

«СИНФАЗНЫЕ БЛОКИ» – КОНСЕРВАТИВНЫЕ УЧАСТКИ ДВОЙНОЙ СПИРАЛИ ДНК МОДУЛЕЙ РЕГУЛЯЦИИ ТРАНСКРИПЦИИ – СБЛИЖЕНЫ В ПРОСТРАНСТВЕ ВСЛЕДСТВИЕ СФАЗИРОВАННОСТИ ОТНОСИТЕЛЬНО ВИТКА СУПЕРСПИРАЛИ ДНК НУКЛЕОСОМЫ

© 2015 г. А.П. Лифанов, В.Ю. Макеев*, Н.Г. Есипова

Институт молекулярной биологии им. В.А. Энгельгардта РАН, 119991, Москва, ул. Вавилова, 32;

*Институт общей генетики им. Н.И. Вавилова РАН, 119991, ул. Губкина, 3

E-mail: johnnie_me@list.ru

Поступила в редакцию 26.09.14 г.

В результате межвидового сравнения гомологичных нуклеотидных последовательностей локусов генов раннего развития *Drosophila* выделены консервативные участки длиной 30–70 нуклеотидов, по длине находящиеся на промежуточном уровне между сайтами связывания факторов транскрипции (обычно около 7–10 нуклеотидов) и нуклеосомным повтором (165–210 нуклеотидов). Найденные участки располагаются в основном в районе известных функциональных элементов локуса (*cis*-регуляторных модулей – энхансеров, проксимального промотора, кодирующего сегмента). Эти участки, занимая в сумме не более половины общей длины энхансера, содержат в себе подавляющее большинство аннотированных сайтов связывания факторов транскрипции. Квазипериодическая картина размещения консервативных участков хорошо согласуется с экспериментально определенным паттерном локализации нуклеосом. Период расположения соседних участков часто составляет примерно 84 нуклеотида – длину витка суперспирали ДНК, входящей в нуклеосому. Это соответствие периодов позволяет назвать выделенные консервативные участки «синфазными» блоками и утверждать, что они, располагаясь на соседних витках нуклеосомной суперспирали, сближены в пространстве.

Ключевые слова: энхансер, модуль регуляции транскрипции, сайты связывания регуляторных факторов, нуклеосома.

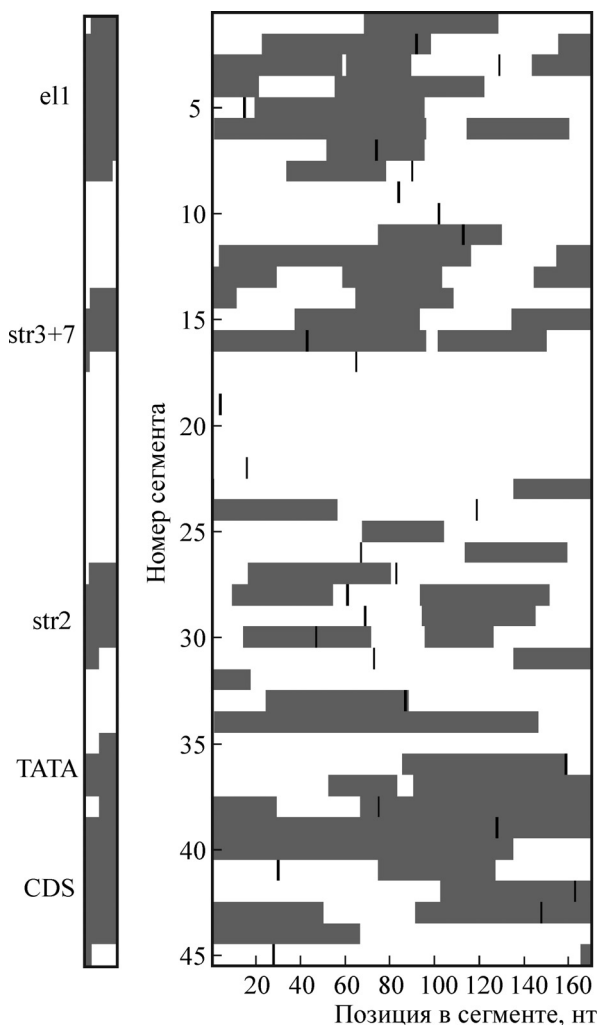
В настоящее время широко распространена как обработка данных об уже известных элементах генома (например, поиск мест «сгущения» и «разрежения» элементов, как CpG-острова [1], кластеры сайтов связывания факторов регуляции транскрипции [2], обогащения или обеднения определенных участков локуса местами связывания нуклеосом [3]), так и подходы, направленные на выявление в геноме новых элементов и закономерностей. Один из таких подходов – сравнение гомологичных нуклеотидных последовательностей для организмов с различным эволюционным расстоянием и выделение похожих участков [4].

