

Поиск пептидов, кодируемых lncRNA человека, на основании анализа данных масс-спектрометрии и рибосомального профилирования

Научный руководитель – Арапиди Георгий Павлович

Попов Алексей Алексеевич

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

E-mail: Popa_5_1@mail.ru

lncRNA, длинные некодирующие РНК, определяются как транскрипты длиной более 200 нуклеотидов, не транслируемые в белки [1]. В настоящее время имеются веские основания считать, что некоторые молекулы lncRNA всё же кодируют аминокислотные последовательности [2]. Продуктами их трансляции могут являться как полноразмерные белки, так и более короткие пептиды. Одной из причин, по которым эти транскрипты ошибочно аннотируют как некодирующие, является их машинная аннотация устаревающими инструментами, имеющими низкую чувствительность. Также в некоторых случаях ошибкой является использование консервативного определения белка, характеризующегося минимальной длиной последовательности, наличием канонического стартового кодона и необходимых сайтов на краях последовательности РНК, из чего вытекает порой безосновательное причисление РНК к некодирующим. В настоящий момент функции белков, кодируемых такими транскриптами изучены явно недостаточно. Данная область представляется интересной и важной, в том числе для медицины.

Целью данного проекта являлось исследование набора lncRNA человека, ранее опубликованных, как связывающиеся с рибосомами с последующей трансляцией [doi:10.1093/nar/gkz644]. Анализ проводился нами с помощью современных инструментов, использующих данные рибосомального профилирования (Ribo-seq) и масс-спектрометрии для наиболее точного предсказания кодирующего потенциала открытых рамок считывания. Результаты Ribo-seq загружали из базы данных RPFDB, а данные масс-спектрометрии - из базы данных PRIDE Archive.

В результате проверки на основании анализа данных Ribo-seq было найдено 237 lncRNA, для которых были предсказаны потенциальные кодирующие короткие открытые рамки считывания. Наличие в базе данных The Human Protein Atlas подтверждённых антителами белков для некоторых “lncRNA” позволило оценить корректность работы инструмента. Анализ масс-спектрометрических данных ранее опубликованных трех глобальных протеомных карт человека [doi:10.1021/acs.jproteome.8b00580, doi:10.1038/nature13302, doi:10.1038/nature13302] подтвердил гипотезу о трансляции с предсказанных коротких рамок считывания.

Все результаты исследования будут собраны в одну базу данных, где для каждого типа клеток или ткани организма будет произведена оценка транскрипции lncRNA и кодирующего потенциала соответствующих открытых рамок считывания, сопоставленные с наличием и количеством белка или пептида. Создаваемая база данных возможно позволит более полно изучить отличительные особенности кодирующих последовательностей и в дальнейшем позволит создать алгоритмы более точного предсказания кодирующей функции транскриптов с возможным применением для любых организмов.

Источники и литература

- 1) Derrien T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression // Genome research. – 2012. – Т. 22. – №. 9. – С. 1775-1789.

- 2) Lu S. et al. A hidden human proteome encoded by 'non-coding' genes // Nucleic acids research. – 2019. – Т. 47. – №. 15. – С. 8111-8125.