

УДК 004.891.2

РАЗРАБОТКА И РЕАЛИЗАЦИЯ АЛГОРИТМА ДЛЯ ДИАГНОСТИРОВАНИЯ ВОСПАЛИТЕЛЬНЫХ ЗАБОЛЕВАНИЙ КИШЕЧНИКА НА ОСНОВЕ НЕИНВАЗИВНЫХ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ.

Иванов А.Б. (Университет ИТМО)

Научный руководитель – канд. техн. наук Ульянов В.И. (Университет ИТМО)

Секция: Технологии программирования, искусственный интеллект, биоинформатика

Проведена разработка методов для диагностирования воспалительных заболеваний кишечника (ВЗК) на основе данных о микробиоте кишечника, полученных в результате неинвазивного взятия биологического материала у пациентов. Реализованная программа выделяет из метагеномных образцов компоненты ДНК, которые классифицируются на основе метагеномных данных, полученных из образцов пациентов с подтвержденным диагнозом. Данный метод может повысить точность и упростить постановку диагноза, что принесет пользу клиническому ведению и лечению ВЗК.

Введение.

Воспалительные заболевания кишечника (ВЗК) – термин, который включает в себя набор хронических воспалительных заболеваний желудочно-кишечного тракта. Двумя клинически определенными проявлениями ВЗК, имеющими отличительные клинические и патологические особенности, являются Язвенный колит (ЯК) и болезнь Крона (БК). ВЗК является глобальной болезнью с максимальным уровнем заболеваемости в Европе и Северной Америке (>0,3 %) и ростом заболеваемости в новых индустриальных странах.

Золотым стандартом для диагностики и мониторинга ВЗК является эндоскопия. Для подтверждения диагноза используется биопсия при колоноскопии, которая может дополняться измерением биомаркеров, таких как фекальный калпротектин, С-реактивный белок и маркеры сывороточных антител. Однако биопсия является инвазивной процедурой, а биомаркеры имеют высокую вариативность и не могут использоваться в качестве единственного признака для постановки диагноза. Поэтому стоит задача выявления новых маркеров и разработки новых методов, которые помогут повысить точность и упростить постановку диагноза, что принесет пользу клиническому ведению и лечению ВЗК.

Основная часть.

ВЗК включает в себя сложные генетические нарушения, однако на развитие заболевания оказывают влияние и другие факторы, такие как диета и микробиота. Микробиом человека состоит из множества микроорганизмов, которые колонизируют определенные ниши в человеческом теле (кишечник, ротовая полость, кожные покровы). Состав микробиома зависит от внешних факторов, включающих в себя среду обитания, прием лекарственных средств, и может меняться со временем, однако бесспорным является тот факт, что он играет важную роль в метаболизме и иммунном ответе организма различным заболеваниям. Текущие исследования выявляют связь между состоянием микробиоты кишечника и ВЗК, однако выявление конкретных маркеров, указывающих на болезнь, и причинно-следственной связи между этими двумя событиями являются открытыми на данный момент вопросами.

Для извлечения информации из образцов микробиоты кишечника человека используются секвенаторы второго поколения. Результатом работы секвенаторов является текстовый файл, содержащий информацию о ДНК организмов, содержащихся в образце, в виде набора коротких (100 – 300 пар нуклеотидов) прочтений.

Предлагаемое решение основано на выделении из исследуемого метагеномного образца длинных неветвящихся путей, которые могут соответствовать линейной части какого-либо

микрорганизма из образца, и их классификации на основе метагеномных данных от пациентов с подтвержденным диагнозом.

Алгоритм работы программы можно разделить на несколько этапов:

1. Разбиение всех прочтений из исследуемого метагеномного образца на k -меры – строки длиной k символов, и построение по ним графа де Брейна. Граф де Брейна содержит в качестве вершин последовательности нуклеотидов длиной k символов. В графе существует ребро между вершинами A и B , если соответствующие последовательности совпадают на длину $(k-1)$ нуклеотида ($A[2\dots k] = B[1\dots k-1]$).
2. Из построенного графа де Брейна выделяются все неветвящиеся пути, длина которых превышает минимальную указанную длину. Каждый такой путь сохраняется и рассматривается в дальнейшем как отдельная компонента.
3. Для набора метагеномов с известными диагнозами производится выделение k -меров из каждого метагенома и рассчитывается относительное покрытие каждой компоненты k -мерами. В результате каждой компоненте сопоставляется численный вектор степени ее представленности в различных образцах.
4. С помощью программы Mash на основании k -мерного спектра определяется близость между исследуемым образцом и известным набором образцов. Это необходимо в связи с тем, что пациенты, на которых оказывают влияние схожие внешние факторы, будут иметь схожий метагеномный состав и, вероятно, одинаковые принципы возникновения ВЗК.
5. Полученные на шаге 3 вектора представленности умножаются на вектор близости и сортируются на основании степени корреляции с заболеваниями (положительная или отрицательная связь с ЯК, БК, отсутствием ВЗК).
6. На основании полученных маркеров пациенту классифицируется в одну из трех категорий: язвенный колит, болезнь Крона, отсутствие ВЗК.

Полученные результаты могут быть использованы врачами для определения дальнейших методов исследования и лечения пациента.

Выводы.

Предложенный алгоритм использует информацию, полученную из образцов пациента в результате неинвазивного метода исследования, и направлен на повышение точности диагностирования ВЗК. На данном этапе обработана метагеномная база от 390 пациентов с известным статусом (174 ЯК, 93 БК, 123 отсутствие ВЗК). Разработанный алгоритм будет применен для постановки диагноза 105 пациентам в «Исследовании по метагеномной диагностике при воспалительных заболеваниях кишечника» (Metagenomics Diagnosis for Inflammatory Bowel Disease Challenge), направленном на определение потенциала метагеномных данных в постановке диагнозов ВЗК.

Иванов А.Б. (автор)

Ульянцев В.И. (научный руководитель)
