

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

Аннотация к дипломной работе

МИШУК
ЕГОР АЛЕКСАНДРОВИЧ

ОБРАБОТКА ДАННЫХ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО
СЕКВЕНИРОВАНИЯ ДЛЯ ПРОВЕДЕНИЯ АНАЛИЗА
ВАРИАЦИЙ ЧИСЛА КОПИЙ В ГЕНОМЕ ЧЕЛОВЕКА

Научный руководитель:
кандидат биологических наук,
О.Д. Левданский

Минск, 2023

РЕФЕРАТ

Дипломная работа 41 с., 12 рис., 5 табл., 63 источника.

Ключевые слова: NGS, CNV, эпилепсия, невынашивание при беременности, нефропатология, некоронарогенные заболевания сердца, кардиопатология.

Объект исследования: данные высокопроизводительного секвенирования (NGS) 262 образцов ДНК. В выборке были представлены образцы со следующими патологиями: эпилепсия – 23 образца; некоронарогенные заболевания сердца – 31 образец; почечная недостаточность – 68 образцов; недоношенные дети – 55 образцов; болезни сердца – 13 образцов; без патологий – 72 образца.

Цель: провести анализ вариаций числа копий в данных расширенных клинических экзомов белорусов и обработать результаты анализа.

Методы исследования: биоинформатические (предобработка сырых данных секвенирования, выравнивание прочтений, CNV анализ).

В результате исследовательской работы был создан набор скриптов для CNV анализа данных высокопроизводительного секвенирования. С помощью разработанных скриптов проведен поиск CNV у 262 пациента из пяти выборок (некоронарогенные заболевания сердца, кардиопатологии других типов, невынашивание при беременности, эпилепсия, нефропатология). Определены множественные CNV, ассоциированные с каждым типом патологии. Выявленные CNV представляют собой потенциальные маркеры для определения предрасположенности к данным типам заболеваний. Показано, что наличие делеций, затрагивающих гены цинковых пальцев способствуют повышению риска развития невынашивания при беременности, в то время как дубликации в данных генах, наоборот, являются протективными в отношении данной патологии.

Выявлено, что дубликации, затрагивающие широкий спектр генов, могут являться маркерами риска развития эпилепсии, в то время как делеции некоторых генов, встречающиеся в человеческой популяции, у пациентов с эпилепсией практически не обнаруживаются.

Для некоронарогенной патологии сердца наиболее значимо ассоциированными локусами, оказались как делеции, так и преобладающие по количеству дубликации, с большей частотой встречающиеся в выборке пациентов.

РЭФЕРАТ

Дыпломная работа 41 старонка, 12 малюнкаў, 5 табліц, 63 крыніцы.

Ключавыя словы: NGS, CNV, эпілепія, невыношванне пры цяжарнасці, нефрапаталогія, некоронарогенныя захворванні сэрца, кардыяпаталогія.

Аб'екты даследвання: дадзеныя высокапрадукцыйнага секвеніравання (NGS) 262 узораў ДНК. У выбарцы былі прадстаўлены ўзоры з наступнымі паталогіямі: эпілепія - 23 ўзору; некаранарогенныя захворванні сэрца - 31 ўзор; нырачная недастатковасць - 68 узораў; неданошаныя дзеці - 55 узораў; хваробы сэрца - 13 узораў; без паталогій - 72 ўзору.

Мэта даследвання: правесці аналіз варыяцый колькасці копій у дадзеных пашыраных клінічных экзумаў беларусаў і апрацаваць вынікі аналізу.

Метады даследвання: біяінфарматычныя (прадапрацоўка сырых дадзеных секвеніравання, выраўноўванне чытанняў, CNV аналіз).

У выніку даследчай працы быў створаны набор скрыптоў для CNV аналізу даных высокапрадукцыйнага секвеніравання. З дапамогай распрацаваных скрыптоў праведзены пошук CNV ў 262 пацыента з пяці выбарак (некаранарогенныя захворванні сэрца, кардыяпаталогіі іншых тыпаў, невыношванне пры цяжарнасці, эпілепія, нефрапаталогія). Вызначаны множныя CNV, асацыяваныя з кожным тыпам паталогіі. Выяўленыя CNV уяўляюць сабой патэнцыйныя маркеры для вызначэння схільнасці да дадзеных тыпаў захворванняў. Паказана, што наяўнасць дзялок, якія закранаюць гены цынкавых пальцаў спрыяюць павышэнню рызыкі развіцця невыношвання пры цяжарнасці, у той час як дуплікацыі ў дадзеных генах, наадварот, з'яўляюцца пратэктывымі ў дачыненні да дадзенай паталогіі.

Выяўлена, што дуплікацыі, якія закранаюць шырокі спектр генаў, могуць з'яўляцца маркерамі рызыкі развіцця эпілепсіі, у той час як дзялок некаторых генаў, якія сустракаюцца ў чалавечай папуляцыі, у пацыентаў з эпілепіяй практычна не выяўляюцца.

Для некаранарогенных паталогій сэрца найболей значна асацыяванымі локусамі, апынуліся як дзялкі, так і пераважныя па колькасці дуплікацыі, з большай частатой сустракаемыя ў выбарцы пацыентаў.

ABSTRACT

Diploma project 41 p., 12 fig., 5 tables, 63 sources.

Key words: NGS, CNV, epilepsy, pregnancy failure, nephropathology, non-coronary heart disease, cardiomyopathy.

Subject: high-throughput sequencing (NGS) data from 262 DNA samples. The samples were from the following pathologies: epilepsy – 23 samples; non-coronary heart disease – 31 samples; renal failure – 68 samples; premature infants – 55 samples; heart disease – 13 samples; and no pathology – 72 samples.

The aim of the research: to analyse copy number variation in the extended clinical exome data of Belarusians and to process the results of the analysis.

The research methods: bioinformatics (preprocessing of raw sequencing data, read alignment, CNV analysis).

As a result of the research work, a set of scripts for CNV analysis of high-throughput sequencing data was created. The developed scripts were used to search for CNVs in 262 patients from five samples (non-coronary heart disease, other types of cardiopathology, pregnancy failure, epilepsy, nephropathology). Multiple CNVs associated with each type of pathology were identified. Identified CNVs are potential markers for the predisposition to these types of diseases. Deletions affecting CNV genes were shown to contribute to the risk of miscarriage while duplications in these genes were shown to be protective against this pathology.

Duplications involving a wide range of genes have been shown to be risk markers for epilepsy, while deletions of some genes found in the human population are virtually undetectable in patients with epilepsy.

For non-coronary heart disease, both deletions and predominantly duplications were found to be the most significantly associated loci, with a higher frequency of occurrence in a sample of patients.