

МЕТОДЫ БИОИНФОРМАТИЧЕСКОГО АНАЛИЗА ВИРУСНЫХ МЕТАГЕНОМОВ КИШЕЧНИКА ЧЕЛОВЕКА

Е.В. Старикова, К.М. Климина, А.А. Хурумова, Д.С. Ульянов, Е.Н. Ильина

ФНКЦ ФХМ ФМБА России, Москва, Россия

Появление методов метагеномного секвенирования сделало возможным анализ и аннотацию некультивируемых микроорганизмов – так называемой “тёмной микробной материи”, а также ранее не известных вирусов. С помощью методов вирусной метагеномики был открыт известный бактериофаг CrAssphage [1], присутствующий в кишечнике большинства людей из различных стран и континентов. Микробиом кишечника человека является одной из наиболее разнообразных по видовому составу микроорганизмов экологических ниш, при этом многие присутствующие в нём виды остаются неизученными.

Классические референсные подходы не позволяют в полной мере оценить видовое разнообразие вирусов, поскольку известные вирусные последовательности составляют, по разным оценкам, от 5 до 20 % секвенируемых вирусных метагеномов. Необходимо расширение используемых биоинформатических инструментов для аннотации новых вирусных последовательностей и лучшего понимания взаимосвязей между вирусами и бактериями в кишечном микробиоме.

Целью нашей работы являлось получение вирусных метагеномов здоровых добровольцев в двух временных точках и биоинформатический анализ полученных вирусных последовательностей.

Из образцов кала здоровых добровольцев, собранных в двух временных точках, была получена фракция вирусоподобных частиц (ВПЧ) путём фильтрации и ультрацентрифугирования в градиенте CsCl и последующей обработкой ДНКазой. Затем было проведено секвенирование тотальной ДНК очищенной фракции вирусоподобных частиц на секвенаторе Illumina HiSeq 2500.

Полученные прочтения подвергались предобработке и фильтрации. Прошедшие фильтрацию прочтения аннотировались программой для таксономической аннотации метагеномов KRAKEN [2], а также картировались на базу известных бактериофагов, а также на базу метагеномных вирусных последовательностей IMG/VR. В результате данного анализа было показано наличие ряда известных фагов бактерий родов *Escherichia*, *Cronobacter* и *Bacteroides*, а также неаннотированных вирусных последовательностей бактерий родов *Parabacteroides*, *Bacteroides*, *Faecalibacterium* и др. Прошедшие фильтрацию прочтения также собирались в контиги *de novo* с помощью метагеномных сборщиков megahit [3] и metaspades [4]. Из собранных контигов отфильтровывались бактериальные контаминанты на основании определения последовательностей 16S.

В результате сборки были получены предположительные вирусные контиги (как линейные, так и циклические) длиной до 109 Кб. Полученные вирусные контиги аннотировались инструментом CAT/BAT. Белковые последовательности, предсказанные для данных контигов, аннотировались с помощью скрытых марковских моделей (HMM) из базы ортологичных вирусных генов pVOG [5], а также с помощью скрытых марковских моделей, сконструированных нами для белков вирусных семейств, не представленных в базе pVOG.

Для выявления предположительных бактерий-хозяев нами было проведено соответствие вирусных контигов спейсерам из CRISPR-систем известных бактерий, в результате чего было установлено полное соответствие ряда последовательностей спейсерам из геномов бактерий родов *Bifidobacterium*, *Roseburia*, *Streptococcus*, *Coprococcus* и *Acinetobacter*.

Дальнейшее изучение вирусных метагеномов кишечника человека и расширение существующих биоинформатических баз (в том числе баз метагеномных вирусных последовательностей, скрытых марковских моделей вирусных белков, баз спейсеров CRISPR-систем) способствует лучшему пониманию взаимосвязей между вирусами и бактериями в кишечнике человека.

Работа была проведена при поддержке гранта РФФИ № 19–34–80033

ЛИТЕРАТУРА

- Dutilh B.E. et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes //Nature communications. 2014. Т. 5. С. 4498
- Wood D.E., Salzberg S.L. Kraken: ultrafast metagenomic sequence classification using exact alignments //Genome biology. 2014. Т. 15. №. 3. С. R46.
- Li D. et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph // Bioinformatics. 2015. Т. 31. №. 10. С. 1674–1676.
- Nurk S. et al. metaSPAdes: a new versatile metagenomic assembler //Genome research. 2017. Т. 27. №. 5. С. 824–834.
- Grazziotin A.L., Koonin E.V., Kristensen D.M. Prokaryotic virus orthologous groups (pVOGs): a resource for comparative genomics and protein family annotation //Nucleic acids research. 2016. С. gkw975.