В первой части настоящей работы проведено сравнение гомологичных нуклеотидных последовательностей и определены положения консервативных участков для локусов группы генов раннего развития нескольких видов *Dro-*

sophila: *D. melanogaster*, *D. pseudoobscura*, *D. erecta*, *D. littoralis*, *D. willistoni*.

Для нахождения консервативных участков использовали компьютерную программу OWEN [5], позволяющую выравнивать две протяженные гомологичные нуклеотидные последовательности и выделять участки с заданной степенью совпадения. Для уточнения параметров выравнивания (размера окна и степени совпадения) было проведено сравнение геномных последовательностей локуса гена *even-skipped* из геномов перечисленных выше видов *Drosophila* с локусом «эталонного» вида *D. melanogaster*. Анализ результатов сравнения показал, что оптимальной для выделения консервативных блоков является пара *D. melanogaster*–*D. pseudoobscura*. С учетом требования по включению в находимые блоки известных сайтов связывания факторов транскрипции (ССФТ) было произведено уточнение параметров выравнивания и степени консервативности находимых отрезков: степень консервативности принята не ниже

Сокращение: ССФТ – сайт связывания факторов транскрипции, нт – нуклеотиды.



Относительное расположение консервативных участков и нуклеосом на ДНК на примере начального отрезка локуса гена *even-skipped* генома *D. melanogaster*. Карта локуса представлена в виде сегментов длиной 170 нт, отображенных один под другим (начало локуса – сверху); серые полосы – консервативные участки, черные вертикальные линии – центры экспериментально определенных мест посадки нуклеосом. Слева – разметка функциональных элементов локуса: энхансеров (e11, str3+7, str2), проксимального промотора (TATA) и кодирующего сегмента (CDS).

90% (совпадение девяти нуклеотидов (нт) в окне, равном 10 нт); значения остальных параметров выравнивания совпадали со значениями по умолчанию. Для выбранной пары видов *Drosophila* положение консервативных участков определено для генов группы *pair-rule* [6]: *even-skipped*, *hairy*, *odd-skipped*, *paired*, *runt*, *fushi tarazu*, *odd-paired*, *sloppy paired* и *ten*. Положение ССФТ и известных регуляторных и кодирующих сегментов в локусах перечисленных генов *Drosophila* взято из работы [2].

Найденные консервативные участки имеют протяженность в среднем от 30 до 70 нт и по длине находятся на промежуточном уровне между сайтами связывания факторов транскрипции (обычно около 7–10 нт) и нуклеосомным повтором (165–210 нт). Эти последовательности, в сумме занимая не более половины общей длины энхансера, включают в себя большинство ССФТ. Неконсервативные участки последовательностей также обычно не похожи на случайные и, по-видимому, могут нести какую-то функциональную нагрузку.

Во второй части работы положения найденных консервативных участков сравниваются с местами локализации нуклеосом на ДНК локусов изучаемых генов, определенными в работе [3].

Для отображения найденных консервативных участков для каждого из перечисленных генов построена карта локуса – числовая последовательность, содержащая ненулевые значения в позициях, лежащих внутри участков, и нули в остальных местах. На карту локуса также нанесены положения центров локализации нуклеосом.

Для установления близости периодов расположения консервативных участков и мест локализации нуклеосом карты локусов разделяются на последовательные сегменты длиной 170 нт; эта величина близка к минимальной длине нуклеосомного повтора [7]. Полученные сегменты располагаются один над другим. Результат такой сегментации для начального участка локуса гена *even-skipped* генома *D. melanogaster* приведен на рисунке. Сравнение картин сегментации локусов указанного и остальных генов позволяет выделить общие характерные черты.

Во-первых, как консервативные блоки, так и места локализации нуклеосом образуют сгущения (кластеры) в местах расположения известных функциональных элементов локуса.

Во-вторых, картины размещения как консервативных элементов, так и мест локализации нуклеосом образуют взаимосогласованный паттерн. Этот паттерн близок к периодическому с периодом минимального нуклеосомного повтора: границы блоков из смежных периодов приблизительно выравниваются по вертикали. Это позволяет сделать вывод, что ДНК, входящая в состав функциональных элементов локуса, упакована в нуклеосомы. Картина подобного чередования с периодом, близким к нуклеосомному, наблюдалась авторами данной работы и для других участков генома – экзонов и интронов кодирующих областей коллагеновых генов [9].

