Modelling Causal Mechanisms in Metabolic Disorders Integrating Omics Data

Gaia Lazzari [a], Ramin Amiri [a], Michele Costanzo [b,c], Margherita Ruoppolo [b,c], Camilla Luni [a]

[a] Department of Civil, Chemical, Environmental, and Materials Engineering (DICAM), Alma Mater Studiorum - University of Bologna, 40131 Bologna, Italy

[b] Department of Molecular Medicine and Medical Biotechnology (DMMBM), University of Naples Federico II, 80131 Napoli, Italy

[c] CEINGE, 80131 Napoli, Italy

E-mail: camilla.luni@unibo.it

Cellular metabolism is a complex network of biochemical reactions that sustains life, converting nutrients into energy and building blocks. It can be viewed as a large-scale reactive system governed by constraints such as mass conservation, stoichiometry and thermodynamic feasibility. To understand how metabolism behaves and is altered in disease states, it is fundamental to analyse biological data by creating computational frameworks.

High-throughput technologies, such as sequencing and chromatography coupled with mass spectrometry, enable collecting multi-omics data capturing different layers of cellular information, including gene expression, protein abundance and metabolite concentrations. While such data provide quantitative representations of cellular states, interpreting them remains a challenge due to their numerosity and interconnectivity. To understand what a specific experimental perturbation implies, it is crucial to develop computational pipelines that integrate omics data with prior biological knowledge. A goal of these pipelines is to reconstruct causal networks, models that represent the relationships between biological components, where each connection identifies a previously experimentally validated causal interaction. These networks help trace how molecular changes propagate through the system, linking upstream regulatory signals to downstream metabolic responses.

While existing methods often analyse omics data separately and do not explain how molecular changes are connected, this study integrates multi-omics data with known interactions to infer context-specific causal mechanisms. This methodology was applied to methylmalonic acidemia (MMA), a rare error of metabolism primarily caused by the deficiency of a specific enzyme. First, a customised prior-knowledge network was constructed by merging information from multiple databases, such as metabolic reactions or protein-metabolite associations. Next, differentially expressed genes and metabolites between healthy and pathological conditions were mapped onto the network. A typical cellular pathway involves genes encoding proteins such as transcription factors, which regulate the expression of metabolic enzymes, and the resulting enzymatic activity influences metabolite levels. Integrating these relationships enabled a precise analysis of the position and role of differentially expressed elements within the network. The final network was reduced by integer linear programming to highlight disease-relevant alterations consistent with known molecular features of MMA, also experimentally observed in other biological contexts. The reduced network can be used to extract information about cascades of perturbed events associated with the disease.

This approach contributes to the broader field of chemical engineering by providing a reproducible and generalisable framework applicable to system-level understanding of metabolism, supporting relevant applications to identify therapeutic points of intervention.

Keywords: *Metabolic engineering, Multiomics data, Therapeutic target identification*