

# Gazing into the Metaverse: Automated exploration and contextualization of metabolic data

*This manuscript was automatically generated on May 21, 2020.*

## Authors

---

✉ Jordan A. Berg<sup>1</sup>, ✉ Youjia Zhou<sup>2,3</sup>, ✉ T. Cameron Waller<sup>4</sup>, ✉ Yeyun Ouyang<sup>1</sup>, ✉ Sara M. Nowinski<sup>1</sup>,  
✉ Tyler Van Ry<sup>1,5</sup>, ✉ Ian George<sup>1</sup>, ✉ James Cox<sup>1,5,6</sup>, ✉ Bei Wang<sup>2,3</sup>, ✉ Jared Rutter<sup>1,6,7</sup>

1. Department of Biochemistry, University of Utah
2. School of Computing, University of Utah
3. Scientific Computing & Imaging Institute, University of Utah
4. Division of Medical Genetics, Department of Medicine, School of Medicine, University of California San Diego
5. Metabolomics Core Facility, University of Utah
6. Diabetes & Metabolism Research Center, University of Utah
7. Howard Hughes Medical Institute, University of Utah

† To whom correspondence should be addressed:

\* This is a draft of the eventual final manuscript and should therefore be treated as such. Conclusions, along with author order and contributions, may change as the manuscript is finalized.

## Abstract

---

Metabolism and other biological interactions and reactions are complex, each with variable inputs, outputs, and modifiers. The harmony between these factors consequently determines the health and stability of a cell or organism. Perturbations to these components often have rippling downstream effects. These effects can be difficult to trace across the global reaction network, particularly when the effects occur between canonical representations of pathways. Researchers have largely utilized reductionist approaches to understanding these systems; however, these methods are limited in their scope. Even the power of systems-centric -omics approaches can be limited when only a handful or “loud” signals in the data are prioritized. To address these challenges, we developed Metaboverse, an interactive desktop app for the interactive exploration and automated extraction of potential regulatory events, patterns, and trends from multi-omic data within the context of the metabolic network and other global reaction networks. This framework will be foundational in increasing our ability to holistically understand static and temporal metabolic events and shifts and gene-metabolite intra-cooperativity. Metaboverse is freely available under a GPL-3.0 license at <https://github.com/Metaboverse/>.

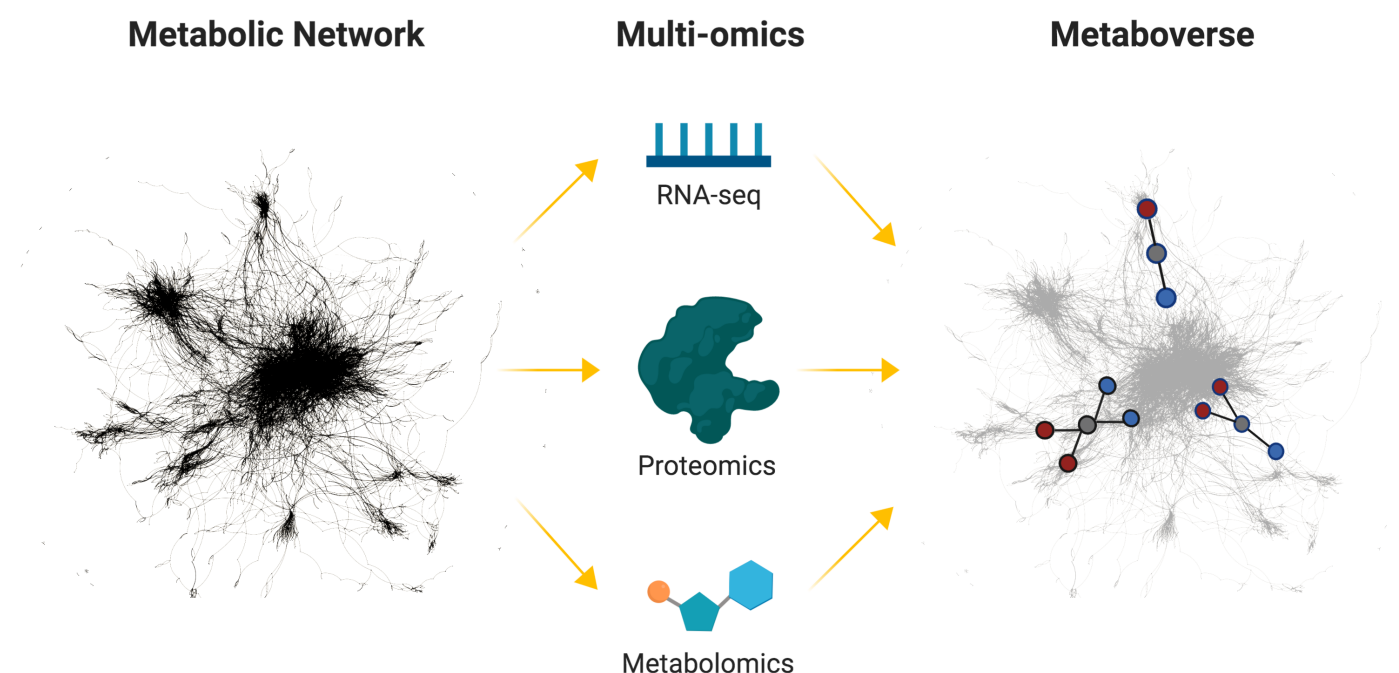
## Introduction

---

Metabolism is a complex network of reactions and interactions between genes, enzymes, complexes, and metabolites [1]. To understand these complex components, researchers normally adopt a reductionist approach to teasing apart the characteristics and mechanics of these processes and how they fit into the larger picture of biology and disease. While a vital component in the scientific process, by doing so, many interesting properties of metabolism can be missed. For example, in differential gene expression analysis, researchers rely on thresholds of magnitude and statistical significance to prioritize genes for follow-up study. However, this can inadvertently limit the scope of study of metabolism when in fact metabolism is a highly interconnected system where distal components and their modulation can have rippling effects across the network. The current approach is analogous to telling the story of Little Red Riding Hood, but only by reading the 20 most frequent words used in the study. Certainly doing so efficiently highlights key words like “wolf” and “little red riding hood,” but also prevents a coherent story from being told and would make it difficult for someone who had never heard the story of Little Red Riding Hood from comprehending the actual plot.

