Rule-based integration of heterogeneous multi-omics data to determine the prognostic markers for relapse in Acute Lymphoblastic Leukemia

Aleksandra Gruca[1,2], Roman Jaksik[3], and Marek Sikora[2,4]

[1]EU COST Action CHARME CA15110
[2]Institute of Informatics, Silesian University of Technology, Gliwice, Poland
[3] Institute of Automated Control, Silesian University of Technology, Gliwice, Poland
[4] Institute of Innovative Technologies EMAG, Katowice, Poland

Public multi-omics repositories allow researchers to extract data for integration and analysis in order to discover the molecular bases of diseases and development of effective treatments. However, the diversity of biological systems, the technological limits, the large number of biological variables and the relatively low number of biological samples make the integration and analysis of multi-omics datasets a challenging task.

In this work we integrate the data from TARGET Acute Lymphoblastic Leukemia (ALL) Phase2 project. In particular, we combine clinical data, gene expression, DNA methylation and copy number variance to find new markers correlated with poor clinical outcome and early bone marrow relapse.

From the cohort of 792 patients we identified 80 patients for which all required multi-omics experimental data is available in TARGET database. In this group, 39 patients suffered from ALL relapse. For the analysis we designed rule-based framework. The main advantage of rule-based approach is the fact that the outcome rule set is not only useful as a classifier but also provides a natural means of understanding which features (and their combinations) influence the outcome. Presented rule-based workflow allows selecting the most important features discriminating between relapse and relapse-free patients providing prognostic multi-markers for relapse in ALL.