Progress of the biological sciences in the era of big data will depend on how we address the following question: '*How do we connect multiple disparate data types to obtain a meaningful understanding of the biological functions of an organism?*'

Multi-omics technologies are a relatively new field of research. They include genomics, transcriptomics (mRNA expression), proteomics and metabolomics. By combining data from various omics technologies it is possible to obtain a more complete picture of complex molecular events and, above all, to increase understanding of how healthy and disease-changed cells function.

High-throughput omics technologies are generating large volumes of multi-omics data at an unprecedented rate. Specialized computational approaches are required to effectively and efficiently carry out the predictions using biomedical data acquired from diverse modalities. Unfortunately, the overwhelming majority of currently developed systems focus on complex decision rules that are obstacles to mature applications. Little emphasis on transparency that emerges from standard machine learning impedes biological understanding, in particular, any mechanistic interpretation. It can be observed that there is a strong need for 'white box', comprehensive prediction models which may actually help in understanding and identifying relationships between specific features and improve biomarker discovery.

The research objective is to test the hypothesis that relative hierarchical and horizontal dependencies analysis can be successfully applied for multi-omics prediction and diagnosis support. In order to verify this, a new ML-empowered integrative analysis focused on finding comprehensible predictions and potential 'omics'-based biomarkers for clinical use will be proposed. The general idea is to find relations between the features within a single individual. Relative dependencies that occur in each sample will be confronted to the rest of the data and the relations will be ranked according to their consistency, complexity and discriminative power. This concept will constitute a new cross-omics prediction model called Relative Dependencies Analysis (RDA) that searches for dependencies and patterns in the data. It is planned to deliberately replace the raw data with the ordering relationships between the features. This step obviously causes some loss of potentially important information, however, it also opens up far-reaching new opportunities. First of all the RDA system will use features that are somehow detached from the raw values of the dataset which may increase robustness to methodological and technical factors, study-specific biases as well as normalization and standardization procedures. Secondly using only ranks within each sample will allow merging the data from different omics, platforms, and experiments into a single set, as well as using obtained data and results in subsequent secondary analyses and meta-analyses.

The pioneering nature of this project concerns two areas. The first one deals with creating a new comprehensive tool called Relative Dependencies Analysis (RDA) for omics data classification and knowledge discovery. It is a novel hybrid solution which combines and extends several data mining algorithms and concepts, including evolutionary algorithms, decision trees, multi-tests splits, and Relative Expression Analysis (RXA) methods. The GPU parallelization will be used to improve the local search of the top dependencies within each splitting node of the tree and will be embedded in specialized variants of mutation operators. The second novelty is adapting and using RDA solution to new omics as well as to multi-omics data. This will require additional research that relates to the data integration, inter- and intra-platforms variabilities as well as understanding the specifics of each omics separately. The RDA system will allow performing full-scale experiments that include even low-ranked features and searching for complex but still comprehensible predictions constituted with hierarchical and horizontal dependencies in multi-class data.

The following tasks are planned:

1. A detailed analysis of the relative dependencies between the features and their application in data classification will be performed. The task involves analyzing the (i) impact of different representation of the relations; (ii) RXA-based hierarchical and horizontal scoring; (iii) complexity, resemblance and discriminative power of the dependencies; (iv) development, implementation and validation of a new ML-empowered system for Relative Dependencies Analysis (RDA).
2. Adaptation of the RDA system to the analysis of various single-omic data.
3. Data mining and knowledge discovery based on integrated multi-omics data with the RDA system. Experimental evaluation of the predictions generated by the RDA system in the context of potential application in molecular biology primary research including biomarker discovery and classification.