

S04-27**АЛГОРИТМ СБОРКИ ТРАНСКРИПТОМА НА ОСНОВЕ АНАЛИЗА КОМПОНЕНТ СВЯЗНОСТИ ГРАФА ДЕ БРЁЙНА****Долганов В.О., Царев Ф.Н.****Университет ИТМО, Санкт-Петербург, Россия**

Технологии секвенирования позволили получать информацию о транскриптомах организмов. Знание транскриптома помогает в анализе экспрессии генов, открытии новых изоформ транскриптов и множестве других задач.

После работы секвенатора на выходе получаются чтения. Чтения бывают разного качества и различных видов. Одним из видов чтений являются парные чтения. Такие чтения содержат начала и концы различных участков транскриптов. Стоимость запуска секвенатора, выдающего парные чтения, меньше, чем секвенатора, выдающего полные чтения. Поэтому данная техника широко распространена. Задачей сборки транскриптома является восстановление исходных транскриптов по чтениям.

В случае отсутствия референса, используется сборка *de novo*. При такой сборке используются только чтения без дополнительной информации. Такая сборка является менее точной, но не требует информации об организме. Нами было разработано программное средство для *de novo* сборки транскриптома по парным чтениям. Преимущества подхода по сравнению с аналогами – скорость и экономия памяти. Алгоритм состоит из следующих этапов: 1. Анализ начальных чтений. Сбор информации о частоте k-меров, фильтрация ошибочных k-меров. 2. Построение графа де Брёйна и выделение в нем компонент связности без учета ориентации рёбер. 3. Разбиение больших компонент связности путем удаления из них ошибочных рёбер. Ошибочные рёбра определялись на основе информации о частоте k-меров соответствующих вершин. 4. Независимый анализ каждой полученной компоненты и выделение из них транскриптов. Выделение транскриптов основано на поиске путей в компоненте. При поиске путей так же учитывалась информация о частоте k-меров.

Разработанный алгоритм был реализован на языке программирования Java и сравнён с одним из популярных сборщиков транскриптома Trinity на различных данных. Во всех экспериментах предложенный метод работал в несколько раз быстрее и требовал меньшее количество памяти. Качество сборки предложенного метода зачастую уступало качеству сборки Trinity. В ряде случаев предложенный метод восстановил большее количество информации. Таким образом, предложенный метод может быть использован для быстрого получения предварительной сборки на системе с небольшим объемом памяти.