Over the past decade, several computational tools have emerged and become popular for their focus on trying to solve these issues in data contextualization. We will highlight four, and while others exist, we focus on tools representative and most popular for their respective properties. First is MetaboAnalyst, which relies largely on set enrichment methods for analysis of data, or looking at the belongingness of sets of significantly changed analytes (i.e. metabolite, protein, or gene measurements), for extracting interesting information. While network visualization is available, it focuses primarily on interaction networks and its ability to extract regulatory information is limited, particularly in an automated fashion [2,3]. Second is Cytoscape, which serves as a general go-to platform for representing biological or other networks. One strength of Cytoscape is its ability for users to design their own apps, or plug-ins to develop customized analyses. However, comprehensive regulatory identification methods are unavailable [4]. MetExplore focuses on the curation of networks, and is particularly useful for collaborative annotation of emerging organisms. It additionally can layer experimental data on the network for visualization [5,6]. Reactome, which Metaboverse uses for the curation of biological networks, also offers analytical tools for user data, but again relies on set enrichment or manual methods for identifying patterns. While all have their respective utility, there is still a need for tools that automate pattern and trend detection, especially when data is sparse, across metabolic networks in order to extract regulatory and other features from data [7,8,9].

In order to address these limitations in current conventions of metabolic data analysis, contextualization, and interpretation, we created the operating-system-independent desktop application, Metaboverse, to aid users in filling in and expanding upon the details of their model's metabolic story. Metaboverse is an interactive tool for exploratory data analysis that searches user data in the context of the metabolic network to identify interesting patterns and trends within the data. Metaboverse will aid scientists in formulating new hypotheses from their data and aid them in designing follow-up experiments for a deeper understanding of their model. Metaboverse operates across the entire metabolic network to quickly and automatically detect patterns and trends from a pre-designed pattern library, or can accept interactive input from the user where they can define certain patterns or trends they would like to identify across the global metabolic network. Figure 1 provides a graphical abstract to illustrate *Metaboverse's* role in the exploratory data analysis of biological data in the context of metabolism.



**Figure 1: Metaboverse conceptual overview.** Illustration summarizing the usage of Metaboverse to use biological data layered on the global reaction network to rapidly identify regulatory hotspots. Traditionally, when a scientist performs an -omics experiment, they tend to focus on a couple of features that are differentially regulated. Metaboverse contextualizes the data across the metabolic network and identifies interesting regulatory patterns in the data.

In order to provide a platform for the exploration of single or multi-omic metabolic data, we developed several computational features to aid in the aims discussed above. First, we developed a pattern search engine for the rapid and automated identification of patterns and trends in -omic data on the metabolic network. Conceptually, this search engine borrows principles of topological motif searching from graph theory [10,11] and builds upon the concept of “activity motifs” [12]. Another feature introduced in Metaboverse allows for the interactive exploration of specific reactions or reaction entities with on-the-fly pattern search analysis. Users can explore specific pathways of

interest and look for other interesting patterns and trends in pathways of interest. Perturbations in the abundance or behavior of a particular metabolic network component can lead to downstream genotypic and phenotypic modulations in a biological system. We include an interactive perturbation connectivity module to allow users to explore the effects of perturbations on their system. Finally, we tackle the challenge of sparsity, particularly in metabolomics datasets, by introducing a reaction collapsing or bridging feature that allows for searching for motifs across multiple reactions, where intermediate steps of the reaction trajectory may not have measured values for their respective input or output components.

Metaboverse is designed to handle standard two-condition experiments, flux metabolomics (in progress), and time-course or multi-condition experiments. Users can provide any combination of datasets from each -omics categories (values mapping to gene, protein, and metabolite identifiers). Time-course inputs can consist of a matched samples for every time point, or can provide a combination of steady-state levels and time points. Users input fold change and appropriate statistical measures with proper multiple-testing adjustments from their respective -omics datasets, which Metaboverse then layers onto the metabolic network. Metaboverse can handle data from a variety of model organisms, including humans, mouse, yeast, zebrafish, and more. The foundational curation of Metaboverse is built on Reactome reaction curations, thus providing 90+ species which the Metaboverse environment can process. In order to validate these methodologies available in Metaboverse we analyzed a variety of datasets to provide representative vignettes that highlight *Metaboverse's* reliability in extracting canonical features, as well as utility in identifying novel features and patterns. We expect that Metaboverse will become a foundational piece of the analytical toolkit and augment one's ability to more deeply and holistically explore metabolism and other reaction networks and aid in our ability to provide clearer context within these models.

