

Изучение прямых, инвертированных и зеркальных повторов в X хромосоме человека

Лыскова Алиса Олеговна

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: alisalyskovaaa@gmail.com

Доступные ранее методы секвенирования не позволяли безошибочно собрать X хромосому человека. Благодаря новому методу нанопорового секвенирования была получена точная последовательность X хромосомы, что позволило изучить в ней повторы. [1] Некоторые повторы необходимы для функционирования генома, но функции большинства из них до сих пор остаются неизвестными. Таким образом, цель нашей работы - изучить прямые, инвертированные и зеркальные повторы в X хромосоме человека.

Чтобы найти прямые, инвертированные и зеркальные повторы, мы написали программу на языке программирования Си и визуализировали результаты с помощью R. Повторов всех трех видов оказалось намного больше, чем можно было бы ожидать. Мы сравнили количество повторов, найденных в сборке illumina 2013 года и в последовательности, полученной с помощью нанопорового секвенирования.

Известно, что полипуриновые-полипиримидиновые зеркальные повторы могут формировать H-DNA. Такие структуры, вероятно, блокируют транскрипцию. [2] В нашей работе мы сосредоточились на изучении полипуриновых зеркальных повторов. Мы показали, что короткие повторы (с длиной плеча до 21 нуклеотида) являются вероятно случайными, образованными в результате мутагенеза. Большая часть повторов с длиной плеча от 22 нуклеотидов являются микросателлитами.

В нашей работе мы показали оверрепрезентацию прямых, инвертированных и зеркальных повторов и изучили полипуриновые зеркальные повторы в X хромосоме человека.

Источники и литература

- 1) Miga KH, Koren S, Rhie A, Vollger MR et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature*. 2020 Sep;585(7823):79-84. doi: 10.1038/s41586-020-2547-7. Epub 2020 Jul 14. PMID: 32663838; PMCID: PMC7484160
- 2) Pandey S, Ogloblina AM, Belotserkovskii BP, et al. Transcription blockage by stable H-DNA analogs in vitro. *Nucleic Acids Res*. 2015;43(14):6994-7004. doi:10.1093/nar/gkv622
- 3) Manor H, Rao BS, Martin RG. Abundance and degree of dispersion of genomic d(GA)n.d(TC)n sequences. *J Mol Evol*. 1988;27(2):96-101. doi: 10.1007/BF02138367. PMID: 3137357
- 4) Saha SK, Goswami A, Dutta C. Association of purine asymmetry, strand-biased gene distribution and PolC within Firmicutes and beyond: a new appraisal. *BMC Genomics*. 2014;15(1):430. Published 2014 Jun 4. doi:10.1186/1471-2164-15-430
- 5) Lavi B, Levy Karin E, Pupko T, Hazkani-Covo E. The Prevalence and Evolutionary Conservation of Inverted Repeats in Proteobacteria. *Genome Biol Evol*. 2018;10(3):918-927. doi:10.1093/gbe/evy044
- 6) Montermini L, Andermann E, Labuda M, Richter A, Pandolfo M, Cavalcanti F, Pianese L, Iodice L, Farina G, Monticelli A, Turano M, Filla A, De Michele G, Coccozza S. The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. *Hum Mol Genet*. 1997 Aug;6(8):1261-6. doi: 10.1093/hmg/6.8.1261. PMID: 9259271

- 7) Rhonda M Clark, Gillian L Dalglish, Dan Endres, Mariluz Gomez, Jim Taylor, Sanjay I Bidichandani. Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. *Genomics*, Volume 83, Issue 3, 2004, Pages 373-383, ISSN 0888-7543
- 8) Ole K Tørresen, Bastiaan Star, Pablo Mier, Miguel A Andrade-Navarro, Alex Bateman, Patryk Jarnot et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Research*, Volume 47, Issue 21, 02 December 2019, Pages 10994–11006
- 9) Gabriel Matos-Rodrigues, Niek van Wietmarschen, Wei Wu, Venu Tripathi, Natasha Koussa et al. Linking Dynamic DNA Secondary Structures to Genome Instability. *bioRxiv*, 10.1101/2022.01.19.476973, 2022.01.19.476973
- 10) Cer RZ, Bruce KH, Mudunuri US, Yi M, Volfovsky N, Luke BT, Bacolla A, Collins JR, Stephens RM. Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D383-91. doi: 10.1093/nar/gkq1170. Epub 2010 Nov 21. PMID: 21097885; PMCID: PMC3013731
- 11) Bacolla A, Tainer JA, Vasquez KM, Cooper DN. Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res.* 2016;44(12):5673-5688. doi:10.1093/nar/gkw261
- 12) Zhou, Y., Bizzaro, J.W. & Marx, K.A. Homopolymer tract length dependent enrichments in functional regions of 27 eukaryotes and their novel dependence on the organism DNA (G+C)% composition. *BMC Genomics* 5, 95 (2004)
- 13) Bacolla A, Collins JR, Gold B, et al. Long homopurine*homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region. *Nucleic Acids Res.* 2006;34(9):2663-2675. Published 2006 May 19. doi:10.1093/nar/gkl354
- 14) Malik HS. Ribonuclease H evolution in retrotransposable elements. *Cytogenet Genome Res.* 2005;110(1-4):392-401. doi: 10.1159/000084971. PMID: 16093691
- 15) Lu JY, Shao W, Chang L, Yin Y, Li T, Zhang H, Hong Y, Percharde M, Guo L, Wu Z, Liu L, Liu W, Yan P, Ramalho-Santos M, Sun Y, Shen X. Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep.* 2020 Mar 10;30(10):3296-3311.e5. doi: 10.1016/j.celrep.2020.02.048. PMID: 32160538; PMCID: PMC7195444
- 16) Paulson H. Repeat expansion diseases. *Handb Clin Neurol.* 2018;147:105-123. doi: 10.1016/B978-0-444-63233-3.00009-9. PMID: 29325606; PMCID: PMC6485936