

Альтернативные структуры ДНК обнаружены вблизи ультраконсервативных некодирующих элементов генома

Макарова И.*, Калюжный Д.

Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия

* irina.vic.makarova@gmail.com

Ключевые слова: ультраконсервативные некодирующие элементы генома (УНЭ); альтернативные структуры ДНК; ДНК квадруплексы; ДНК триплексы

Мотивация и цель: Некодирующие, но при этом консервативные элементы генома сохранились практически неизменными от курицы и других эволюционно более удаленных позвоночных до человека. Причины такой эволюционной стабильности исследуют с точки зрения их нуклеотидного состава, эпигенетической регуляции, ассоциированности с сайтами связывания транскрипционных факторов и границами топологически ассоциированных доменов [1, 2]. Наша цель – расширить анализ и определить возможную роль примыкающих участков в эволюционной консервативности данных элементов.

Методы и алгоритмы: Мы проанализировали ультраконсервативные некодирующие элементы (УНЭ) в трех организмах – человеке (*Homo sapiens*), мыши (*Mus musculus*) и рыбке Данио (*Danio rerio*). Положения УНЭ были использованы из UCNEbase [3]. Положения последовательностей, способных формировать структуры, отличные от В-формы, были получены из следующих источников:

- Для квадруплексных участков использовались данные эксперимента G4-seq [4].
- Для предсказания мотивов ДНК триплексов (Н-мотивов) и крестообразных структур применялся Inverted Repeats Finder версии 3.08 [5], где они были определены как зеркальные полипиримидиновые/ полипуриновые и инвертированные повторы соответственно.

Статистический анализ проводился относительно перемешанных положений структур, отличных от В-формы. Для предобработки и анализа данных использовались скрипты, написанные на языке Python версии 3.11 в редакторе кода Visual Studio Code версии 1.87.0. Анализ обогащения по функциональной принадлежности был выполнен с помощью STRING-db версии 12.0.

Результаты: При анализе нуклеотидного состава участков, примыкающих к ультраконсервативным элементам, мы обнаружили обогащение гуанином и цитозином участков, смежных с УНЭ во всех трех геномах. Мы также заметили асимметрию обогащения гуанином с 3' конца элементов. Наиболее выраженное обогащение было отмечено для рыбки Данио.

Отмеченная неравномерность нуклеотидного состава подтолкнула нас проанализировать структуры, отличные от В-формы, специфичные к геномной последовательности, которые могли бы образовываться на участках внутри и в непосредственной близости к УНЭ.

Мы обнаружили, что участки, примыкающие к ультраконсервативным элементам с 3' конца, значимо обогащены квадруплексами в геномах человека и мыши. В

геноме рыбки Данио достоверное обогащение характерно для примыкающих участков в обоих направлениях, при этом 3' конец обогащен в большей степени. Н-мотивы также значимо обогащали участки, прилегающие к ультраконсервативным элементам, в геномах всех трех организмов. Для Н-мотивов не характерна асимметричность обогащения с одного конца, но при этом они преимущественно локализованы на кодирующей цепи ДНК. Интересно отметить, что разные ультраконсервативные элементы фланкированы квадруплексами и Н-мотивами.

Анализ мотивов крестообразных структур не показал наличия паттерна ассоциированности с ультраконсервативными элементами.

Выводы: Мы впервые провели анализ ассоциации ультраконсервативных некодирующих элементов с альтернативными структурами ДНК. Значимое обогащение квадруплексными структурами и триплексными мотивами может являться ключом к пониманию устойчивости нуклеотидных последовательностей в эволюции. Наше исследование открывает перспективы изучения функциональной значимости и эволюционной стабильности ультраконсервативных некодирующих элементов.

Non-B DNA structures adjacent to ultraconserved non-coding elements unveiled

Makarova I.*, Kaluzhny D.

Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

* irina.vic.makarova@gmail.com

Key words: ultraconserved non-coding elements (UCNEs); non-B DNA structures; DNA quadruplexes; DNA triplexes

Motivation and Aim: Non-coding, yet highly conserved genomic elements persist virtually unchanged from chickens and other evolutionarily distant vertebrates to humans. The reasons for such evolutionary stability are explored in terms of their nucleotide composition, epigenetic regulation, association with transcription factor binding sites and boundaries of topologically associated domains [1, 2]. Our aim is to expand the analysis and determine the potential role of adjacent regions in the evolutionary conservation of these elements.

Methods and Algorithms: We analyzed ultraconserved non-coding elements (UCNEs) in three organisms – human (*Homo sapiens*), mouse (*Mus musculus*), and zebrafish (*Danio rerio*). Positions of UCNEs were retrieved from UCNEbase [3]. Positions of sequences capable of forming non-B DNA structures were obtained from the following sources:

- For quadruplex regions, data from G4-seq experiment [4] were utilized.
- For predicting regions of DNA triplexes (H-motifs) and cruciform structures Inverted Repeats Finder version 3.08 [5] was used. H-motifs and cruciform motifs were defined as mirror polypyrimidine/polypurine repeats and inverted repeats, respectively.

Statistical analysis was carried out regarding the shuffled positions of non-B DNA structures. Data preprocessing and analysis were conducted using Python 3.11 scripts in

the Visual Studio Code editor. Gene-set enrichment analysis was performed using STRING-db version 12.0.

Results: In all three genomes, we observed enrichment of guanine and cytosine in regions adjacent to UCNEs. We also noted an asymmetry in guanine enrichment at the 3' end of the elements, with the most pronounced enrichment observed in zebrafish.

The observed unevenness in nucleotide composition prompted us to examine non-B DNA structures that are specific to the genomic sequence and could form within or in close proximity to UCNEs. We found significant enrichment of quadruplexes in regions adjacent to ultraconserved elements at the 3' end in the human and mouse genomes. In the zebrafish genome, enrichment was observed in both directions, with a greater extent at the 3' end.

H-motifs were also significantly enriched in regions adjacent to UCNEs in all three organisms. Unlike quadruplexes, H-motifs did not exhibit enrichment asymmetry from one end but were predominantly localized on the coding DNA strand. Interestingly, different ultraconserved elements were flanked by quadruplexes and H-motifs.

Analysis of cruciform structure motifs did not reveal a pattern of association with ultraconserved elements.

Conclusion: We conducted, for the first time, an analysis of the association between ultraconserved non-coding elements and non-B DNA structures. Significant enrichment with quadruplex structures and triplex motifs may be key to understanding the stability of nucleotide sequences in evolution. Our study opens up perspectives for investigating the functional significance of ultraconserved non-coding elements.

Список литературы/References

1. Polychronopoulos D., King J.W.D., Nash A.J., Tan G., Lenhard B. Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res.* 2017;45(22):12611-12624. doi 10.1093/nar/gkx1074
2. Fedorova L., Crossley E.R., Mulyar O.A., Qiu S., Freeman R., Fedorov A. Profound non-randomness in dinucleotide arrangements within ultra-conserved non-coding elements and the human genome. *Biology (Basel)*. 2023;12(8):1125. doi 10.3390/biology12081125
3. Dimitrieva S., Bucher P. UCNEbase-A database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res.* 2013;41:D101-D109. doi 10.1093/nar/gks1092
4. Marsico G., Chambers V.S., Sahakyan A.B., McCauley P., Boutell J.M., Antonio M.D., Balasubramanian S. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* 2019;47(8):3862-3874. doi 10.1093/nar/gkz179
5. Warburton P.E., Giordano J., Cheung F., Gelfand Y., Benson G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.* 2004;14(10A):1861-1869. doi 10.1101/gr.2542904