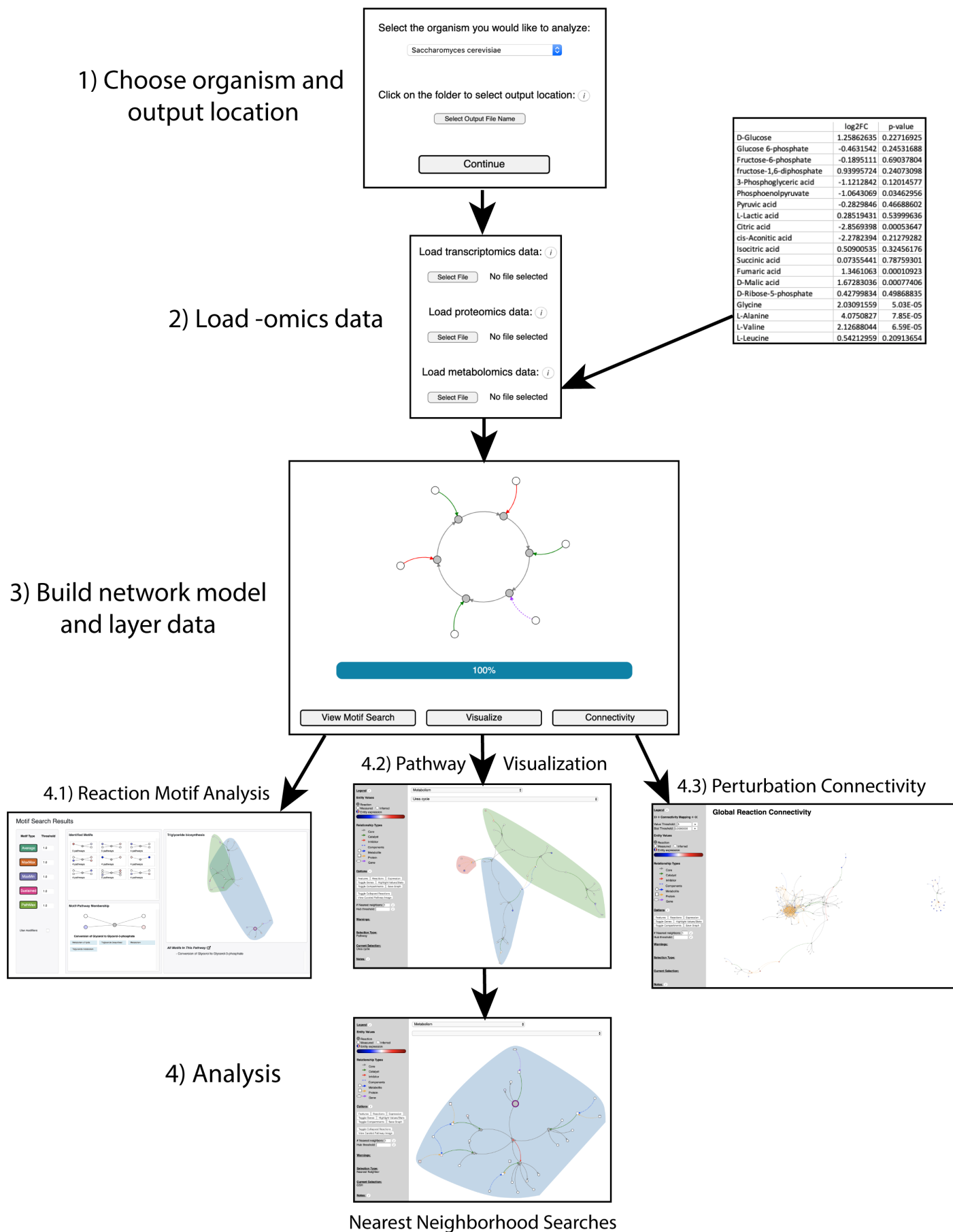
## Results

---

### **Metaboverse is a dynamic, user-friendly tool for the exploration of high-throughput biological data in organism-specific pathways.**

#### **Overview.**

We designed Metaboverse as a light-weight, self-contained app for the dynamic exploration of high-throughput biological data. The pathway curations are derived from Reactome, and as of writing, is capable of analyzing data for over ninety species. A user begins by providing a previously curated Metaboverse file, or the desired output location for a new curation and organism of interest. Next, the user provides the relevant transcriptomic, proteomic, and/or metabolomic datasets to layer onto the global reaction network of their organism of interest. These data categories can be extended to any dataset that uses the relevant mapping IDs; for example, one could provide ribosome profiling translation efficiency values mapped to the appropriate gene IDs for layering onto the network and downstream analysis. During this step, the user will also specify a few experimental parameters for consideration during downstream analysis and visualization. Following these user inputs, the organism's network, data processing, and motif analysis (discussed further below) is curated and a curation file is output for future analysis (see [Figure 2](#)).



**Figure 2: Metaboverse package overview.** Illustration comparing the data input, processing, and exploration steps of the Metaboverse package.

## **Handling data sparsity within the biological global network.**

Missing data points, particularly in metabolomics experiments, are frequent and can make analysis of pathways and identification of regulatory patterns in the network challenging. While thousands of metabolites are known to participate in human metabolism, the current state of the technology for determining which mass spectra belong to which metabolite can be challenging and often results in a limited number of data points being quantified. These can lead to gaps in the metabolic network which can be challenging to explore and analyze. We therefore developed a reaction compression algorithm (detailed more in the Methods section) that collapses up to three reactions with missing data points if they can be bridged with known data on the distal ends of the reaction path. These reactions, or pseudo-reactions are visually distinct to show users where reactions were collapsed and understand what that pseudo-reaction summarizes.

## **Rapid identification of interesting regulatory patterns in the reaction network.**

Following network curation, the user can visualize available reaction motifs identified across the global reaction network. In a computational science context, a motif is a re-occurring pattern in network structure, or the organization of network entities and their relationships to one another. However, with -omic data, we are more interested in identifying patterns in expression or abundance of genes, proteins, and metabolites. Previous work by Checkik, et. al. introduced the concept of “activity motifs”, where instead of identifying motifs based on network structure, they were identified using the expression characteristics of nodes in transcription factor binding signaling networks [12]. We adapted this methodology to identify and interactively display interesting regulatory hotspots within the global reaction network. For example, a reaction’s inputs may exhibit increased levels and its outputs low levels, indicating a regulatory event occurring at the reaction.

In Metaboverse, we define a motif as a regulatory pattern identified across a reaction or pseudo-reaction. Metaboverse contains a library of default motifs to search the network for, and users can define custom motifs they would like to identify across the global network (please refer to the documentation for more details on available motifs [13]). Metaboverse will search the global network from a pre-defined library of regulatory patterns (regulatory hotspots, or motifs) and return a graphical stamp view of conserved patterns. These motifs by default are ordered by the statistical values associated with the entity components of the motif, but can be ordered by the strength of the motif based on the magnitude of change. Statistical values associated with each measurement for a measured entity as provided by the user are used to weight these motifs and prioritize returned results. We use a three-tiered sorting strategy when sorting by the associated statistical values. The highest prioritized motifs are those where the relevant components on either side of the reaction (inputs vs. outputs) that determined the reaction motif are statistically significant. Of these motifs from the first tier, motifs are sorted by lowest to highest cumulative p-value or other relevant statistical value. In the second sorting tier, reactions where at least one side of the reaction had a statistically significant motif component are sorted by statistical strength. Finally, all other motifs are sorted by the cumulative statistical value for the motif components.

