

Mechanism based biomarker discovery by integrating multi-omics data into biological models

Asier Antoranz^{1,2}, Leonidas G. Alexopoulos^{1,2}

¹National Technical University of Athens, Athens, Greece

²ProtATonce Ltd, Athens, Greece

Mechanism-based biomarker discovery aims to identify molecular signatures that not only capture the underlying biology of the disease, but also allow to identify its diversity, distinguishing the subtype of origin and permitting more effective therapies. The present project, exhibits a proteomic biomarker discovery scheme that feeds on transcriptomic data using CKD as case study. The pipeline combines biological models extracted from publicly available databases with state-of-the-art algorithms that aim to capture meaningful biology for the disease. The goal is to suggest a number of proteins as candidate biomarkers to be measured in blood samples that are able to asymptotically diagnose a myriad of subcategories and use machine learning methods to integrate them into a panel of increased performance and stability. The cornerstone is to molecularly determine the underlying pathophysiological mechanism causing CKD and detect it before the damage is done. Briefly, our methodology includes 5 steps:

- 1) Integrate protein-protein interaction networks with annotated gene-sets to construct our knowledge-base (KB) which comprises contextual information about the biological entities participating in our model.
- 2) Feed this KB with CKD data obtained from the Gene Expression Omnibus (GEO) to create disease-specific models.
- 3) Apply different state-of-the-art omic-analysis methods (network analysis, pathway analysis, etc.) to describe a molecular profile for each biological entity included in the model.
- 4) Select and optimize the features included in the above-described molecular profile by applying an iterative active-learning approach in order to identify the best quality candidate biomarkers. The evaluation of performance as biomarkers is assessed by using plasma samples.
- 5) Integrate the individual candidates into a biomarker panel with increased performance and stability.

The motivation behind the use of blood samples is that, for monitoring human health, measuring protein biomarkers in blood is considered a very attractive solution because the pathology of almost every body tissue can affect the blood proteome on top of the simplicity of obtaining the sample.