

ПРИМЕНЕНИЕ ГИПЕРГРАФОВ К ГАПЛОТИПИРОВАНИЮ ВИРУСОВ

Лебедев Р.В. Университет ИТМО, г. Санкт-Петербург

Научный руководитель – Алексеев Н.В., к.ф.-м.н., ведущий научный сотрудник ФИТиП, Университет ИТМО, г. Санкт-Петербург

Данная работа посвящена разработке нового метода гаплотипирования вирусов, базирующегося на применении теории гиперграфов.

Введение. Каждый организм описывается своим геномом, набором генетического материала, который можно представить в виде строки определенного размера из четырех чередующихся символов, соответствующих азотистым основаниям. По такой последовательности можно определить различные характеристики организма и использовать их для ряда прикладных задач.

В результате секвенирования биологических данных из генома прочитываются определенные подстроки, называемые ридами, которые выравниваются на референсный геном при его наличии. Далее задача заключается в кластеризации данных ридов таким образом, чтобы каждый кластер описывал отдельный гаплотип вируса из данных.

Существующие подходы к решению описываемой задачи основываются обычно на графовых методах (например, CliqueSNV), вероятностном выводе (например, aBayesQR) или их комбинации, и до сих пор нуждаются в повышении качества работы. Многие методы при построении своей математической модели теряют часть информации из исходных данных, хотя могли бы использовать ее для разрешения различных неоднозначностей. В данной работе производится попытка устранения этой проблемы.

Основная часть. Гиперграфом называется обобщение понятия графа, в котором ребро может соединять произвольное множество вершин и называется гиперребром.

Суть предлагаемого метода заключается в построении из ридов гиперребер, их последующем сжатии и выделении итоговых гаплотипов. А именно производится поиск всех позиций с мутациями в геноме. Каждый рид преобразуется в гиперребро, состоящее из покрываемых им мутационных позиций, а также ему назначается определенный вес, отражающий частотную представленность этого гиперребра на данном участке покрытия генома. Одинаковые гиперребра уплотняются, сливаясь в одно с соответствующим увеличением веса. Далее пересекающиеся непотиворечивые гиперребра по специально определенной метрике сливаются либо в одно, либо в несколько гиперребер с перераспределением частотных весов. В итоге получаем набор гаплотипов и оценки их представленности.

Выводы. Разработанный метод гаплотипирования позволяет извлекать из исходных данных более полную информацию и использовать ее для восстановления близкородственной популяции вирусов, которым специфична высокая скоростью мутации.

Лебедев Р.В.

Алексеев Н.В.