Reaction motifs are visually distinct during network visualization to quickly draw the user to these interesting patterns. For a given pattern, the user can then explore each pathway this particular motif is found in. Motif analysis of the global regulatory network will allow users to rapidly identify interesting features within the data, particularly patterns between canonical pathways or in other pathways that may seem tangential in their research, but are instead rather striking based on these contextualized analytical method. In the data vignettes below, we demonstrate this utility further. Users will also be able to design their own patterns through an interactive pattern drawing tool, and even design specific scenarios that are cognizant of feature type (in progress). For example, one might



be interested in a pattern where a protein displays higher expression, but the resulting metabolite is decreased. This could then be coded interactively into the Metaboverse framework by the user.

### **Dynamic visualization of organism-specific reaction pathways.**

Following curation of the global network as described above, the user can manually search individual canonical pathways or individual entities and their reaction neighborhoods. For a given selection, all relevant reactions that are annotated as a part of that pathway will be graphed, along with their core input (reactant) and output (product) components. In addition to these core elements, known catalysts and inhibitors are included, as well as the component proteins, genes, and metabolites that form functional complexes involved in a particular reaction. Labels can be toggled on or off in the display, and the user can switch between viewing the values or statistics associated with each data point with their relevant color mapping. In cases where a gene value is known, but its protein value is unmeasured, the protein value will be inferred using aggregated gene component values. The same is then done for functional complexes using their inferred or known component values. Relevant pathway and analytical metadata is also displayed. Other information, such as identified motifs found in the pathway can also be found and expanded for further exploration in a new window. Aids for visualization are also available, such as the ability to remove nodes from visualization that contain a high number of relationships to other network features such as that these nodes, which act as hubs in the network, do not lead to cluttered representations of the network. Often, these hub nodes are ubiquitous features such as water and proton which may be of limited interest to the user during data visualization. Compartment domains are also graphed to include a relevant reactions and their components that occur in a given cellular compartment.

### **Visualization of downstream effects of network perturbations.**

Users may be interested a particular metabolite or protein and the downstream effects its perturbation has on related pathways. By double-clicking a node of interest, or by selecting the entity name from the drop-down menu, the user can explore all downstream effects across all pathways in the global network. The user can also define how many neighborhoods to display such that one can visualize two or more reaction steps downstream of the selected entity. This functionality moves the analysis away from our traditional, strictly defined pathway approach to analysis, and helps contextualize the far-reaching effects changes in metabolism or other biological systems can have across classical pathways.

### **Exploring perturbation connectivity within the global network**

Abundance or behavioral changes of a metabolic network component can lead to downstream genotypic and phenotypic modulations in a biological system. One important measure of robustness of a network is “connectivity”. In a biological context, an example of connectivity, or the loss of network connectivity, is easily grasped when considering a transport reaction from the cytosol into the mitochondria of a critical metabolite [[14](#),[15](#),[16](#),[17](#)]. When the transport hub is abolished and network connectivity is lost, the critical metabolite cannot participate in the required downstream reactions. For example, an example from our laboratory established the consequences of the ablation of the mitochondrial pyruvate carrier on downstream citric acid cycle processes and the corresponding increase in lactate production [[18](#)] and leads to serious cellular complications and dysfunction [[19](#),[20](#)].

The importance of connectivity could also be considered in a treatment context. For example, a druggable and critical metabolite may be perturbed in a particular disease context. However, if a metabolite participating in a neighboring, downstream reaction is also perturbed in a way not related

to the perturbation of the first metabolite, the efficacy of the drug treatment could be severely impaired [21,22,23]. To aid in exploration of the connectivity of the biological network, we also developed a connectivity module where users can display all reactions that have at least one involved component perturbed based on either an abundance or statistical level. By doing so, when the graph is constructed, proximal reactions that were perturbed independently will be “sewn” together to reconstruct a perturbation connectivity map in the user’s model.

## Data vignettes

In order to demonstrate the added utility of Metaboverse to the community that is not currently available in other tools, we used Metaboverse to analyze a series of public and new datasets. From the vignettes provided below, we show that Metaboverse not only is able to identify points of interest previously described or expected, but can rapidly identify for the user unexpected and systematic regulatory patterns in a reaction network context.

