

Разработка программного конвейера с целью мутационного профилирования при онкологии на примере данных рака кишечника

Научный руководитель – Власенкова Рамиля Ахметовна

Кудрявцева Анастасия Александровна

Студент (бакалавр)

Казанский (Приволжский) федеральный университет, Институт фундаментальной медицины и биологии, Кафедра биоинформатики и медицинской кибернетики, Казань, Россия

E-mail: anastasiaflex@mail.ru

Исходя из данных Всемирного Фонда Исследования Рака, колоректальный рак является одним из трех наиболее распространенных типов. Было насчитано 1,8 миллион заболевших в 2018 году. Системный подход к анализу данных помогает изучить ключевые онкогены и находить взаимосвязи с мутациями в других генах.

Для реализации поставленной задачи были собраны данные из следующих баз данных: cBioPortal, ICGC, GENIE. Данные содержали в себе клиническую информацию, а также информацию по мутациям. Далее данные были объединены в общую таблицу для проведения последующего анализа. В качестве основного инструмента применялась программа RStudio. Используя собранные данные, с помощью машинного обучения, был предсказан отсутствующий микросателлитный статус. В качестве подбора критериев модели использовался метод 10-кратной кросс-валидации. Для сравнения результатов, микросателлитный статус был предсказан с помощью анализа распределения выборки. Выборка анализировалась по критериям нормального распределения, хи-квадрату, распределению Пуассона и распределению Рэлея. По лучшему результату строился график, показывающий порог по количеству мутаций. По найденному порогу определялся перцентиль, в дальнейшем помогающий классифицировать статус пациента по количеству найденных у него мутаций. Последним этапом проводился анализ встречаемости. Для этого использовался точный тест Фишера, bootstrap, а также метод отношения правдоподобия.

Таким образом, был разработан алгоритм и программный конвейер для определения мутационного профиля пациентов с использованием методов машинного обучения, статистических критериев, а также метода отношения правдоподобия.

Источники и литература

- 1) Gupta P, Chiang S, Sahoo PK, Mohapatra SK, You J, et al. Prediction of Colon Cancer Stages and Survival Period with Machine Learning Approach // Cancers. 2019. С. 2-16.
- 2) Salod Z, Singh Y. Comparison of the performance of machine learning algorithms in breast cancer screening and detection: A protocol// J. Public Health Res. 2019. Vol. 8(1677). С.112-118.