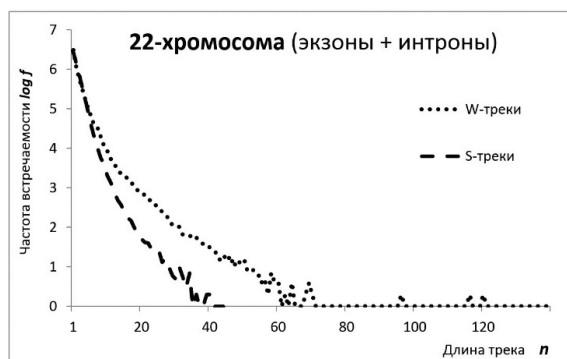


Рисунок 11. Частота встречаемости смешанных W- и S-треков в генах 22-хромосомы

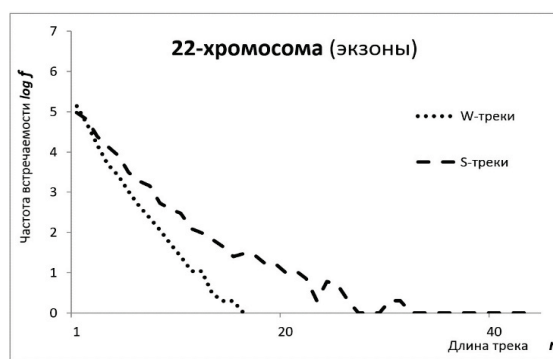


Рисунок 12. Частота встречаемости смешанных W- и S-треков в кодирующей части 22-хромосомы

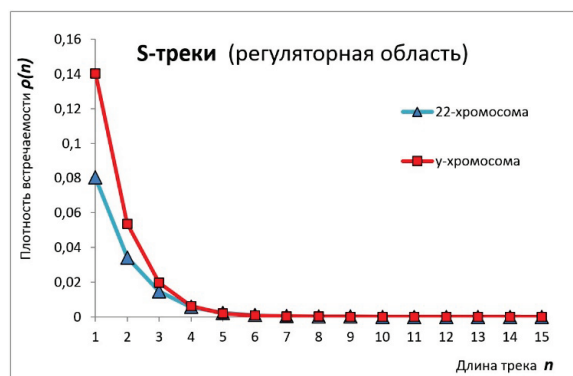


Рисунок 13. Плотность встречаемости смешанных S-треков в регуляторной области 22- и Y-хромосомы

Таким образом, представленные результаты позволяют сделать вывод о важной специфике нуклеотидной организации половой Y-хромосомы. Она, по-видимому, эволюционно минимизировала кодирующие области ДНК к возникновению в них спонтанных генных патологий и при этом повысила свою общую активность к точечным мутациям за счет регуляторных участков. Это, в целом, хорошо коррелирует с имеющимися наблюдениями.

Список литературы / References:

1. Ali S., Hasnain S.E. Genomics of the human Y-chromosome-1. Association with male infertility. *Gene*, 2003, vol. 4 (321), pp. 25-37.
2. Lindblad-Toh K. [et al.] Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 2005, vol. 438, pp. 803-819.
3. Graves J.A.M. Sex Chromosome Specialization and Degeneration in Mammals. *Cell*, 2006, vol. 124 (5), pp. 901-914.
4. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002, vol. 420, pp. 520-562.
5. Wen-Hsiung Li [et al.] Male-driven evolution. *Current Opinion in Genetics & Development*, 2002, vol. 12, pp. 650-656.

