

# СВЯЗЬ ФУНКЦИОНАЛЬНОГО РАЗНООБРАЗИЯ ПРОМОТОРОВ *E. COLI* И ИХ ПРОФИЛЕЙ ВЫЗВАННОЙ СУПЕРСПИРАЛИЗАЦИЕЙ ДЕСТАБИЛИЗАЦИИ ДУПЛЕКСА (SIDD)

Орлов М.А., Рясик А.А., Зыкова Е.А., Ермак Т.В.<sup>1</sup>, Сорокин А.А.

Институт биофизики клетки РАН

<sup>1</sup>Институт цитологии и генетики СО РАН

Современные методы секвенирования предоставили доступ к исключительно большой информации о первичной структуре ДНК, что делает необходимой ее автоматизированную обработку. Традиционно для аннотации геномов используют алгоритмы, анализирующие непосредственно нуклеотидную последовательность (генетический “текст”). Однако в случае предсказания положения регуляторных областей (в частности, промоторов) они демонстрируют недостаточно высокую специфичность. Это может быть связано с тем, что процессы регуляции реализуются при взаимодействии ДНК с белками в ходе молекулярного узнавания, которое определяется не последовательностью ДНК, а ее физическими свойствами. В связи с этим наиболее перспективны алгоритмы, анализирующие различные физические характеристики ДНК. В данной работе в качестве такой характеристики рассмотрена вызванная суперспирализацией дестабилизация дуплекса (SIDD, Stress-Induced Duplex Destabilization). Расчет SIDD позволяет количественно оценить влияние топологии на энергетику молекулы ДНК, определяющую вероятность плавления дуплекса по заданному положению. Ранее было показано, что профили SIDD промоторов, а также ряда других регуляторных областей имеют характерные пики [1].

В данной работе использован полный набор экспериментально подтвержденных промоторов *E. coli* (штамм K12) из базы данных RegulonDB версии 8.5. Вычисления SIDD выполнены для интервалов [-750;750] п.о. относительно точки старта транскрипции и значений плотности суперспиральности  $\sigma = 0,4; 0,5; 0,6; 0,7; 0,8$ . Использование разных значений  $\sigma$  может позволить выявить чувствительные к дестабилизации дуплекса группы промоторных последовательностей. Помимо этого, для наборов профилей SIDD, соответствующих каждому значению проведен кластерный анализ по методу Уорда. В результате получено 3 компактных устойчивых кластера, объединяющие профили с характерными элементами профилей. В обоих случаях для полученных группировок проведен анализ обогащения функциональными группами генов по GeneOntology.

Работа поддержана грантом РФФИ №16-37-00303 мол\_а.

## Литература.

1. Wang, H.Q. and Benham, C.J. Promoter prediction and annotation of microbial genomes based on DNA sequence and structural responses to superhelical stress // BMC Bioinformatics Vol. 7, 2006, pp. 248-262.