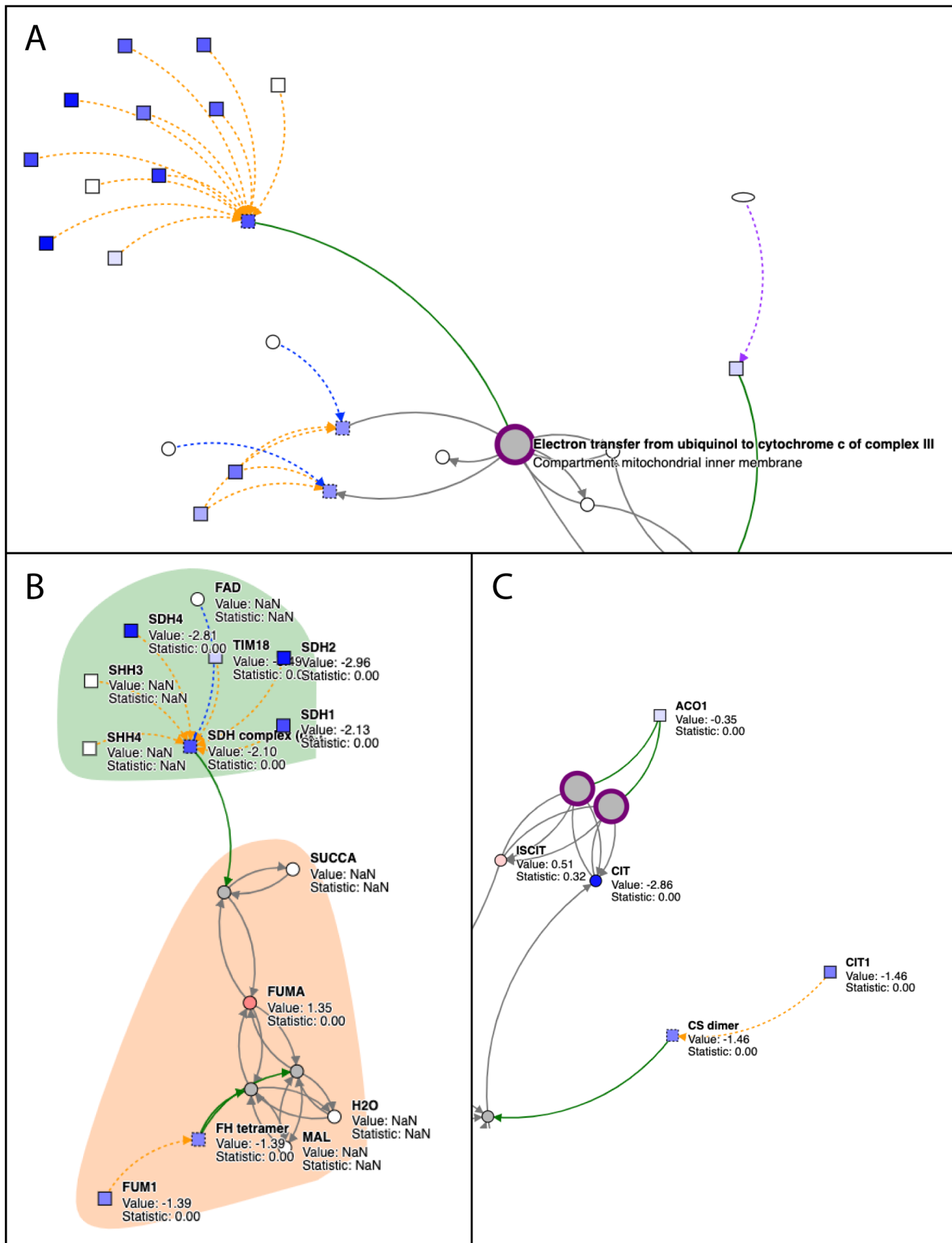
### 1. Perturbation of mitochondrial fatty acid synthesis across time reveals expected and unexpected reaction motifs.

Mitochondrial fatty acid synthesis (mtFAS) acts as an important metabolic pathway long known to function to produce lipoic acid, a critical cofactor for several metabolic enzymes. Recent work has begun to uncover additional, important roles for this pathway. For example, we now know that mtFAS can perform metabolic nutrient sensing roles to coordinate lipoic acid synthesis with iron sulfur (FeS) cluster biogenesis, assembly of oxidative phosphorylation complexes, and more [24,25]. Additionally, this pathway has received increased physiological focus with the discovery of patients with mutations in key mtFAS enzymes [26].

*MCAT*, or *MCT1* in *Saccharomyces cerevisiae*, is an acyltransferase responsible for the transfer of a malonyl group from malonyl-CoA to *NDUFAB1*, the mitochondrial acyl carrier protein. By knocking out this gene, one can simulate the effects of a disruption in mtFAS.

In order to probe the relationship between mtFAS-related protein concentration and the effects of its perturbation on downstream metabolic processes from a systematic perspective, we used a *mct1Δ* model in *Saccharomyces cerevisiae* to measure steady state proteomics and transcriptomics at 12 hours after a shift to a non-fermentable carbon source, as well as steady state metabolomics at 0, 15, 30, 60, and 180 minutes after this shift in carbon source. By layering these data onto the *Saccharomyces cerevisiae* metabolic network using Metaboverse, we observed interesting respiratory signatures as expected based on our previous work [24,25]. For example, using Metaboverse, we noticed a strong pattern in the electron transfer from ubiquinol to cytochrome C of complex III of the electron transport chain (ETC) (Figure 3 A). At the proteomics level, cytochrome C component proteins, CYC1 and CYC7 are both significantly reduced in concentration compared to wild-type cells ( $\log_2$ (fold change): -1.57 & -0.88; p-value: < 0.01 & < 0.01; respectively). This reduction in cytochrome C concentration is paired with a marked reduction in concentration of the protein components of cytochrome C reductase, which is a catalyst of this electron transfer reaction (average  $\log_2$ (fold change) of all measured protein components = -2.03, where 9/11 component proteins were measured. Range of  $\log_2$ (fold change) values: -0.34 to -3.21; all p-values below 0.01).





A second pattern of interest that we expect in this model is the conversion of Succinate to Fumarate by the Succinate dehydrogenase complex (SDH, or Complex II of the ETC) (Figure 3 B). We know from previous work that complex assembly and mtFAS are intricately connected [24,27]. However, this particular behavior when pairing the temporal metabolomics data with the steady-state proteomics data is interesting and acts as a potential jumping-off point for further follow up experimentation. Across each time point, we see a steady accumulation of fumarate ( $\log_2(\text{fold change}) \geq 1$ ). We pair this with the proteomics data and observe a strong reduction in both SDH and FUM1, which encodes the fumarase needed to convert fumarate to malate in the TCA cycle (-0.49 to -2.96). It is possible that the reduction of FUM1 and build-up of fumarate is acting as a bottleneck and signaling a reduction in Complex II component production and assembly. However, as measurements for succinate and malate are missing, it is difficult to clarify this regulation more without further follow-up experiments. An additional hypothesis may involve SFC1, the mitochondrial succinate-fumarate transporter, which we previously noticed from the analysis of the proteomics data was reduced in concentration 2.5-fold. Unfortunately, at this time this transport reaction is not annotated in the Reactome database for *Saccharomyces cerevisiae*, which raises an important goal for the Metaboverse project, which is to augment these pathway representations to fill in the gaps of knowledge using additional sources. However, with the prior knowledge, it is interesting to notice this pattern with fumarate as the build-up of fumarate may reflect the inability of the mitochondria to export fumarate.

