Modeling and Identifying Regulatory Patterns within Chaotic Metabolic Networks

Jordan A. Berg[1], Youjia Zhou[2,3], Bei Wang[2,3], Jared Rutter[1,4]
1. Department of Biochemistry, University of Utah; 2. School of Computing, University of Utah; 3. Scientific Computing and Imaging Institute, University of Utah; 4. Howard Hughes Medical Institute, University of Utah

Metabolism is a complex system of reactions, each with inputs, outputs, and modifiers, that together determine the health and stability of a cell. Perturbations in these reactions can lead to rippling downstream effects that determine an organism's physiological well-being. Perturbations to these systems are often the stimulus of disease. For example, in colon adenomas, perturbations to the transport of the metabolite pyruvate across the inner mitochondrial membrane lead to a remodeling of glucose metabolism and an increase in metabolites associated with the Warburg effect. This perturbation alone is sufficient to increase tumorigenesis significantly[1].

Analysis of metabolism under different biological conditions generally involves limited, reductionist approaches where a specific metabolite or protein abundance is measured, or systems analysis via metabolomics. However, there are many challenges in metabolomics data analysis. The first challenge is related to sparsity of metabolite measurements. While metabolomics can identify several metabolites in a biological system, often a good metabolomics dataset will at best measure 100–200 metabolites out of the 1,500+ metabolites currently known to be involved in human metabolism[2,3]. This leads to gaps in metabolism as one may only measure those involved in a percentage of the total steps in a given metabolic pathway. The second challenge hinges on the analysis of metabolic datasets. Traditional approaches for metabolomics analysis center on identifying high-magnitude changes in the dataset, often through methods such as generating a volcano plot of the data. Other methods, such as set enrichment, can tell the user if there is an enrichment of a given list of metabolites in a certain metabolic pathway. Others allow for overlaying user-measured data on the metabolic network. However, none of these approaches allows for the automated identification of metabolic regulatory patterns within their data. This generally results in the user focusing on a few high-magnitude changes in the dataset, or the user manually picking out a pattern in a metabolic pathway they are already interested or familiar with.

In order to address these limitations in metabolic network analysis, we created *Metaboverse*, a computational platform for the integrated analysis and dynamic visualization of user data on the metabolic network. *Metaboverse* is written using a combination of Python and Javascript and is packaged using Electron to create a portable, cross-platform application. Users specify their organism of interest and experimental parameters. They also provide relevant transcriptomic, proteomic, and/or metabolomic datasets for their biological model. Once these inputs have been provided using the user-friendly graphical user interface, *Metaboverse* builds the relevant metabolic network using Reactome database. Reactome curates the metabolic network using well-vetted sources; such as HMDB (the Human Metabolism Database), KEGG (the Kyoto Encyclopedia of Genes and Genomes), and Recon (Global Human Metabolic Reconstruction of Metabolism). With the integration of these resources, a rich database of catalysts and inhibitors of metabolic reactions is also available.

After the appropriate metabolic network is built, user provided data is layered onto the network as node weights for the corresponding input and/or output nodes. For nodes that represent a protein complex, available measurements for the complex components are used to infer the abundance of the given complex. Gene expression values can also be used to infer values for proteins if measured values for that protein are unavailable. While concerns might arise for the correlation between genes and protein levels, metabolic enzyme levels have been shown to better correspond with gene measurements[4]. At this point, *Metaboverse* utilizes a novel approach for handling sparsity or measurements in the metabolic network. In order to collapse reactions with missing input or output values into pseudo-reactions, we use a collapse paradigm as follows (summarized in Fig. 1). If a reaction has at least one known input and output measurement, the reaction is maintained as is in the network (Fig. 1a). If a reaction has at least one known input and no known output, or vice versa, *Metaboverse* searches the global reaction network for all reactions with matching inputs or outputs to the reaction in question. If a matched reaction has at least one

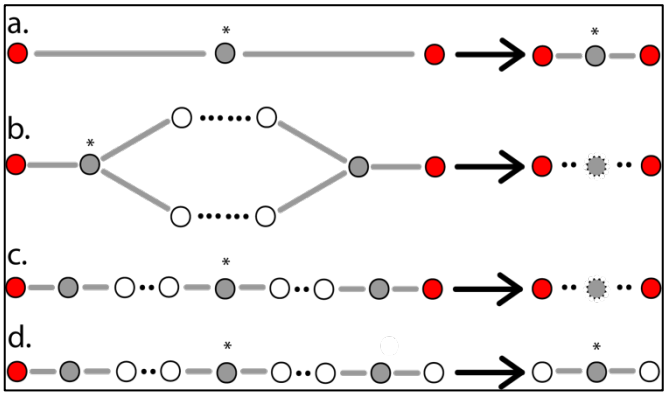

Figure 1: Reaction collapse schematic. An asterisk indicates a reaction being considered for collapse. A grey node indicates a reaction. A red node indicates a measured component node. A white node indicates an unmeasured component node. A solid edge indicates a curated relationship. A dashed edge indicates a bridged, pseudo-relationship.

