SILESIAN UNIVERSITY OF TECHNOLOGY

Faculty of Automatic Control, Electronics and Computer Science

Mathematical modelling in comparative analysis of methylation profiles of *de novo* and therapy-related AML

Doctoral Dissertation

by

Agnieszka Cecotka

Supervisors

Professor Joanna Polańska, PhD, DSc

Christophe Badie, PhD

2022

Gliwice, POLAND

Mathematical modelling in comparative analysis of methylation profiles of *de novo* and therapy-related AML

Doctoral Dissertation by **Agnieszka Cecotka**

Supervisors: **Professor Joanna Polańska, PhD, DSc**, **Christophe Badie, PhD**

Abstract

The expansion of high throughput experimental techniques leads to collecting a massive amount of data that must be processed. These data's characteristics are the huge number of features and the nontypical statistical distribution of the obtained signal. Hence, there is a need to develop new algorithms and pipelines to process them. One of phenomena measured with high throughput techniques is DNA methylation - epigenetic modification crucial for gene expression control and cancer development. Aberrations of DNA methylation in Acute Myeloid Leukaemia (AML) might be a reason for differences in treatment response and survival time between patients with *de novo* and therapy-related AML (AML being a side effect of previous malignancy treatment). Therefore, this thesis aims to develop pipelines for processing the DNA methylation data and investigate the DNA methylation profile in AML patients. The greatest approaches were developed, selected, and described from data preprocessing, statistical analysis, mathematical modelling, and functional analysis of detected features to validate the results with different experimental platforms.

Initially, the original algorithm for finding DNA methylation profile of AML is presented and implemented for data obtained with methylation microarrays. It is a composition of mathematical modelling and statistical approaches to conclude about low, medium, high and extremely high hyper- or hypomethylation of genome sites or regions. Subsequently, the detection of differences between genders in DNA methylation levels and survival factors in specific genomic regions in AML patients is investigated. It uses the integration of results obtained in a comparative and survival analyses. Moreover, the pipeline for detecting aberrations in DNA methylation in *de novo* and chemo- or radiotherapy-related AML is described. It is drawn upon supervised and unsupervised feature selection in epigenetics and functional analysis domains. The result is the detection of several biomarkers of therapy-related AML, confirmed in an independent, pyrosequencing experiment.