A final observation of interest from analysis through Metaboverse is that of citrate and isocitrate. Normally, the TCA cycle generates citrate from malate to oxaloacetate via MDH1, then oxaloacetate to citrate via CIT1 and CIT3. Building off the observations discussed above, we expect lower levels of malate as due to decreased MDH1. We also observe lower levels of CIT1 (-1.46) and reciprocal decreases in citrate that worsens across the time-course. However, at 30 minutes post-carbon source shift, isocitrate levels are decreased by -1.38  $\log_2(\text{fold change})$  (p-value < 0.01). By 60 and 180 minutes, there is no significant change in isocitrate levels compared to wild-type cells, while citrate levels remain significantly reduced (Figure 3 C). Does this indicate some sort of adaptation to a respiratory-required environment with ETC-impaired cells?

By analyzing this multi-omics dataset using reaction motif analysis and other interactive visualization, interesting questions arise. We see several reaction motifs that are expected based on prior knowledge of this biological model, as well as other interesting behaviors worthy of further follow-up. This demonstrates the potential Metaboverse has to act as a valuable hypothesis-generation tool, as well as a convenient platform for visualization and analysis of a user's dataset in the context of the metabolic or other reaction networks.

## Discussion

---

In this manuscript, we introduce a new software tool for analysis and exploration of user data layered on the metabolic and global reaction networks. To improve on previous tools with similar capabilities, we introduced several new analytical tools and methods to aid users in the automated identification and discovery of regulatory patterns within their data in a reaction network context. These include the automated ability to identify reaction motifs across the global reaction network, such as a reaction where an input is low abundance and an output is high. Metaboverse also provides dynamic and interactive visualization capabilities to search for patterns and features within the user data on a reaction network within classical pathways. If a user is interested in how a reaction motif is propagating across the global reaction network and not just a single pathway, a user can explore a reaction component's nearest reaction neighborhood. The user can also explore the connectedness of perturbations across the global network and begin to explore hypotheses of the role of redundancy within a biological phenomenon.

In order to handle the challenge of sparsity, particularly in regard to metabolomics data and the metabolic reaction network, we introduce a reaction collapsing feature which summarizes a series of connected reaction where values may be missing for a number of reactions, but where the terminal ends of the reaction path have measured values. Importantly, this augments the capabilities available within Metaboverse, especially in the identification of additional reaction motifs that may be of interest to the user.

We demonstrated the utility of Metaboverse in exploring single- and multi-omic datasets. We analyzed previously published studies from a variety of organisms (in progress), as well as generated a novel dataset that highlights the time-course capabilities of this framework.

While Metaboverse aims to enhance the computational toolkit for data analysis and hypothesis generation in metabolic and other experiments, a number of challenges still remain, which we intend to address as we continue to maintain and build upon this software. For example, while the reaction collapsing features of Metaboverse aid in identifying patterns across several reactions where data may be missing, there are a variety of biological and technical edge cases that need to be considered in future implementations of this feature. This is particularly challenging in datasets where few metabolites were measured. Hopefully, as technical limitations in metabolomics are also overcome, more complete snapshots of metabolism will be visible within this framework. Additionally, while we take a more straightforward and somewhat rudimentary approach to statistical significance integration in the reaction motif searches, more holistic platforms for cross-omics integration are needed and remain a consequential challenge in multi-omics research.

In summary, we hope that Metaboverse will bring a new perspective to users' data. We envision Metaboverse to be a staple tool in the metabolic research toolkit that will help researchers critically and holistically consider their data in the context of biological network interactions and help draw the connections needed to aid them in extracting new and exciting hypotheses from their data that would be difficult to do without this tool.

## Methods

---

A tutorial for how to use Metaboverse can be found at [metaboverse.readthedocs.io/getting-started](https://metaboverse.readthedocs.io/getting-started).

### 1. Network Curation

Biological networks are curated using the current version of the Reactome database. In particular, the pathway records for each species, complex component and interaction data, Ensembl, and UniProt Reactome mapping tables are integrated into the network database for Metaboverse. Additionally, the ChEBI database names table ([ftp://ftp.ebi.ac.uk/pub/databases/chebi/Flat\\_file\\_tab\\_delimited/names.tsv.gz](ftp://ftp.ebi.ac.uk/pub/databases/chebi/Flat_file_tab_delimited/names.tsv.gz)) is integrated. These data are used to generate a series of mapping dictionaries for entities to reactions and reactions to pathways for curation of the global network.

After the relevant information is parsed from each table or record, the global network is propagated using the NetworkX networking framework [cite:networkx] to generate nodes for each reaction and reaction component, and edges connecting components to the appropriate reactions. In some cases, a separate ID is used to generate two nodes for the same metabolite within two separate compartments to aid in visualization downstream; however, user data for the given entity would be properly mapped to both nodes.

After the network is curated for the user-specified organism, each node's degree (or magnitude of edges or connections) is determined to aid in the user's downstream ability to avoid visualizing high-degree components, such as a proton or water, on the metabolic network, which can lead to graphical entanglement and cluttering and a decrease in computational performance [cite:Waller;GigaScience; 2020].