6. Ballantyne K.N. [et al.] Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. *The American Journal of Human Genetics*, 2010, vol. 87, pp. 341-353.
7. Lynch M., Sung W., Morris K., Coffey N., Landry C.R., Dopman E.B., Dickinson W.J., Okamoto K., Kulkarni S., Hartl D.L., Thomas W.K. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. USA*, 2008, vol. 105 (27), pp. 9272-9277.
8. Dillon M.M., Sung W., Lynch M., Cooper V.S. The Rate and Molecular Spectrum of Spontaneous Mutations in the GC-Rich Multichromosome Genome of *Burkholderia cenocepacia*. *Genetics*, 2015, vol. 200, pp. 935-946.
9. Mehrotra S., Goyal V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics Proteomics Bioinformatics*, 2014, vol. 12 (4), pp. 164-171.
10. Forster H., Röhl A., Lünemann P., Brinkmann C., Zerjal T., Tyler-Smith C., Brinkmann B. A Short Tandem Repeat-Based Phylogeny for the Human Y Chromosome. *The American Journal of Human Genetics*, 2000, vol. 67 (1), pp. 182-196.
11. Jobling M.A., Tyler-Smith C. Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Science International*, 2001, vol. 118, pp. 158-162.
12. Jobling M.A., Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age". *Nature Reviews. Genetics*, 2003, vol. 4, pp.598-612.
13. Subirana J.A., Messeguer X. The distribution of alternating AT sequences in eukaryotic genomes suggests a role in homologous chromosome recognition in meiosis. *J. Theor. Biol.*, 2011, vol. 283, pp. 28-34.
14. Marx K.A., Zhou Y., Kishawi I.Q. Evidence for Long Poly(dA).Poly(dT) Tracts in *D. Discoideum* DNA at High Frequencies and Their Preferential Avoidance of Nucleosomal DNA Core Regions. *J. Biomol. Struct. & Dyn.*, 2006, vol. 23, pp. 429-446.
15. Subirana J.A., Messeguer X. The most frequent short sequences in non-coding DNA. *Nucleic Acids Res.*, 2010, vol. 38 (4), pp. 1172-1180.
16. Yagil G. The frequency of two-base tracts in eukaryotic genomes. *J. Mol. Evol.*, 1993, vol. 37, pp. 123-130.
17. Yagil G. DNA tracts composed of only two bases concentrate in gene promoters. *Genomics*, 2006, vol. 87, pp. 591-597.
18. Shomer D., Yagil G. Long W tracts are over-represented in the *E.coli* and *H.influenzae* genomes. *Nucl.AcidsRes.*, 1999, vol. 27, pp. 4491-4480.
19. Киселев С.С., Комаров В.М., Масулис И.С., Озолин О.Н. Распределение мононуклеотидных повторов в бактериальных хромосомах. А/Т-треки преобладают над G/C-треками. *Компьют. исследов. моделир.*, 2010, т. 2, с. 183-187. [Kiselev S.S., Komarov V.M., Masulis I.S., Ozoline O.N. Distribution of mononucleotide repeats in bacterial chromosomes. Redundancy of A/T-tracts under G/C-tracts. *Computer researches & modelling*, 2010, vol. 2, pp. 183-187. (In Russ.)]
20. Piazza F., Lio P. Statistical analysis of simple repeats in the human genome. *Physica A*, 2005, vol. 347, pp. 472-488.
21. Decherig K.J., Cuelenaere K., Konings R.N.H., Lennissen J.A.M. Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. *Nucleic Acids Research*, 1998, vol. 26, pp. 4056-4062.
22. Самченко А.А., Киселев С.С., Кабанов А.В., Кондратьев М.С., Комаров В.М. О природе доминирования олигомерных (dA:dT)_n треков в структуре геномов эукариот. *Биофизика*, 2016, т. 6, с. 1045-1058. [Samchenko A.A., Kiselev S.S., Kabanov A.V., Kondratjev M.S., Komarov V.M. On the nature of oligo (dA:dT)_n tracks domination in the structure of eukaryotic genomes. *Biofizika*, 2016, vol. 6, pp. 1045-1058. (In Russ.)]
23. Mackewicz D. [et al.] Distribution of Recombination Hotspots in the Human Genome – A Comparison of Computer Simulations with Real Data. *PLOS ONE*, 2013, vol. 8, e65272.
24. Cooper D.N., Youssoufian H. The CpG dinucleotide and human genetic disease. *Hum Genet.*, 1988, vol. 78, pp. 151-155.
25. Sved J., Bird A. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc. Natl. Acad. Sci USA*, 1990, vol. 87, pp. 4692-4696.
26. Fryxell K.J., Moon W.J. CpG mutation rates in the human genome are highly dependent on local GC content. *Mol. Biol. Evol.*, 2005, vol. 22, pp. 650-658.

ON THE PECULIARITIES OF THE MECHANISM OF POINT SPONTANEOUS MUTATIONS IN HUMAN Y-CHROMOSOME STRUCTURE

Samchenko A.A., Kiselev S.S., Kondratyev M.S., Komarov V.M.

Institute of Cell Biophysics of RAS

Institutskaya str. 3a, Pushchino, 142290, Russia; e-mail: komarov_vm@mail.ru

Abstract. The specific character of distribution of mononucleotide tracts and mixed W or S tracts in human Y and 22 chromosomes was analysed by comparative genomic approaches. It was found that in regulatory regions of Y chromosome anomalously high density of S (G/C)-tracts consisting of relatively short sequences (up to 6 base pairs) was observed. In primary structure of coding regions (exons) of 22 chromosome only a small redundancy of S- tracts was discovered. Previously in our papers we pointed to the non-trivial, increased ability of G/C pairs to a spontaneous change in the geometry of its complementary H-binding (because of the different dimensions of the hidden structural polymorphism of Watson-Crick G/C and A/T base pairs) in comparison with A/T pairs. In this case, revealed high density of S(G/C)-tracts in intergenic regions of Y chromosome indicated to possibility of increasing in the number of events of spontaneous complementarity lesions in this areas in comparison with similar events in the autosomic 22 chromosome. That is, in this way an abnormally high frequency of spontaneous point mutations in the Y chromosome compared with the frequency of similar mutations in other human chromosomes can be initiated. This is observed experimentally.

Key words: *DNA, chromosome structure, redundancy of S-tracts, structural polymorphism of Watson-Crick base pairs, frequency of spontaneous mutations.*