

## Optimization of multi-omic genome-scale models

*Metabolism* is the set of biochemical reactions in a cell that maintain its living state. As these reactions are essential, it is vital that metabolic networks in all living organisms are well-characterized. This can be achieved by constructing a *genome-scale metabolic model*, a computational representation of every metabolic reaction taking place within a living system and the chemical compounds which are involved in these reactions (metabolites). Being downstream of gene expression, metabolism is being increasingly used as an indicator of the outcome for drugs and therapies, as well as for cancer studies.

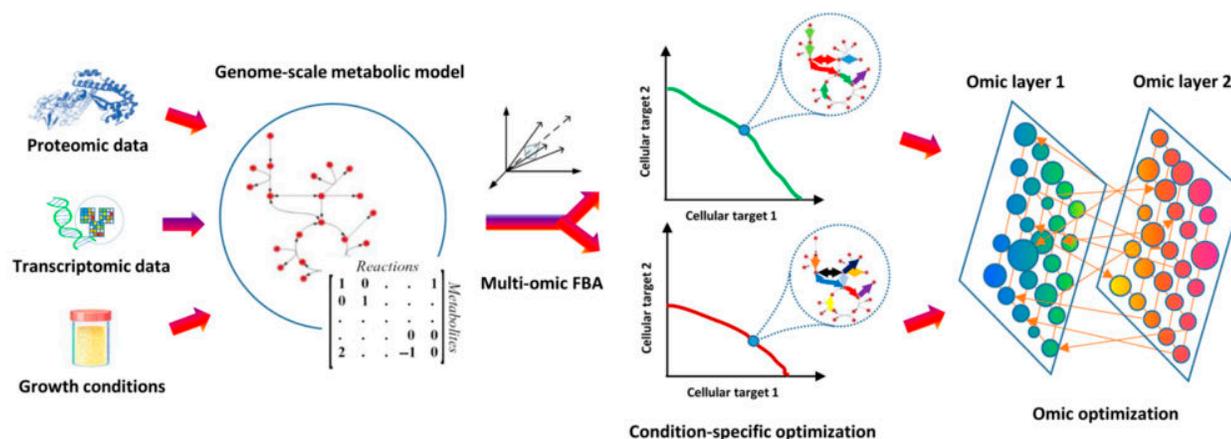


Fig. 1. Through the collection of transcriptomic, proteomic and other omic data across various growth conditions from experiments and existing literature, a multi-omic genome-scale metabolic model can be constructed. The simulation of growth under different conditions allows for condition-specific optimization of each of the omic layers, which can then be combined to form a multi-omic network.

*Flux balance analysis* (FBA) is commonly used to mathematically express the flow of metabolites (flux) through a network of biochemical pathways. FBA predicts rates to represent amounts of metabolites that are consumed or produced in every reaction in the metabolic network. Constraints can be imposed on the system to identify a range of values representing the entire distribution of fluxes under a specific set of conditions. A *multi-omic* model introduces additional constraints in the form of data spanning entire biological fields: for example, genes (genome), RNA (transcriptome), proteins (proteome), metabolites (metabolome), etc.

The integration of multi-omic data from existing literature and *in-vivo* experiments generates *condition-specific* models that represent the metabolic state more precisely. To extract the most meaning from multi-omic models, fusion of multiple data types into a single, cohesive network is essential for measuring responses at multiple omic levels.

Simulation of growth under different conditions allows for condition-specific optimization of each of the omic layers, which can be combined to form a multi-omic network (Fig. 1). This enables detection of coordinated responses shared between different data types as well as variation in responses across different growth conditions.

In many cases, an optimal flux value is sought that best satisfies a single metabolic objective (e.g. maximization of cellular growth rate) by using a process known as linear optimization. Owing to the fact that organisms often have multiple objectives to satisfy, *multi-objective optimization* presents a valuable alternative. Trade-offs between conflicting metabolic objectives can be resolved through simulating a series of non-dominated vectors, which are lists of values where there is no better solution that exists for a given objective without sacrificing the performance of another. A series of non-dominated vectors is called a *Pareto front* and enables consideration of multiple conditions and constraints affecting each objective in a multi-objective optimization problem (Fig. 2).

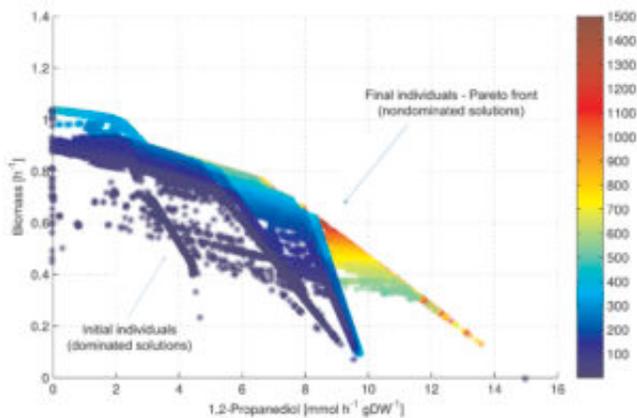


Fig. 2. Pareto front produced by METRADE when maximizing for 1,2-propanediol and biomass in *Escherichia coli*. The trade-off sheds light on the regions where the bacterium operates. Solutions are asterisks denoting potential growth conditions.

METRADE is a method for mapping transcriptomic or proteomic data onto a multi-omic metabolic model, and allows performing multi-objective optimization to identify optimal states through the comparison of predicted flux rates for multiple objectives. This is achieved by constructing a Pareto front displaying these reaction flux rates in a space where each profile is associated with a growth condition. The comparison of objectives is used to identify the best trade-off, where the maximal number of cellular objectives are simultaneously optimized.

The fusion of multiple data types into a single, cohesive network is a challenge, for which many factors must be considered. Generation of intermediate data structures or correlation of multi-omic data with model outputs can identify the circumstances under which certain variables are important

and what their effects are. Regression and machine learning techniques will be increasingly used to predict the effect sizes of the multiple omic variables involved (genes, transcripts, proteins and reaction fluxes), towards optimization of the cellular goal.

***Supreeta Vijayakumar***<sup>1</sup>, ***Max Conway***<sup>2</sup>, ***Pietro Lio***<sup>2</sup>, ***Claudio Angione***<sup>1</sup>

<sup>1</sup>*Department of Computer Science and Information Systems, Teesside University, United Kingdom*

<sup>2</sup>*Department of Computer Science and Technology, University of Cambridge, United Kingdom*

## **Publication**

[Optimization of Multi-Omic Genome-Scale Models: Methodologies, Hands-on Tutorial, and Perspectives.](#)

Vijayakumar S, Conway M, Lió P, Angione C

*Methods Mol Biol.* 2018