

Анализ корреляции между данными CAGE и концами генов

Научный руководитель – Ставровская Елена Дмитриевна

Маргасюк Сергей Дмитриевич

Студент (специалист)

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

E-mail: smargasyuk@airmail.cc

CAGE (кэп-анализ экспрессии генов) - протокол секвенирования, предназначенный для прочтения коротких участков последовательности мРНК, начиная с кэпа [3]. Технология CAGE позволяет довольно точно определять старты транскрипции, а также уровень экспрессии генов. Однако при анализе картированных чтений CAGE для клеток крови человека (проект FANTOM 4) с помощью программы StereoGene обнаружено, что значительное число чтений расположено близко к концам генов, причем большая часть совпадает с концами генов с точностью до нуклеотида [5]. Представляет интерес как механизм кэпирования данных областей генома, так и функция полученных кэпированных мРНК.

Непонятно, почему происходит кэпирование данных участков, которые не являются стартами транскрипции. Согласно наиболее распространенной модели терминации транскрипции у эукариот, модели торпеды, после разрезания транскрипта и полиаденилирования полученного 3'-конца, РНК-полимераза II продолжает синтез мРНК. При этом высвободившийся при разрезании некэпированный 5'-конец узнается высокопроцессивной рибонуклеазой Xrn2 и разрушается ею [6]. Возможной альтернативой является кэпирование 5'-конца нового транскрипта. Например, рекэпирование показано для некоторых некодирующих РНК, не проходящих до конца путь нонсенс-опосредованного распада мРНК (NMD) [4].

Цель работы - выявить функцию кэпируемых РНК, синтезируемых с участков генома вблизи концов генов. Первичный анализ показал, что распределение числа ридов, покрывающих кластеры CAGE на концах генов, не отличается от соответствующего распределения для всех CAGE кластеров. Это свидетельствует в пользу того, что наблюдаемый эффект не является результатом экспериментальной ошибки. Мы также показали, что в то время как уровень экспрессии кластера CAGE в начале гена является мерой уровня экспрессии гена [2], для кластеров CAGE в конце гена отсутствует зависимость уровня экспрессии от уровня экспрессии соответствующих генов. Анализ перепредставленных GO-аннотаций с помощью сервиса David [1] выявил, что большинство генов связаны с альтернативным сплайсингом (p-value 2.2E-35).

Источники и литература

- 1) Huang, DW, Sherman, BT, Lempicki, RA 2009, 'Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources', Nature Protoc. 2009;4(1):44-57.
- 2) Kawaji, H et al. 2014, 'Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing', Genome Research 2014 Apr;24(4):708-17.

- 3) Kodzius, R et al. 2006, 'CAGE: cap analysis of gene expression', Nature Methods 3, 211 - 222 (2006)
- 4) Schoenberg, DR, Maquat, LE 2009, 'Re-capping the message', Trends in Biochemical Sciences, Volume 34, Issue 9, 435 – 442
- 5) Stavrovskaya, ED, Mironov, AA et al. 2016, 'StereoGene: Rapid Estimation of Genomewide Correlation of Continuous or Interval Feature Data', preprint
- 6) Watson J et al. 2013, 'Molecular Biology of the Gene (7th ed.)'