

Различия в организации структуры цис-регуляторных областей генов человека в зависимости от их тканеспецифичности

Слюсарев С.С.*, Бутенко Е.В., Шкурят Т.П.

Южный федеральный университет, Ростов-на-Дону, Россия

* slusarev-sergey@yandex.ru

Ключевые слова: тканеспецифичные гены; CpG островки; Alu-элементы; организация генома

Мотивация и цель: Гены человека можно разделить на повсеместно экспрессируемые гены домашнего хозяйства и тканеспецифичные гены, которые экспрессируются только в одном типе клеток или тканей. Генетический механизм регуляции этих типов генов может быть ассоциирован с организацией цис-регуляторного региона. CpG-островки, расположенные в цис-регуляторных областях, могут изменять уровни экспрессии генов в зависимости от паттерна их метилирования.

Целью данной работы был поиск структурных отличий функциональных элементов в цис-регуляторных регионах тканеспецифичных генов от генов повсеместно экспрессирующихся (генов домашнего хозяйства).

Методы и алгоритмы: Материалом исследования послужили 977 тканеспецифичных генов человека из базы Human Protein Atlas и 554 гена «домашнего хозяйства» [1]. Анализ функциональных элементов проводили с помощью CpGplot, NCBI Genome browser, SINE Base, а также баз данных об экспрессии генов GTEx, Illumina, BioGPS. Исследовали наличие CpG-островков в цис-регуляторных областях (2000 п. н. до сайта начала транскрипции) изучаемых генов.

Результаты: В TSS200 области генов домашнего хозяйства в 88.08 % были представлены GC островками, в TSS1500 области они составляли только 31.04 % от общего числа последовательностей. Для тканеспецифичных генов было обнаружено значительно меньшее число GC островков в TSS200 области (19.55 %), и 9.31 % этих генов имели GC островки в TSS1500 области. Отмечается существенная разница и в длине островков между исследуемыми типами генов. Обнаружено, что средняя длина островка равнялась 566.51 п. н. для тканеспецифичных генов и 719.71 п. н. для генов домашнего хозяйства. С помощью базы данных SINE Base осуществляли поиск гомологии для коротких диспергированных повторов в цис-регуляторных областях изучаемых генов. Alu-элементы присутствовали в 49.84 % цис-регуляторных областей тканеспецифичных генов и в 65.52 % цис-регуляторных областей генов домашнего хозяйства. Обнаружено, что усредненное расстояние от Alu-элемента до сайта начала транскрипции генов домашнего хозяйства меньше, чем это же расстояние для тканеспецифичных генов. Биоинформатический анализ показал, что пары Alu-элементов присутствовали в 13.2 и 23.1 % цис-регуляторных областей тканеспецифичных генов и генов домашнего хозяйства соответственно. Важно отметить, что все Alu-пары имели инвертированный характер ориентации. Среднее расстояние между инвертированными парами Alu-элементов, найденных в цис-регуляторных

областях тканеспецифичных генов было меньше, чем аналогичный показатель для генов домашнего хозяйства.

Выводы: Обнаружены существенные отличия в локализации CpG-островков и Alu-элементов цис-регуляторных регионов тканеспецифичных генов от генов повсеместно экспрессирующихся (генов домашнего хозяйства) в TSS200 и TSSS1500.

Финансирование: Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках государственного задания в сфере научной деятельности № FENW-2023-0018.

Differences in the organisation of the structure of cis-regulatory regions of human genes depending on their tissue specificity

Slusarev S.S.*, Butenko E.V., Shkurat T.P.

Southern Federal University, Rostov-on-Don, Russia

*slusarev-sergey@yandex.ru

Key words: tissue-specific genes, CpG islands, Alu elements, genome organization

Motivation and Aim: Human genes can be divided into ubiquitously expressed housekeeping genes and tissue-specific genes that are expressed in only one type of cell or tissue. The genetic mechanism of regulation of these types of genes may be associated with the organisation of the cis-regulatory region. CpG islands located in cis-regulatory regions may change the expression levels of genes depending on their methylation pattern. The aim of this work was to find structural differences between the functional elements in the cis-regulatory regions of non-specific genes and genes that are universally expressed (housekeeping genes).

Methods and Algorithms: The study material included 977 human tissue-specific genes from the Human Protein Atlas database and 554 housekeeping genes (Eisenberg E., Levanon E. Y., 2013). Analysis of functional elements was carried out using Cpgplot, NCBI Genomebrowser, SINE Base, as well as gene expression databases GTEx, Illumina, BioGPS. The presence of CpG islands in cis-regulatory regions (2000 bp before the transcription start site) of the studied genes was studied.

Results: 88.08 % of housekeeping genes contained CpG islands in the TSS200 region, and only 31.04 % of ubiquitously expressed genes contained a CpG island in the TSS1500 region. For tissue-specific genes a significantly smaller number of CpG islands were found in the studied regions. 19.55 % of tissue-specific genes contained CpG islands in the TSS200 region, and only 9.31% of these genes contained CpG islands in the TSS1500 region. There is also a significant difference in the length of islands between the types of genes studied. It was found that the average length of the island was 566.51 bp. for tissue-specific genes and 719.71 bp for housekeeping genes. Using the SINE Base database, we searched for homology for short dispersed repeats in the cis-regulatory regions of the studied genes. Alu elements were present in 49.84 % of the cis-regulatory regions of tissue-specific genes and in 65.52 % of the cis-regulatory regions of housekeeping genes. It was found that the average distance from the Alu element to the transcription start site of home genes is less than the same distance for tissue-specific genes. Bioinformatics analysis showed that Alu element pairs were

present in 13.2 and 23.1 % of cis-regulatory regions of tissue-specific and housekeeping genes, respectively. It is important to note that all Alu pairs had an inverted orientation. The average distance between inverted pairs of Alu elements found in cis-regulatory regions of tissue-specific genes was less than that for housekeeping genes.

Conclusion: Significant differences were found in the localization of CpG islands and Alu elements of cis-regulatory regions in tissue-specific genes from genes ubiquitously expressed in the TSS200 and TSSS1500 regions.

Funding: The work was carried out with financial support from the Ministry of Science and Higher Education of the Russian Federation within the framework of the state assignment in the field of scientific activity No. FENW-2023-0018.

Список литературы/References

1. Eisenberg E., Levanon E.Y. Human housekeeping genes, revisited. *Trends Genet.* 2013;29(10):569-574