

Кластерный анализ транскрипционного ответа *R.sphaeroides*  
на изменение концентрации кислорода

Гудимчук Никита Борисович

Студент 4-го курса

Московский Государственный Университет им. Ломоносова, физический факультет  
nikita\_gb@mail.ru

*Rhodobacter sphaeroides* – факультативная фотогетеротрофная бактерия, отличающаяся высокой гибкостью своего метаболизма. В зависимости от внешних условий *R. sphaeroides* может использовать различные способы получения энергии: аэробное дыхание, фотосинтез, брожение [1]. Одним из важнейших сигналов, определяющих переключение этой бактерии с одного типа метаболизма на другой, является концентрация кислорода. Эффективным инструментом исследования таких переключений на уровне всего генома является технология генных микрочипов, разработанная в середине 90-х годов.

В настоящей работе выполнен статистический анализ результатов микрочиповых экспериментов, проведенных на кафедре молекулярной биологии Университета шт. Вайоминг (США). Эти результаты представляют собой значения экспрессии генов всех 4282 генов *R. sphaeroides* при шести различных концентрациях кислорода: 0,5, 2, 3, 10, 20 и 30% [2]. Для выделения генов с наибольшей вариабельностью значений экспрессии к исходным экспериментальным значениям экспрессии была применена процедура фильтрации SAM [3]. В результате был получен массив из 260 генов, для каждого из которых значение экспрессии по крайней мере при одной концентрации кислорода отлично от среднего значения экспрессии данного гена с уровнем значимости (FDR) не ниже 0,001. Отобранные дифференциально-экспрессированные гены после стандартной нормировки на нулевое среднее и единичную дисперсию были подвергнуты кластеризации двумя иерархическими методами – агломеративным и дивизивным (DiAna). Основанием для выбора указанных алгоритмов послужили высокая стабильность и детеминированность получающихся разбиений.

Качество полученных разбиений оценивалось методом SyR (Systematic Resampling), разработанным на кафедре биофизики физического факультета МГУ. В основе метода лежит последовательное применение процедуры бутстрепа к исследуемым данным в целях получения выборочных распределений параметров разбиения (числа классов, стабильности отдельных классов и т.п.). В результате применения метода SyR показано, что при варьировании числа классов от 2 до 16 наиболее достоверное разбиение массива дифференциально-экспрессированных генов агломеративным иерархическим методом кластеризации состоит из восьми классов, а дивизивным – из двух классов.

Полученные результаты определяют группы генов, сходным образом реагирующих на изменения концентрации кислорода во внешней среде, и могут служить отправной точкой для поиска участков связывания с ДНК регуляторных белков, ответственных за реакцию генома *R. sphaeroides* на изменение условий окружающей среды.

#### Литература

1. Bauer C. *et al.* (2003) *Phil. Trans. R. Soc. Lond., Ser. B*, **358**, 147–154.
2. Pappas T. *et al.* (2004) *J. Bacteriol.*, **184**, 4748–4758.
3. Tusher V.G. *et al.* (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 5116–5121.