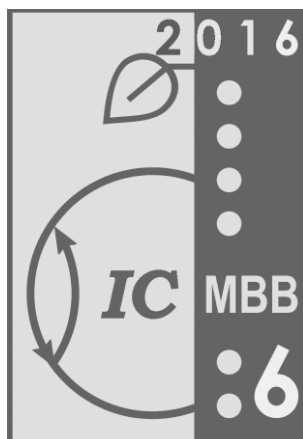


**РОССИЙСКАЯ АКАДЕМИЯ НАУК
«ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
ИМЕНИ М.В.КЕЛДЫША»
ИНСТИТУТ МАТЕМАТИЧЕСКИХ ПРОБЛЕМ БИОЛОГИИ**



МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ И БИОИНФОРМАТИКА

ДОКЛАДЫ VI МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ

Пушино, 16-21 октября 2016 г.

Под редакцией д.ф.-м.н. В.Д.Лахно

Пушино 2016

Математическая биология и биоинформатика: Доклады VI Международной конференции.

16-21 октября 2016 г., г. Пущино

В сборнике представлены доклады VI-й Международной конференции «Математическая биология и биоинформатика», проводимой Институтом математических проблем биологии РАН при участии Научного совета по математической биологии и биоинформатике РАН. Конференция проводится при финансовой поддержке Российского фонда фундаментальных исследований (проект 16-07-20497).

Состав Оргкомитета

Лахно В.Д., д.ф.-м.н., – председатель
Устинин М.Н., д.ф.-м.н. - зам.председателя
Махортых С.А., к.ф.-м.н. – ученый секретарь
Александров А.А., д.б.н.
Арсеньев А.С., д.х.н.
Васин А.А., д.ф.-м.н.
Ильин В.А., д.ф.-м.н.
Козлов Н.Н., д.ф.-м.н.
Коротков Е.В., д.б.н.
Марков А.В., д.б.н.
Миронов А.А., д.ф.-м.н.
Назипова Н.Н., к.ф.-м.н.
Ризниченко Г.Ю., д.ф.-м.н.
Романюха А.А., д.ф.-м.н.
Самсонова М.Г., к.б.н.
Туманян В.Г., д.ф.-м.н.
Шабанов Б.М., к.т.н.
Шайтан К.В., д.ф.-м.н.

Редакционная группа:

Махортых С.А., к.ф.-м.н.
Назипова Н.Н., к.ф.-м.н.
Якушева А.А.
Махортых Е.С.

Состав Программного комитета

Журавлёв Ю.И., д.ф.-м.н., академик РАН –
председатель
Колчанов Н.А., д.б.н., академик РАН -
зам.председателя
Рудаков К.В., д.ф.-м.н., член-корреспондент РАН -
зам.председателя
Арчаков А.И., д.б.н., академик РАН
Бердышев В.И., д.ф.-м.н., академик РАН
Дегерменджи А.Г., д.ф.-м.н., академик РАН
Евтушенко Ю.Г., д.ф.-м.н., академик РАН
Жижченко А.Б., д.ф.-м.н., академик РАН
Левин В.К., д.ф.-м.н., академик РАН
Матвеев С.В., д.ф.-м.н., член-корреспондент РАН
Рубин А.Б., д.ф.-м.н., член-корреспондент РАН
Федотов А.М., д.ф.-м.н., член-корреспондент РАН
Фрисман Е.Я., д.б.н., член-корреспондент РАН
Четверушкин Б.Н., д.ф.-м.н., академик РАН
Шокин Ю.И., д.ф.-м.н., академик РАН

DNAStructProfiler: автоматизированное построение профилей консервативности вторичных структур ДНК/РНК

Гречишникова Д.А.¹, Попцова М.С.¹

Московский Государственный Университет им. М.В.Ломоносова, физический факультет¹

maria.poptsova@gmail.com

DNAStructProfiler: An automated pipeline for reconstruction of DNA/RNA secondary structures conservation profiles

