

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
Кафедра системного анализа и компьютерного моделирования**

**БУРЫЙ
Александр Викторович**

**ПРЕДСКАЗАНИЕ АЛЬТЕРНАТИВНЫХ ТРАНСКРИПТОВ
ОНКОГЕНОВ ЧЕЛОВЕКА**

Аннотация к магистерской диссертации

**Научный руководитель:
Кандидат физико-математических наук,
доцент Н. Н. Яцков**

Допущен к защите

«___» _____ 2022 г.

**Зав. Кафедрой системного анализа
и компьютерного моделирования
кандидат физ.-мат. наук, доцент В. В. Скакун**

Минск, 2022

РЕФЕРАТ

Магистерская работа состоит из реферата, 15 иллюстраций, 6 таблиц, списка литературы (26 источников), объем работы – 50 страниц.

Ключевые слова: гены человека, экзон, интрон, гибридный онкоген RUNX1/RUNX1T1, альтернативный сплайсинг, интеллектуальный анализ данных, метод главных координат, Big Data, Apache, Hadoop, Spark, Storm.

Магистерская диссертация посвящена исследованию модели предсказания альтернативных транскриптов онкогенов на основе методов анализа больших данных.

Объект исследования – альтернативные транскриптомы онкогенов, существующие методы предсказания альтернативного сплайсинга, фреймворки распределенных систем.

Целью исследования является исследовать и реализовать модели предсказания альтернативных транскриптов онкогенов на основе методов анализа больших данных.

В соответствии с целью исследования были получены следующие результаты:

1. Сделать обзор литературы по теме;
2. Исследовать методы интеллектуального анализа данных (метод главных координат, спектральный алгоритмы, подходы на основе вычисления расстояний);
3. Изучить инструменты анализа больших данных, программные средства и фреймворки R/Java/Scala/Hadoop/Spark для реализации алгоритмов анализа больших данных;
4. Рассмотреть существующие методы анализа экзонных последовательностей транскриптов онкогенов;
5. Разработать и реализовать алгоритмы кодирования и векторизации экзонных последовательностей транскриптов онкогенов, алгоритмов больших данных, анализа моделей предсказания альтернативных транскриптов онкогенов.
6. Разработка критериев качества моделей предсказания альтернативных транскриптов онкогенов.
7. Исследование разработанных алгоритмов на экспериментальных данных о модельных генах.
8. Предсказание альтернативных транскриптов онкогена RUNX1/RUNX1T1 с использованием разработанных алгоритмов.

Магистерская диссертация выполнена автором самостоятельно.

ABSTRACT

Master's thesis consists of an abstract, 15 illustrations, 6 tables, a list literature (26 sources), the volume of work - 50 pages.

Keywords: human genes, exon, intron, hybrid oncogene RUNX1 / RUNX1T1, alternative splicing, data mining, principal coordinate method, Big Data, Apache, Hadoop, Spark, Storm.

The master's thesis is devoted to the study of a model for predicting alternative transcripts of oncogenes based on big data analysis methods.

The object of research is alternative transcriptomes of oncogenes, existing methods for predicting alternative splicing, distributed systems frameworks.

The aim of the study is to investigate and implement models for predicting alternative transcripts of oncogenes based on big data analysis methods.

In accordance with the purpose of the study, the following results:

1. Make a literature review on the topic;
2. Explore methods of data mining (method of principal coordinates, spectral algorithms, approaches based on distance calculation);
3. To study big data analysis tools, R/Java/Scala/Hadoop/Spark software tools and frameworks for implementing big data analysis algorithms;
4. Consider existing methods for the analysis of exon sequences of oncogene transcripts;
5. To develop and implement algorithms for coding and vectorization of exon sequences of oncogene transcripts, big data algorithms, analysis of models for predicting alternative oncogene transcripts;
6. Development of quality criteria for models for predicting alternative transcripts of oncogenes;
7. Study of the developed algorithms on experimental data on model genes;
8. Prediction of alternative transcripts of the RUNX1/RUNX1T1 oncogene using the developed algorithms.

Master's thesis was done by the author independently.