

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра молекулярной биологии**

Аннотация к дипломной работе

ШУМСКИЙ
Пётр Валерьевич

**АНАЛИЗ ДАННЫХ ДНК СЕКВЕНИРОВАНИЯ ГЕТЕРОГЕННЫХ
ПОПУЛЯЦИЙ НА ПРИМЕРЕ ОБНАРУЖЕНИЯ
СОМАТИЧЕСКИХ МУТАЦИЙ ПРИ ОСТРОМ МИЕЛОИДНОМ
ЛЕЙКОЗЕ**

Научный руководитель:
с.н.с. лаборатории молекулярно-
генетических исследований ГУ
«РНПЦ детской онкологии,
гематологии и иммунологии»
А. А. Мигас

Минск, 2020

РЕФЕРАТ

Дипломная работа: 38 страниц, 11 рисунков, 10 таблиц, 55 источников.

Ключевые слова: Высокопроизводительное секвенирование, панельное секвенирование, острый миелобластный лейкоз (ОМЛ).

Объект исследования: образцы ДНК, полученные из моноклкераных клеток (МНК) пациентов детского возраста с ОМЛ.

Цель: разработка автоматизированного решения (пайплайна) для анализа данных высокопроизводительного секвенирования по заданному генетическому профилю (141 ген) для пациентов с ОМЛ.

Методы исследования: физические (спектрофотометрия), молекулярно-генетические (выделение ДНК, полимеразная цепная реакция, высокопроизводительное секвенирование) и биоинформационные.

В ходе выполнения работы проведена экстракция ДНК из образцов МНК костного мозга 12 пациентов детского возраста с ОМЛ с последующим секвенированием 141 гена, ассоциированного с неоплазиями миелоидного ростка кроветворения. Проведена проверка качества полученных прочтений с учетом таких показателей, как значение качества определения по шкале Фред, ГЦ-состав, наличие технических последовательностей, длина прочтений. Для каждого этапа анализа подобраны программы и их конфигурации с учётом особенностей первичных данных.

В результате разработанное автоматизированное решение (пайплайн) для обработки биоинформационных данных, позволяющее из первичных прочтений, полученных в ходе высокопроизводительного секвенирования, получить набор аннотированных вариантов. С использованием разработанного пайплайна проанализировано 12 образцов. В итоге получен финальный набор из 64 аннотированных вариантов, в среднем 5.3 на образец ($\pm 4,4$), из них 41 однонуклеотидная замена (37 миссенс-мутаций, 4 нонсенс-мутаций) и 23 делеции.

РЭФЕРАТ

Дыпломная работа: 38 старонак, 11 малюнкаў, 10 табліц, 55 крыніц.

Ключавыя словы: высокапрадукцыйнае секвеніраванне, панэльнае секвеніраванне, востры міелабластны лейкоз (ВМЛ).

Аб'ект даследавання: узоры ДНК, атрыманыя з монануклеарных клетак (МНК) пацыентаў дзіцячага ўзросту з ВМЛ.

Мэта: распрацоўка аўтаматызаванага рашэння (пайплайна) для аналізу дадзеных высокапрадукцыйнага секвеніравання па зададзеным генетычным профілі (141 ген) для пацыентаў з ВМЛ.

Метады даследавання: фізічныя (спектрафатометрыя), малекулярна-генетычныя (вылучэнне ДНК, палімеразная ланцуговая рэакцыя, высокапрадукцыйнае секвеніраванне) і біяінфармацыйныя.

У ходзе выканання работы праведзена экстракцыя ДНК з узораў МНК касцявога мозгу 12 пацыентаў дзіцячага ўзросту з ВМЛ з наступным секвеніраваннем 141 гена, асацыяванага з неаплазіямі міелоіднага парастка крыватвору. Праведзена праверка якасці атрыманых чытанняў з улікам такіх паказчыкаў, як значэнне якасці вызначэння па шкале Фрэд, ГЦ-склад, наяўнасць тэхнічных паслядоўнасцяў, даўжыня чытанняў. Для кожнага этапу аналізу падабраныя праграмы і іх канфігурацыі з улікам асаблівасцяў першасных дадзеных.

У выніку распрацавана гатовае аўтаматызаванае рашэнне (пайплайн) для апрацоўкі біяінфармацыйных дадзеных, якое дазваляе з першасных чытанняў, атрыманых у ходзе высокапрадукцыйнага секвеніравання, атрымаць набор анатаваных варыянтаў. З выкарыстаннем распрацаванага пайплайна прааналізавана 12 узораў. У выніку атрыманы фінальны набор з 64 анатаваных варыянтаў, у сярэднім 5.3 на ўзор ($\pm 4,4$), з іх 41 аднануклеотыдная замена (37 міссэнс-мутацый, 4 нонсэнс-мутацый) і 23 дзялок.

SUMMARY

Diploma project: 38 pages, 11 figures, 10 tables, 55 sources.

Key words: High-throughput sequencing, targeted sequencing, acute myeloid leukemia (AML).

The object of the research: DNA samples obtained from mononuclear cells (MNCs) of pediatric patients with AML.

The aim of the research: development of an automated solution (pipeline) for analyzing high-throughput sequencing data for a given genetic profile (141 genes) for patients with AML.

The research methods: physical (spectrophotometry), molecular genetic (DNA isolation, polymerase chain reaction, high-throughput sequencing) and bioinformational techniques.

In the course of the work, DNA was extracted from samples of bone marrow MNCs of 12 pediatric patients with AML, followed by sequencing of 141 genes associated with neoplasia of the hematopoietic myeloid germ. The quality of the reads was verified taking into account such indicators as Phred-scaled quality scores, GC composition, the presence of technical sequences, the length of reads. For each stage of the analysis, tools and their configurations were selected taking into account the characteristics of the primary data.

As a result, a ready-made automated solution (pipeline) for processing bio-information data was developed. It allows to obtain a set of annotated variants from the raw reads obtained during high-throughput sequencing. Using the developed pipeline, 12 samples were analyzed. As a result, a final set of 64 annotated variants was obtained, on average 5.3 per sample (± 4.4), of which 41 single-nucleotide variants (37 missense mutations, 4 nonsense mutations) and 23 deletions.