Наконец, на длине 170 нт обычно помещаются два консервативных блока и две промежуточные неконсервативные вставки. Характерное расстояние между соседними консервативными участками составляет примерно 84 нт – длину витка суперспирали ДНК, входящей в нуклеосому [6,7]. Сходство этих длин позволяет утверждать, что указанные участки расположены на соседних витках нуклеосомной суперспирали, и позволяет назвать их «синфазными блоками».

Характерная длина триплета, состоящего из двух последовательно расположенных синфазных блоков, разделенных неконсервативной вставкой, составляет примерно 1,5 витка суперспирали ДНК, входящей в нуклеосому, что лишь немногим меньше полной длины такой ДНК (1,65–1,8 витка [7,8]). Таким образом, ДНК такого триплета почти полностью «закрывает» гистоновое ядро нуклеосомы. Центральный неконсервативный участок оказывается расположенным приблизительно на оси симметрии нуклеосомной частицы, в месте связывания одной нити ДНК. Синфазные блоки локализируются на противоположной стороне гистонового ядра, связывающего две нити ДНК, и оказываются, таким образом, сближенными в пространстве.

Обращение к одной из общепринятых трехмерных моделей нуклеосомы, включающей в себя также гистон H1, стабилизирующий нити

ДНК на краях нуклеосомной частицы [10], позволяет объяснить неконсервативность центральной части триплета и, как следствие, отсутствие ССФТ на этом участке его экранированием белком H1.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, гранты 14-04-90034-Бел_а, 12-04-01776-а.

СПИСОК ЛИТЕРАТУРЫ

1. F. Antequera, *Cell. Mol. Life Sci.* **60**, 1647 (2003).
2. A. P. Lifanov, V. J. Makeev, A. G. Nazina, and D. A. Papatsenko, *Genome Res.* **13**, 579 (2003).
3. T. N. Mavrich, et al., *Nature* **453**, 358 (2008).
4. J. Touchman, *Nature Education Knowledge* **3** (10), 13 (2010).
5. A. Y. Ogurtsov, M. A. Roytberg, S. A. Shabalina, and A. S. Kondrashov, *Bioinformatics* **18**, 1703 (2002).
6. L. Wolpert and C. Tickle, *Principles of Development* (Oxford University Press, 2002).
7. G. Felsenfeld and M. Groudine, *Nature* **421** (6921), 448 (2003).
8. L. Marino-Ramirez, M. G. Kann, B. A. Shoemaker, and D. Landsman, *Expert Rev Proteomics* **2**, 719 (2005).
9. А. П. Лифанов, П. К. Власов, В. Ю. Макеев и Н. Г. Есипова, *Биофизика* **53**, 524 (2008).
10. N. Happel and D. Doenecke, *Gene* **431** (1–2), 1 (2009).

“Co-Phased Blocks” – Conservative Regions of Double-stranded DNA of Transcriptional Regulatory Modules – are Close in Space due to Phasing at Nucleosome DNA Super Helix

A.P. Lifanov*, V.J. Makeev**, and N.G. Esipova*

**Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova 32, Moscow, 119991 Russia*

***Vavilov Institute of General Genetics, Russian Academy of Sciences, ul. Gubkina 3, Moscow, 119991 Russia*

Conservative regions varying in length from 30 to 70 nucleotides (nt) are revealed as a result of interspecies comparison of homologous nucleotide sequences of the loci, genes controlling early development in *Drosophila* species. These regions are observed at the intermediate level between transcription factor binding sites (usually about 7–10 nt) and the nucleosome repeat unit (165–210 nt). Segments found are located mainly in the area of known functional elements of the locus (*cis*-regulatory modules: enhancers, a proximal promoter, a coding segment). Conservative domains occupy less than a half of the full enhancer length and contain practically all annotated transcription factor binding sites. A quasi periodical pattern of conservative region disposition is in accord with the experimental nucleosome localization pattern. The distance between neighboring domains is about 84 nt, i.e. a pitch of nucleosome DNA super helix. A repeatable accuracy of the distance permits of treating the delineated conservative regions as co-phased blocks and stating that these domains are close in the space being located at neighboring coils of the nucleosome DNA super helix.

Key words: enhancer, transcriptional regulatory module, regulatory transcription factor binding sites, nucleosome