

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к дипломной работе

**Разработка классификатора лекарственной устойчивости
на основе анализа полногеномных данных микобактерии
туберкулеза**

Шабан Светлана Леонидовна

Научный руководитель – член-корреспондент НАН Беларуси,
доктор физико-математических наук, профессор
Тузиков А.В.

Консультант – научный сотрудник ОИПИ НАН Беларуси
Сергеев Р.С.

Минск, 2020

Реферат

Дипломная работа, 50 страниц, 22 рисунка, 9 таблиц, 9 формул, 29 источников.

МИКОБАКТЕРИЯ ТУБЕРКУЛЁЗА, ЛЕКАРСТВЕННАЯ УСТОЙЧИВОСТЬ, СЕКВЕНИРОВАНИЕ, КЛАССИФИКАЦИЯ, МАШИННОЕ ОБУЧЕНИЕ.

Объект исследования – полногеномные данные, маркеры лекарственной устойчивости.

Цель работы – разработка алгоритма классификации лекарственной устойчивости полногеномных данных с помощью методов машинного обучения.

В ходе работы изучены методы полногеномного секвенирования и сборки генома *de novo*, методы машинного обучения для классификации данных. Проанализированы последние результаты по исследованию поставленной задачи. Опираясь на изученный материал, проведены эксперименты для классификации лекарственной устойчивости микобактерии туберкулёза для каждого препарата по-отдельности и для выявления мультирезистентного туберкулёза на основе анализа мутаций и k-меров. Проведено сравнение разработанных классификаторов, показана их эффективность. Описаны способы возможного улучшения классификаторов для дальнейших исследований.

Областью применения является альтернативный фенотипическому тесту быстрый способ предсказания лекарственной устойчивости.

Abstract

Diploma thesis, 50 pages, 22 figures, 9 tables, 9 formulas, 29 sources.

MYCOBACTERIUM TUBERCULOSIS, DRUG RESISTANCE, SEQUENCING, CLASSIFICATION, MACHINE LEARNING.

Object of research – whole-genome data, drug resistance markers.

Objective – development of algorithm for classifying drug resistance of whole-genome using machine learning methods.

Methods of whole-genome sequencing and sequence assembly, machine learning methods for data classification were studied. The latest research results obtained for the described problem were analyzed. Based on the studied material, experiments were performed to classify the drug resistance of Mycobacterium tuberculosis for each drug separately and to identify multi-resistant tuberculosis based on the analysis of mutations and k-mers. Developed classifiers were compared and their effectiveness was shown. Methods of possible improvement of classifiers were described for further research.

The scope is fast way to predict drug resistance, which is an alternative to the phenotypic test.