

Designing analytical pipeline for identification of epigenetic marks significantly affecting gene expression patterns in rheumatoid arthritis patients.

Michał Dąbrowski
Instytut Podstaw Informatyki Polskiej Akademii Nauk
m.dabrowski@ipipan.waw.pl

Description of the subject

Rheumatoid arthritis (RA) is an autoimmune inflammatory disease characterized by inflammation of the synovium in the joint capsules. The number of patients increases every year, they struggle with burdensome symptoms and so far, there is no solution to completely cure RA patients. In the last decades more indications about the importance of DNA methylation in the course of the disease appeared and showed that the relationship between DNA methylation and gene expression should be further explored and may bring important answers.

The main goal of this project is to determine the impact of epigenetic marks, especially DNA methylation on the expression of genes related to rheumatoid arthritis (RA). Three areas of tasks are assumed: **1)** differential analysis of molecular features (NGS and single cell); **2)** generating new features (eg. chromatin states) based on the combination of various molecular data available in our group and publicly available repositories; **3)** comparative analysis between different algorithms dedicated to detect gene expression regulatory regions and developing new approaches;

First, based on public RNA-seq data the level of gene expression will be compared between RA confirmed patients and control group. For the returned differentially expressed genes (DEGs) the regulatory regions will be assigned and their epigenetic landscape compared. DEGs will be analyzed using Natural Language Processing methods and in result grouped into functionally similar groups. Epigenetic marks and genomic data will be used to learn how to predict various regulatory regions as well as to predict distribution of differentially expressed genes.

Requirements

- completed second-cycle studies in biological, mathematical or IT-related fields
- good knowledge of data analysis, statistics
- good knowledge of molecular mechanisms related to the regulation of gene expression
- very good command of the English language
- high programming skills in at least one language (e.g. R, Python)
- knowledge of Next Generation Sequencing is welcome

Before submitting the documents, I insist candidates contact me individually by email.