MASTER'S THESIS IN BIOINFORMATICS

T-CELL CLONOTYPING FROM BULK WHOLE TRANSCRIPTOME RNA-SEQ DATA

Sergio Fernández Adán and Sergio Galera Alquegui
Bioinformatics Research Centre (BiRC),
Aarhus University

Master's thesis' supervisors:
Christian Storm Pedersen, BiRC - Aarhus University
Paula Tataru Bjerg, QIAGEN - Aarhus

TABLE OF CONTENTS

0. *Abstract*

Our immune system is a compound of different types of cells, tissues, and organs that help us to get protection against harmful infections and diseases. T-cells are one of the cells that take a role in this system, and it is defined as a cellular receptor associated with an intrinsic enzymatic response that acts in diverse intracellular pathways. Usually, we require targeted sequencing data to extract those sequences that have been classified as TCR data to start analysing the immune system of a sample. In this report, we present the *CLC Genomics Workbench* as a tool for immune repertoire extraction out of bulk RNA-seq data. We benchmarked our results against *MiXCR*, open-source software that recalls the correct results with high accuracy and precision and that has been widely used by scientists to extract immune data. We have obtained plenty of results from filtering data by applying different methods after correctly pre-processing all the data. In the process of creating a set of results, we have also developed an algorithm that will be explained in further detail in this report. Observed results show that most of the methods used have a proper performance relative to *MiXCR* protocols while others did not correctly extract all the repertoires.