

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

DOI 10.21661/r-468091

АНАЛИЗ МЕР БЛИЗОСТИ ДЛЯ ОБРАБОТКИ ПУЛОВ АПТАМЕРОВ

Аннотация: в статье рассмотрена технология секвенирования аптамеров, рассмотрены существующие меры близости для обработки пулов аптамеров и выявлены их достоинства и недостатки.

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

Сложная структура нуклеотидов позволяет обеспечивать аптамерам связываться с белками. От последовательности образующих молекулу ДНК олигонуклеотидов зависит конформация (геометрические формы, которые могут принимать молекулы) аптамера, а значит, и его сродство к мишени (то есть сила связывания с молекулой выбранного исследователем белка).

Подобрав нужную последовательность из большой библиотеки аптамеров, ученые получают те, что им нужны, например, те, что связываются с клетками злокачественных опухолей. Благодаря способности связываться со своей мишенью аптамеры могут изменять функциональное состояние клетки [3].

Стандартным методом получения аптамеров к заданным мишениям является технология селекции SELEX (in vitro) [4].

Селекция начинается с генерации библиотеки, которая включает в себя 10^{12} различных последовательностей (видов аптамеров), а для решения поставленной задачи подходят лишь единицы.

Выбор аптамеров основан на скрининге большого числа последовательностей в библиотеках *in vitro*, благодаря чему удается подобрать аптамеры с высокой специфичностью практически к любой мишени.

В процессе отбора селекции происходит обогащение библиотеки последовательностями, имеющими повышенное сродство к мишени. Для получения аптамеров проводят от 8 до 16 раундов, т.е. пока связывание не ухудшится, обогащенную библиотеку клонируют и определяют последовательности индивидуальных аптамеров. Затем анализируют гомологию между индивидуальными аптамерами, по результатам анализа аптамеры группируют в несколько классов и оценивают их способность взаимодействовать с мишенью. Для дальнейших исследований отбирают последовательности с максимальным сродством к мишени. На следующем этапе определяют минимальный размер аптамера, необходимый для специфичного узнавания мишени. С этой целью синтезируют серию укороченных вариантов аптамера и определяют их способность связываться с мишенью [7].

Технология секвенирования позволяет получать последовательности нуклеотидов для всех аптамеров из пула. (Сначала пул – это библиотека, затем пул – это аптамеры, которые связались). Таким образом, каждому аптамеру сопоставляется слово в алфавите {A – аденин, T – тимин, G – гуанин, C – цитозин} [1].

Чтобы определить «похожесть» аптамеров, необходимо определить метрику. Имеющееся программное обеспечение в качестве метрики использует расстояние Хэмминга (Aptamercluster) или расстояние Левенштейна (FASTaptamer).

Получающиеся при этом кластеры содержат цепочки, «похожие» по всей длине [1].

Рассмотрим каждый из подходов:

1. Расстояние Хэмминга.

Расстояние Хэмминга – это количество различающихся позиций для строк с одинаковой длинной.

Расстояние Хэмминга уже довольно широко используется для различных задач, таких как поиск близких дубликатов, распознавание образов, классификация документов, исправление ошибок, обнаружения вирусов и т. д.

2. Расстояние Левенштейна.

Расстояние Левенштейна между двумя строками – это минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую.

Редакционное предписание – это последовательность операций (действий), необходимых для получения из первой строки второй кратчайшим образом. Обычно используют 4 действия: D (delete) – удалить, I (insert) – вставить, R (replace) – заменить, M (match)- совпадение.

При обработке строк, символы не всегда уместно считать числами, несмотря на то, что коды символов являются ими. Кроме того, часто требуется сравнивать строки разной длины. Поэтому для сравнения строк обычно используют метрики, оценивающие максимальную стоимость преобразования одной строки в другую. В общем случае, операциям редактирования, используемым в этом преобразовании, а именно замене символов, их вставке и удалению, можно назначить разные цены.

Расстояние Хемминга между двумя строками одинаковой длины определяется как число позиций, в которых символы не совпадают. Это эквивалентно минимальной цене преобразования первой строки во вторую в случае, когда разрешена только операция замены с единичным весом.

Если допускается сравнение строк разной длины, то, как правило, требуется также вставка и удаление. Если придать им тот же вес, что и замене, минимальная общая цена преобразования будет равна одной из метрик, предложенных Левенштейном.

Такие метрики применяются в качестве мер сходства в самых разных областях, например, для оценки степени гомологичности генетических последовательностей. Заметим, однако, что не все функции, измеряющие расстояние между строками, являются метриками.

Список литературы

1. Вепринцев Д.В. Обработка результатов NGS-секвенирования пулов аптамеров [Текст] / Д.В. Вепринцев; ФГБОУ ВО Красноярский государственный медицинский университет имени проф. В.Ф. Войно-Ясенецкого.
2. Кульбачинский А.В. Методы отбора аптамеров к белковым мишням [Электронный ресурс]. – Режим доступа: <http://www.inbi.ras.ru/ubkh/46/kulbachinsky.pdf>. (дата обращения: 10.12.2017).
3. Давыдова А.С. Эскорт-аптамеры: новые инструменты для направленной доставки лекарственных препаратов в клетки [Текст] / А.С. Давыдова, М.А. Воробьева, А.Г. Веньяминова // Медицина и здравоохранение. – 2006. – Т. 46. – С. 193–224.
4. Замай А.С. Технологии получения и использования ДНК-аптамеров для разработки новых средств диагностики и терапии [Текст] / А.С. Замай // Биохимия. – 2014. – С. 1–35.
5. Аптамеры к белкам: общие сведения [Электронный ресурс]. – Режим доступа: <http://medbiol.ru/medbiol/kulb/00024730.htm> (дата обращения: 10.12.2017).
6. AptamerCluster – A Method to Cluster HT-SELEX Aptamer Pools and Lessons from Its Application [Электронный ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4281958/>. (дата обращения: 09.11.2017).

7. Замай Т.Н. Новый перспективный способ идентификации возбудителя сальмонеллеза [Текст] / Т.Н. Замай, А.Б. Салмина, О.С. Замай // Вестник. новых медицинских технологий. – 2011. – №4. – С. 36–39.
8. 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).