Discovery of the source of gene expression variation in cell populations – analysis of the associations between epigenetic signals and the three-dimensional structure of chromatin

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

Description of the subject

Thanks to the analysis of gene expression at the level of single cells (single-cell RNA-seq), it is possible to identify specific cell clusters - populations that differ in the pattern of gene expression. The aim of this project is to discover groups of genes with a characteristic expression profile for individual cell populations, and then to investigate the mechanisms modulating their expression at the level of epigenetic changes (e.g., DNA methylation, histone modifications) and the three-dimensional (3D) structure of DNA (Hi-C).

Implementation of this project includes identification and integration of data sets containing various types of single-cell data and data describing the three-dimensional structure of DNA. The expected effect of this step will be the assignment of genes and epigenetic modifications to individual domains, as well as a description of the 3D contacts between enhancers and promoters. In the next step, methods/algorithms dedicated for identification of cell populations based on single-cell RNA-seq data analysis, will be compared. As a results, marker features (genes) for individual cell populations will be identified. Afterwards, the expression pattern of identified marker genes, characteristic for individual cell populations, in individual domains, as well as the variability of epigenetic signals in regulatory regions assigned to these marker genes, will be analysed. One of the specific tasks at this stage will be to find gene promoters having confirmed 3D contacts with multiple enhancers, and then to establish an association between the ChIP-seq (H3K27ac) signal level of these enhancers and the variability of target gene expression. In this way, we will determine the relationship between enhancer activity and the level of gene expression.

The analysis of various single-cell data sets, e.g., ATAC-seq, DNA methylation will allow to determine gene expression regulatory mechanisms keeping homogenous expression level of genes within domains, as well as to unveil mechanisms specific for regulatory regions of marker genes identified for individual cell populations. In the absence of important single-cell data, we will use Next Generation Sequencing data obtained for a pool of cells. Finally, at an advanced stage of bioinformatics analyses, it is planned to prepare a grant application to validate the obtained computational results in an *in vitro* model.

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 shills in at least one language, such as R, Perl, Python
- knowledge of Next Generation Sequencing is welcome