Grechishnikova D.¹, Poptsova M.¹

Lomonosov Moscow State University, Physics Department¹

A usual way to determine the evolutionary relatedness, or homology, of two DNA sequences is to search for traces of sequence conservation. However the lack of sequence conservation does not necessarily mean the lack of homology. Not detectable at the primary sequence level, it can be inferred when moving to the level of DNA/RNA secondary structures, which were shown to play an important role in many processes of genome functioning. Here we implemented DNAStructProfiler, an automated pipeline for reconstruction of DNA/RNA secondary structure conservation profiles that will allow researchers to reveal position-specific secondary structures in a set of DNA sequences. The tool can be used with any program that searches for DNA/RNA secondary structures, such as stem-loops, quadruplexes, triplex DNA, and any other structures of interest and is freely available at www.dnapunctuation.org/DNAStructProfiler.html. We demonstrate how the tool can be used to reveal evolutionary conserved stem-loop structures in human L1 retrotransposons.

1 Вступление

Для большинства функционально важных некодирующих последовательностей обнаруживается слабая или не обнаруживается вообще консервативность на уровне последовательностей. Однако белки безошибочно распознают точное местоположение некодирующих участков, для которых гомологичность на уровне первичной последовательности не выявлена. Примеры включают в себя распознавание транскрипционными факторами специфичных сайтов промоторных областей [1], распознавание сплайс-сайтов на границе интронов-экзонов [2], распознавание обратной транскриптазой транспозонной РНК [3], и многие другие. Исследования обнаружили, что функционально важные некодирующие последовательности часто содержат специфичные к местоположению, эволюционно консервативные вторичные структуры.

Со времен известного открытия Уотсоном и Криком структуры ДНК в виде двойной спирали, известной как каноническая правозакрученная В-форма, или В-ДНК конформация, последовали экспериментальные доказательства конформаций, отличных от В-ДНК конформации. Такие структуры включают в себя А-ДНК, Z-ДНК, триплексы, структуры стебель-петля (известные также как крестообразные структуры или шпильки), узловое ДНК, G4 тетрады (квадруплексы), и некоторые другие неканонические структуры [4]. Структуры стебель-петля представляют собой самый простой

структурный элемент, который образуется из палиндромных последовательностей, разделенных промежутком. Экспериментально было подтверждено, что структуры стебель-петля могут функционировать как транскрипционные терминаторы [5], аттенуаторы [6], промоторные элементы [7, 8], регуляторы РНК-сплайсинга [2] и рекомбинации [9].

До сих пор нет программных средств в открытом доступе, которые бы сравнивали последовательности на уровне вторичной структуры, так как методы сравнения и визуализации требуют дальнейшего развития. Мы разработали программное средство, DNAStructProfiler, которое позволяет исследователям обнаруживать консервативные позиции вторичных структур ДНК/РНК, такие как структуры стебель-петля, квадруплексы, триплексная ДНК, и любые другие структуры. Программный пакет с открытым кодом находится в открытом доступе на сайте по адресу www.dnapunctuation.org/DNAStructProfiler.html.

Функциональность разработанного программного продукта рассматривается на примере поиска консервативных позиций структур стебель-петля транспозонов L1, самого распространенного семейства транспозонов в геноме человека.