## 2. Data overlay and broadcasting for missing entities

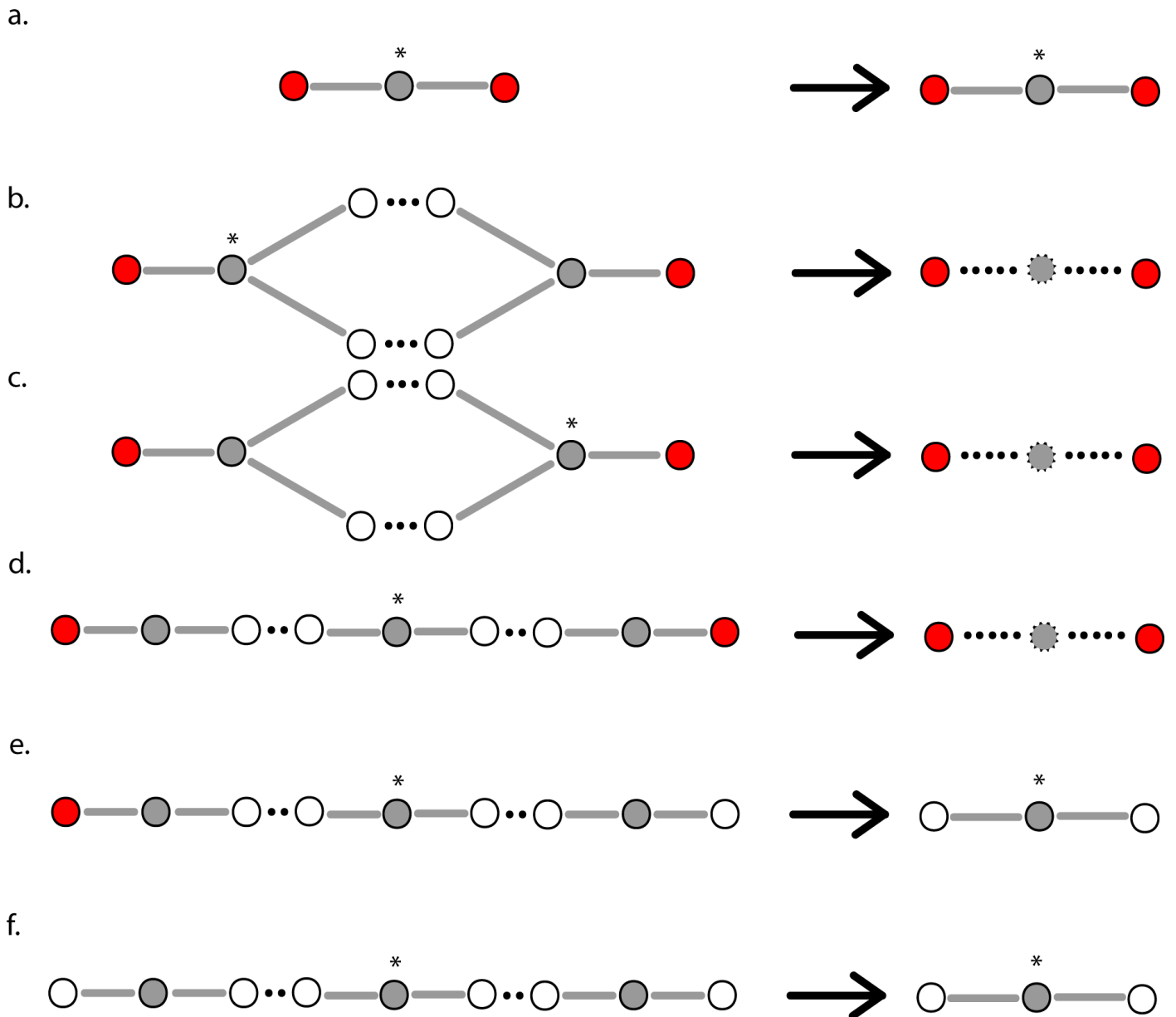
In order to overlay user data on the global network, first, user-provided gene expression, protein abundance, and/or metabolite abundances' names are mapped to Metaboverse compatible identifiers. For components that Metaboverse is unable to map, a list will be returned to the user so they can provide alternative names to aid in mapping. Second, provided data values are mapped to the appropriate nodes in the network. In cases where gene expression data is available, but protein abundance data is missing, Metaboverse will take the average (or user-defined??) of the available gene expression values to broadcast to the protein node. For complexes, all available component values (metabolites, proteins, etc.) are averaged (or user-defined??). Nodes for which values were inferred will be marked by a dashed border during visualization to clearly show which values are known and which were inferred.

## 3. Collapsing reactions with missing expression or abundance user data

After data mapping is complete, Metaboverse will generate a collapsed network representation for optional viewing during later visualization. We did, however, choose to enforce a limit of up to three reactions that can be collapsed as data down a pathway should only be inferred so far. We also enforced certain parameters for reaction collapse as follows:

1. If a reaction has at least one known or inferred value for inputs (reactants) and one known or inferred value for outputs (products), the reaction will be left as is. During the entire reaction collapse step, known catalysts are included when assessing whether a reaction has measured output values (more of a catalyst should lead to more output in most cases) and inhibitors are included when assessing whether the reaction has measured input values (more inhibitor should lead to accumulation of input in most cases). Catalysts and inhibitors are not included when determining reaction neighbors as described below.
2. If a reaction has at least one known input, the input is left as is, and each reaction that shares the same input with the assessed reaction inputs are determined whether they have a measured output. If the neighbor reaction does not contain an known output value, the reaction is left as is. If the neighbor reaction does contain a measured output, the original reaction's inputs and the neighbor reactions outputs are collapsed to form a single, pseudo-reaction between the two. If the reaction has at least one known output, the inverse is performed where neighbors with identical components as the reactions inputs are assessed for whether a collapsed reaction can be created.
3. If a reaction has no measured values, it is determined if the neighboring reactions on both sides (one sharing the reaction's inputs and other sharing the reaction's outputs) have measured values. If both neighbors contain a measured value, a collapsed pseudo-reaction is created summarizing all three reactions.

For pseudo-reactions, appropriate notes are included to describe the collapse. During visualization, these pseudo-reactions are marked by black dashed edges and dashed node borders. A graphical representation of how this reaction collapse is performed can be found in Figure [4](#).



**Figure 4: Reaction node collapse schematic.** (a) For reactions where at least one input and at least one output component contain a measured value from the user data, the reaction will be maintained as is. (b) Where an input of a reaction is known but no output has a known value, Metaboverse will search for all neighboring reactions that contain identical inputs. If the neighboring reaction has a known output value, the two reactions will be merged into one pseudo-reaction. (c) Where an output of a reaction is known but no input has a known value, Metaboverse will search for all neighboring reactions that contain identical outputs. If the neighboring reaction has a known input value, the two reactions will be merged into one pseudo-reaction. (d) For reactions with no known values, neighbor pairs that match the inputs and outputs of the considered reaction will be evaluated for whether their respective outputs and inputs both have known values. If values are known for both neighbors, the three reactions will be merged into one pseudo-reaction. (e) As in (d), but if one neighbor does not contain a value, the one does contain a value, no reaction merging will be performed. (f) As in (d), but if neither neighbors contain known values, no reaction merging will be performed. An asterisk (\*) indicates the target reaction being considered for a given reaction collapse. A red node indicates a reaction input or output with a measured value. A white node indicates a reaction input or output with no measured value. A grey node indicates a reaction. A grey node with a dashed border indicates a pseudo-reaction. A solid edge indicates a known relationship. A dashed edge indicates a relationship inferred via reaction merging.

