

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

ЛИТВИНЕНКО
Виктория Леонидовна

**РАЗРАБОТКА АЛГОРИТМОВ ПРЕДСКАЗАНИЯ
ПОЯВЛЕНИЯ ТРАНСКРИПТОВ РНК ГЕНОВ ЧЕЛОВЕКА**

Аннотация к дипломной работе

Научный руководитель:
кандидат физико-математических наук,
доцент Яцков Николай Николаевич

Допущена к защите

«___» _____ 2022 г.

Зав. кафедрой системного анализа
и компьютерного моделирования,
кандидат физико-математических наук,
доцент В.В. Скакун

Минск, 2022

РЕФЕРАТ

Дипломная работа: 62 страницы, 22 иллюстраций, 10 таблиц, 30 источников, 6 приложений

РАЗРАБОТКА АЛГОРИТМОВ ПРЕДСКАЗАНИЯ ПОЯВЛЕНИЯ ТРАНСКРИПТОВ РНК ГЕНОВ ЧЕЛОВЕКА

АЛЬТЕРНАТИВНЫЙ СПЛАЙСИНГ, ДНК, РНК, ЭКЗОНЫ, ТРАНСКРИПТЫ, ЭКЗОННАЯ ГРАФ-МОДЕЛЬ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, МЕТОД ГЛАВНЫХ

КООРДИНАТ, ОНКОГЕН RUNX1/RUNX1T1

Объектом исследования являются экспериментальные данные об основных объектах, участвующих в процессе альтернативного сплайсинга: экзонах и результирующих транскриптах, а также химерный онкоген RUNX1/RUNX1T1.

Предмет исследования – алгоритмы предсказания появления транскриптов РНК генов человека.

Цель работы – разработать модель на основе алгоритмов интеллектуального анализа данных для предсказания появления альтернативных транскриптов молекул РНК онкогенов человека.

Методы исследования – алгоритмы интеллектуального анализа данных – главных компонент и координат, иерархическая кластеризация данных, сравнения текстовых последовательностей, поиска ассоциативных правил.

Проведен анализ модельных генов, представляющих собой набор данных, описывающих признаки экзонов и структуры транскриптов, которые принимали участие в процессе альтернативного сплайсинга, и созданных для его исследования в клетках. Разработана модель предсказания появления альтернативных транскриптов и применена к исследованию химерного онкогена человека RUNX1/RUNX1T1.

РЭФЕРАТ

Дыпломная праца: 62 старонкі, 22 ілюстрацый, 10 табліц, 30 крыніц, 6 прыкладанняў

РАСПРАЦОЎКА АЛГАРЫТМАЎ ПАДКАЗАНИЯ З'ЯЎЛЕННЯ ТРАНСКРЫПТАЎ РНК ГЕНАЎ ЧАЛАВЕКА

АЛЬТЭРНАТЫЎНЫ СПЛАЙСІНГ, ДНК, РНК, ЭКЗОНЫ, ТРАНСКРЫПТЫ, ЭКЗОННАЯ ГРАФ-МАДЭЛЬ, ІНТЭЛЕКТУАЛЬНЫ АНАЛІЗ ДАНЫХ, МЕТАД ГАЛОЎНЫХ КААРДЫНАТ, АНКАГЕН RUNX1/RUNX1T1

Аб'ектам даследавання з'яўляюцца эксперыментальныя даныя аб асноўных аб'ектах, якія ўдзельнічаюць у працэсе альтэрнатыўнага сплайсінгу: экзонах і выніковых транскрыптах, а таксама хімерны анкаген RUNX1/RUNX1T1.

Прадметныя даследаванні – алгарытмы прадказання з'яўлення транскрыптаў РНК генаў чалавека.

Мэта працы – распрацоўка мадэлі на аснове алгарытмаў інтэлектуальнага аналізу даных для прадказання альтэрнатыўных транскрыптаў малекул РНК анкагенаў чалавека.

Метады даследавання – алгарытмы інтэлектуальнага аналізу даных – галоўныя кампаненты і каардынаты, іерархічная кластарызацыя даных, параўнанне тэкставых паслядоўнасцяў, пошук асацыятыўных правілаў.

Праведзены аналіз мадэльных генаў, якія ўяўляюць сабой даныя, каторыя апісваюць прыкметы экзонаў і структуры транскрыптаў, якія ўдзельнічаюць падчас альтэрнатыўнага сплайсінгу, і створаных для яго даследавання ў клетках. Распрацавана мадэль выяўлення набору альтэрнатыўных транскрыптаў і выкарыстована для даследавання хімічнага анкагену чалавека RUNX1/RUNX1T1.

ABSTRACT

The project contains 62 pages, 22 illustrations, 10 tables, 30 sources, 6 applications

DEVELOPMENT OF ALGORITHMS FOR PREDICTING THE APPEARANCE OF RNA TRANSCRIPTS OF HUMAN GENES

Keywords: alternative splicing, DNA, RNA, exons, transcripts, exon graph
model, data mining, principal coordinates analysis, oncogene RUNX1/RUNX1T1

The object of the investigation is experimental data on the main objects, participation in the process of alternative splicing: exons and resulting transcripts, as well as the RUNX1/RUNX1T1 chimeric oncogene.

Case Studies – algorithms for predicting the RNA transcripts of human genes discovered. The aim of the work is to develop a model based on data mining algorithms for predicting alternative transcripts of human oncogene RNA molecules.

Research methods – data mining algorithms – principal components and coordinates, hierarchical data clustering, comparison of text sequences, search for association rules.

The analysis of model genes, which are a set of data describing the features of exons and the structure of transcripts that took part in the process of alternative splicing, and created for its study in cells, has been carried out. A model for predicting the appearance of alternative transcripts has been developed and applied to the study of the chimeric human oncogene RUNX1/RUNX1T1.