

The use of machine learning methods to detect mutations in DNA sequences and changes in epigenetic signals important for the regulation of gene expression in selected diseases

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

Description of the subject

The goal of this interdisciplinary project is to uncover significant perturbations in regulatory regions that alter the level of expression of genes giving rise to a pathogenic state. Thanks to extensive molecular research over the past two decades, it has been shown that the analysis of protein coding regions cannot provide an answer to the causes of many diseases common to the human population. In this topic, we propose, on the basis of publicly available databases and our own databases, to verify which elements in regulatory areas are important in pathogenesis and to develop a network of connections between them.

The project involves the analysis of data from Next Generation Sequencing, including DNA sequences, transcriptome, methylome and a set of selected ChIP-seqs.

Research findings will help to uncover the mechanisms of pathogenesis and help to elucidate the dysfunction of the networks that control gene expression regulation and epigenetic linkages in these processes.

Requirements

- completed second-cycle studies in biological, mathematical or IT-related fields
- basic knowledge of data analysis, statistics
- good command of the English language
- knowledge of programming in any language, such as R, Perl, Python, is welcome