measured input or output, the two reactions are collapsed into a pseudo-reaction summarizing the two reactions (Fig. 1b). If a reaction has no measured inputs or outputs, the global network is searched for reactions that match both the inputs and outputs of the given reaction for other reactions that each have at least one measured input or output and a pseudo-reaction is created to summarize these reactions (Fig. 1c). All other reactions are maintained as is in the network (Fig. 1d). This process creates a collapsed metabolic network where up to three reactions can be condensed to a single summary, or pseudo- reaction that contains at least one known input and one known output. We limit the reaction collapse to three reactions to reduce loss of information across a metabolic pathway as condensing too many reactions can increase the probability of oversimplifying the characteristics of a given set of reactions.

After these curation steps are complete, a constructed metabolic network with the provided user data is archived for future reference in JSON format, and the user can dynamically explore their data. Metabolic pathways are visualized using the D3.js (Data Driven Documents) framework, which allows for the creation of interactive networks. At this point the user can explore patterns and features of different networks and how their data behaves across these pathways. If a user notices an interesting value or pattern, they can expand that particular point into a graph of the point of interest (a metabolite, protein, etc.) and each reaction the selected component is involved in across the global network, or its nearest neighborhood. This view can be expanded to *n* nearest neighborhoods. As hub metabolites which are involved in many reactions (e.g. water, protons), we limit the visualization of these high-degree nodes to prevent excessive crowding of components that are only connected via these high-degree entities and that may not be relevant to model.

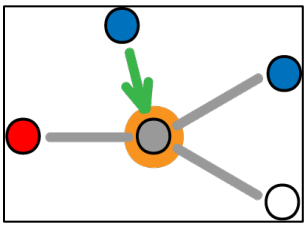

Figure 2: Example motif. Grey node indicates a reaction. White node indicates no measured value, blue node indicates abundance decrease, red node indicates abundance increase. Green arrow edge indicates a catalyst. An orange outline indicates presence of a motif.

In order to rapidly identify all interesting regulatory patterns in the global network based on the user-provided data, we introduce an application of activity motif searches[5]. Classically, motif searches search for repeated structural patterns in a network. We adapted the activity motif search methodology to focus on node weights as determined by the user-provided -omics values for a given input or output node in a reaction. This adapted motif search method searches the global reaction network for a variety of possible patterns based on the provided data. For example, a motif might be a reaction where an input shows a 2-fold decrease in abundance compared to the control, and an output shows a 2-fold increase in abundance compared to the control (Fig. 2). This pattern might indicate perturbations at this reaction that are relevant to phenotypes in the biological system and may be of interest for the user to perform further follow-up studies. *Metaboverse* requires the user to provide appropriate statistical measurements associated with each measurement. These are used to weight each motif's priority when returned to the user. Regulatory information is also used to inform the importance of a motif. For example, in the example explained above, there may also be a measured catalyst or inhibitor of the reaction that also undergoes a significant shift in abundance or expression and may explain the mechanism behind the observed change at the reaction (see green edge in Fig. 2). Results for the various motifs are returned to the user in an interactive stamp view. The user can then select one of these reaction motifs, select from any of the pathways the reaction is a part of, and dynamically explore the pathway or nearest neighborhood of the reaction's inputs or outputs. For other motifs outside of the generic motifs provided and searched by default in *Metaboverse*, a user can interactively generate a new motif to be searched in the network. For example, the user may be interested in a reaction where an enzyme input to a reaction is up-regulated, and an output metabolite is down-regulated. By drawing the desired pattern in the interactive motif builder, *Metaboverse* can then additionally search the global network for matching patterns.

The new functionalities and interactive capabilities provided in *Metaboverse* will revolutionize our ability to understand biological systems. These features will allow users to quickly formulate new hypotheses from their data related to the regulation of metabolism and the consequences of these changes that may be missed without appropriate systems-wide approaches. With these added capabilities, scientists will begin to extract more information from their data and will be able to better contextualize their data in the framework of metabolism.

1. Bensard, C. L. *et al.* Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier. *Cell Metab.* **0**, (2019).
2. Pinu, F. R., Goldansaz, S. A. & Jaine, J. Translational Metabolomics: Current Challenges and Future Opportunities. *Metabolites* **9**, (2019).
3. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
4. Zelezniak, A., Sheridan, S. & Patil, K. R. Contribution of Network Connectivity in Determining the Relationship between Gene Expression and Metabolite Concentration Changes. *PLOS Comput. Biol.* **10**, e1003572 (2014).
5. Chechik, G. *et al.* Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat. Biotechnol.* **26**, 1251–1259 (2008).