

Гриценко Екатерина Михайловна

канд. техн. наук, доцент, заведующая кафедрой

Бабатенко Дмитрий Геннадьевич

студент

Каверзин Евгений Викторович

аспирант

ФГБОУ ВО «Сибирский государственный университет

науки и технологий им. академика М.Ф. Решетнева»

г. Красноярск, Красноярский край

АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТЕЙ АПТАМЕРОВ В ПРОГРАММНОМ ОБЕСПЕЧЕНИИ FASTAPTAMER

Аннотация: в статье рассмотрена технология секвенирования аптамеров, проведен анализ последовательностей пула аптамеров в программном обеспечении Fastaptamer, выявлены достоинства и недостатки данного ПО.

Ключевые слова: аптамеры, ДНК, селекция, кластеризация, пул, FASTAptamer.

Аптамеры – это одноцепочечные молекулы ДНК, связывающиеся с нужными молекулами-мишениями. От последовательности образующих молекулу ДНК зависят геометрические формы, которые могут принимать молекулы аптамера, а значит, и его сродство к мишени.

На основе аптамеров создан ряд лекарственных средств, также их используют для изучения механизмов взаимодействия белков с нуклеиновыми кислотами и т. д. Стандартным методом получения аптамеров к необходимым мишениям является технология селекции. Технология секвенирования позволяет получать последовательности нуклеотидов для всех аптамеров из пула {A, T, G, C}.

После того, как аптамеры прошли технологию секвенирования методом SELEX, синтезатор и, в нашем случае, программное обеспечение в лаборатории бимолекулярных и медицинских технологий, возникает задача

биоинформационической обработки этого набора с целью выделения аптамеров, наилучшим образом связывающихся с мишенью.

Был проведен анализ информационного пула, содержащего в себе 83021 последовательностей аптамеров формата fasta (7.fasta). Содержимое файла представлено на рисунке 1, где первая строка – строка описания, а вторая сама последовательность.

```
1 >1-1297-2595.86
2 TCGCGGGTAGGGGGAGGGCCGAGGAGGCTGTAGGTGGGTG
```

Рис. 1. Исходный пул 7.fasta

Проанализируем данный набор сценария в программном обеспечении FASTAptamer, который использует в качестве метрики расстояние Левенштейна. Для анализа исследуем самые «распространенные» 800 последовательностей. Ниже приведены таблицы исследования обработки последовательностей аптамеров, с различными значениями расстояния Левенштейна (d), чтение фильтра (f) и значения кластера (c).

Таблица 1

Исследование обработки последовательностей аптамеров
с использованием различных значений расстояния Левенштейна

d	1	5	10	15	20	25	Количество цепочек
	1405 сек	1400 сек	1270 сек	45 сек	3 сек	3 сек	800

В данном примере представлен расчет 800 последовательностей аптамеров, которые обрабатывались, используя расстояние Левенштейна равным 1, 23.42 минуты на средней персональной машине, но таких последовательностей может быть и миллионы, и обработка таких данных может занять огромное количество времени, что в нашем случае неприемлемо. Поэтому возникает вопрос оптимизации данного метода.

Также были проанализированы разные значения расстояния Левенштейна (d), а также разбивка на кластеры (c) и использование фильтра чтения (f).

Проведя анализ полученных данных в программном обеспечении FASTAptamer были выявлены следующие недостатки:

- при указании значения фильтра, кластеры обрабатываются значительно быстрей, так как они используют в качестве критерия параметр reads, с суммарным чтением, превышающим указанное значение f. Но по данной картине нельзя увидеть, соседей, которые расположены ниже указанного значения f в пуле. А они могут существенно изменить весь анализ. Поэтому встает необходимость оптимизации данного метода;
- при обработке данных используя только расстояние Левенштейна без использования кластера и фильтра, возникает вопрос оптимизации алгоритма с целью сокращения времени обработки данных.

Список литературы

1. Вепринцев Д.В. Обработка результатов NGS-секвенирования пулов аптомеров [Текст] / Д. Вепринцев //ФГБОУ ВО Красноярский государственный медицинский университет имени проф. В.Ф. Войно-Ясенецкого.
2. Кульбачинский А.В. Методы отбора аптомеров к белковым мишням [Электронный ресурс]. – Режим доступа: <http://www.inbi.ras.ru/ubkh/46/kulbachinsky.pdf> (дата обращения: 10.12.2017).
3. Аптомеры к белкам: общие сведения [Электронный ресурс]. – Режим доступа: <http://medbiol.ru/medbiol/kulb/00024730.htm> (дата обращения: 10.12.2017).
4. FASTAptamer: A Bioinformatic Toolkit for High-Throughput Sequence Analysis of Combinatorial Selections [Электронный ресурс]. – Режим доступа: <http://burkelab.missouri.edu/fastaptamer.html> (дата обращения: 09.11.2017).