

## Секция «Биоинженерия и биоинформатика»

### Аннотирование гипотетических белков секвенированного субштамма *Mycobacterium bovis* BCG-Russia

Аксенова Е.И.<sup>1</sup>, Кунда М.С.<sup>2</sup>, Семенов А.Н.<sup>3</sup>

1 - ФГБУ "НИИЭМ им. Н.Ф.Гамалеи" Минздрава России, , 2 - ФГБУ "НИИЭМ  
им.Н.Ф.Гамалеи" Минздрава России, , 3 - НИУ ВШЭ, , Москва, Россия

E-mail: eaksenova@list.ru

Развитие технологии секвенирования нуклеиновых кислот позволило с 1995 г. получить большой объем данных о геномах бактерий, представленных только в базе GenBank более чем 17000 проектах. Однако при завершении работы, предполагающем не только сборку геномов, но и аннотирование, позволяющее построить карту метаболизма, возникает ряд трудностей, например, определение функции гипотетических белков.

На геномном анализаторе 454 GS Junior (Roche) секвенировали последовательность полного генома отечественного вакцинного субштамма *M. bovis* BCG Russia. Результаты прочтений позволили получить 44,5 кратное покрытие генома.

Сборку проводили с помощью пакета GS Junior Software 2.7 как *de novo*, так и по референсному штамму *M. bovis* BCG str. Токуо 172. В скаффолды собрано 91,4% генома. Анализ серверов для аннотирования показал, что требованиям для описания G+C богатых геномов микроорганизмов отвечает RAST (Rapid Annotation using Subsystem Technology, rast.nmpdr.org).

По данным RAST в последовательности BCG Russia найдено 4381 ORF, более 26% из которых кодируют гипотетические белки различной длины от 37 до 855 а.о. 92% гипотетических белков присвоен FIG (Fellowship for Interpretation of Genomes) номер, поскольку они имели аналоги в базе данных RAST, и только 8% были вновь зарегистрированы. После дополнительного анализа с помощью сервисов KEGG, PFAM, COGs, BLAST гипотетические белки разделили на две группы: с консервативными доменами (54%) и без них (46%). Две трети продуктов без доменов имели длину до 85 а.о., тогда как белки с консервативными участками в основном были длиннее: 200-500 а.о.. Среди консервативных доменов чаще всего выявляли домены ферментов разных классов, токсинов/ антитоксинов, факторов транскрипции, а также DUF домены (domain of unknown function). Таким образом, дополнительное исследование аннотированных гипотетических белков позволяет определить в их структуре смысловые домены, предположить роль белков в путях метаболизма микроорганизмов, выявить новые антигенные детерминанты и факторы, возможно, участвующие во взаимодействии со структурами клетки и патогенезе микобактерий.

Работа осуществляется в рамках гранта Президента Российской Федерации для государственной поддержки молодых российских ученых.

### Слова благодарности

Выражаю слова искренней благодарности своему научному руководителю О.Л. Ворониной за всестороннюю помощь при проведении и оформлении работы.