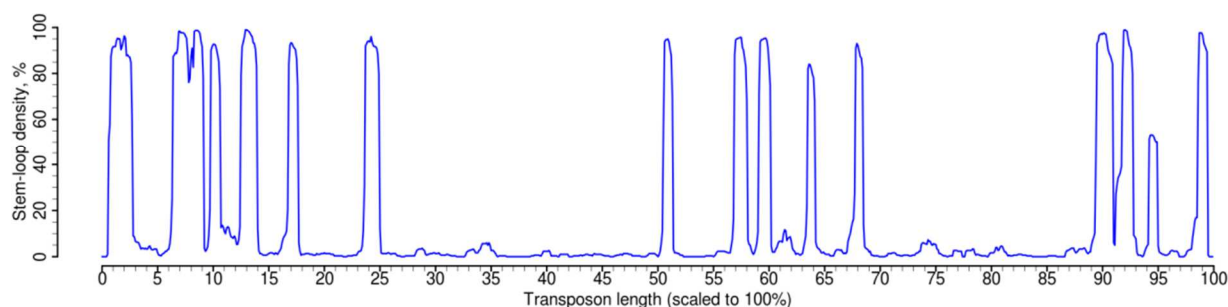


Рис.1. Профиль консервативности позиций структур стебель-петля транспозонов семейства L1P1.

2 DNAStructProfiler

Программа доступна в двух вариантах. Первый вариант создан для совместной работы с программой ДНК-пунктуация (<http://www.dnapunctuation.org/>) и осуществляет аннотацию исходных последовательностей структурами стебель-петля. Во втором варианте пользователь может предоставить любую другую аннотацию в стандартном формате bedGraph, который содержит координаты структур. Таким образом любая программа, аннотирующая последовательности вторичными структурами ДНК/РНК может работать с DNAStructProfiler. Предоставляемый пакет содержит обе версии с соответствующими примерами использования.

3 Профили структур стебель-петля для L1 транспозонов человека

Семейство транспозонов L1 – самое распространенное семейство LINE ретротранспозонов человека, присутствующих в более чем 500 000 копий, которые занимают приблизительно 17% полной длины генома человека [10]. Мы отобрали 27 семейств L1, полученных в результате филогенетического анализа, представленного в [11], аннотировали их структурами стебель-петля и построили профили покрытия с помощью программы DNAStructProfiler. На рис. 1 представлен профиль наиболее молодого семейства L1PA1, к которому принадлежат еще активные транспозоны. На рисунке четко видны консервативные позиции для структур стебель-петля для различных областей ретротранспозона, включая 5'UTR, ORF1 и ORF2, 3'UTR. Для обнаруженных консервативных позиций, программа предоставляет абсолютные координаты нахождения вторичных структур в исходной последовательности.

4 Заключение

Обнаружение консервативных позиций вторичных структур ДНК/РНК представляет интерес для эволюционных исследований последовательностей при отсутствии гомологии на

уровне первичной последовательности. Также данная информация может быть полезна экспериментаторам для проведения мутагенного анализа для выявления функции конкретной структуры в заданной позиции последовательности.

Список литературы

- [1] Butler, J.E., Kadonaga, J.T.: The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev*, 2002, **16**, 2583-2592.
- [2] Warf, M.B., Berglund, J.A.: Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem, Science*, 2010, **35**, 169-178.
- [3] Ohshima, K., Okada, N.: SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet Genome Res*, 2005, **110**, 475-490.
- [4] Smith, G.R.: Meeting DNA palindromes head-to-head. *Genes Dev*, 2008, **22**, 2612-2620.
- [5] Wilson, K.S., von Hippel, P.H.: Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci U S A*, 1995, **92**, 8793-8797.
- [6] Henkin, T.M., Yanofsky, C.: Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays*, 2002, **24**, 700-707.
- [7] Haasnoot, P.C., Brederode, F.T., Olsthoorn, R.C., Bol, J.F.: A conserved hairpin structure in Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. *Rna*, 2000, **6**, 708-716.
- [8] Wyrwicz, L.S., Gaj, P., Hoffmann, M., Rychlewski, L., Ostrowski, J.: A common cis-element in promoters of protein synthesis and cell cycle genes. *Acta Biochim Pol*, 2007, **54**, 89-98.
- [9] Bacolla, A., Wells, R.D.: Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog*, 2009, **48**, 273-285.
- [10] Cordaux, R., Batzer, M.A.: The impact of retrotransposons on human genome evolution. *Nat Rev Genet*, 2009, **10**, 691-703.
- [11] Khan, H., Smit, A., Boissinot, S.: Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates, *Genome Res*, 2006, **16**, 78-87.