## 4. Regulatory pattern (motif) searches

## 5. Nearest neighborhood searches

In order to visualize all global connections, a user can select an entity (a gene, protein, or metabolite) and visualize all reactions that the component is involved in. By doing so, the user can visualize other downstream effects a change of one entity might have across the global network, which consequently aids in bridging and identifying any motifs that may occur between canonically annotated pathways. These neighborhoods can be expanded to view multiple downstream reaction steps and their accompanying genes, proteins, and metabolites by modulating the appropriate user option in the app.

Users can also limit which entities are shown by enforcing a degree threshold. By setting this value at 50, for example, the graph would not show nodes that have 50 or more connections. One caveat, however, is that this will occasionally break synchronous pathways into multiple pieces if one of these high-degree nodes was the bridge between two sides of a pathway.

## 6. Perturbation connectivity

Perturbation connectivity networks are generated by searching each reaction in the global network for any reaction where at least one component is significantly perturbed. Users can modify the necessary criteria to base the search on the expression or abundance value or the statistical value, and can choose the thresholding value to be used. For the expression thresholding, the provided value is assumed to be the absolute value, so a thresholding value of 3 would include any reactions where at least one component, showed a greater than 3 measured change or less than -3 measured change, the value of which is dependent on the data provided by the user, thus could represent  $\log_2$  fold changes, z-scores, or any other unit appropriate the biological context.

Once a list of perturbed reactions is collected, the network is constructed included each of these reactions and their components. Perturbed neighboring reactions that share components are thus connected within the graph and perturbed reactions that are not next to other perturbed reactions are shown as disconnected sub-graphs.

## 7. Network visualization and exploration

### 7.1 Dynamic network plotting

Users interact with Metaboverse through an interactive app interface. The app uses Electron, a cross-platform app framework that uses JavaScript, HTML, and CSS to design the interface. Metaboverse thus comes packaged as a single executable app with all necessary dependencies included for running on Linux, MacOS, and Windows.

Interactive graphing is handled using the D3 and JQuery JavaScript libraries. Force-directed graphs are constructed by taking the user selection for a pathway or entity and determining the reactions that are components of that pathway. All inputs, outputs, modifiers, and other components of these reactions, along with edges where both source and target are found in the sub-graph as nodes, are included and plotted. Relevant metadata, such as user-provided data and reaction descriptions, can be accessed by the user in real time. Metadata for categorical displays, such as edge or node type, are extracted from the metadata during graphing of the sub-network.



Some performance optimization features are included by default to prevent computational overload. For example, nearest neighbor sub-graphs with more than 1500 nodes, or nodes with more than 500 edges will not be plotted as plotting of this information in real-time is computationally prohibitive.

## 7.2 Visualizing pathways and super-pathways

In order to visualize a pathway, a user selects their pathway of choice and all component reactions and their reactants, products, modifiers, and metadata are parsed from the global network. Super-pathways help categorize these pathways and are defined as any pathway containing more than 200 nodes.

## 7.3 Visualizing compartments

Compartments are derived from Reactome annotations. Compartment visualizations are generated using D3's hull plotting feature. Compartment boundaries are defined at the reaction levels and made to encompass each reaction's reactants, products, and modifiers for that given compartment.

## 7.4 Annotations

Annotations for each reaction are derived from the Reactome database. Pseudo-reactions annotations do not include this information; instead they include notes on which reactions were collapsed to create the selected pseudo-reaction. All inferred pseudo-reactions and protein or complex values are displayed with dashed edges to differentiation from measured values.

## 7.6 Additional features

While Metaboverse will continue to undergo development and new features will be added, we will briefly highlight some additional features available at time of publication. We encourage users to check the documentation for more current updates and information regarding use of Metaboverse [[13](#)].

### 7.6a Toggle genes

As gene components can crowd the graph space, users can toggle gene display on and off using the appropriate button. The graph is then refreshed to either include or ignore gene components based on their node meta-tag.

### 7.6b Toggling values

Users can switch between coloring nodes based on the value or statistic provided by toggling the appropriate button. Colorbar information for the dataset is saved in the graph metadata during curation and used to generate a colorbar. The colorbar for statistics is represented using a  $-\log_{10}$  scale for a statistic value originally ranging between 0 and 1.

### 7.6c Toggling features/labels

By default, reaction and feature labels are displayed by hovering the mouse over the node. Reaction or feature nodes can have the labels statically displayed by selecting the appropriate button. An event watch function is used to watch for this user selection and update the display of the node labels.

### 7.6d Toggling collapsed reactions

By selecting the appropriate button, users can toggle between displaying a full or collapsed pathway representation of the sub-network. By selecting this button, the graph is refreshed using the appropriate reaction dictionary, where for graphing of the collapsed representation, a reaction with available pseudo-reactions substituted for the original reactions are included for graph propagation.

### 7.6e View curated pathway image

While Metaboverse graphs networks dynamically, users may be more familiar or comfortable with classical, curated pathway layouts when exploring their data. For a given pathway graph, the user can select the appropriate button and Metaboverse will open a new window with the Reactome curated pathway layout.

### 7.6f Saving graphs

Users can generate a PNG output file for any network created in Metaboverse by selecting the appropriate button.

### 7.6g Nearest neighbor and hub thresholding

The number of nearest neighbors to graph, or the limit to number of edges a graphed node can have, can be modulated by the user using the appropriate input spaces. When graphing a nearest neighbors network, Metaboverse will recursively fetch related reactions and their neighbors until a node display threshold is reached. This allows the user to visualize downstream effects of a change that may propagate across several reactions. The hub threshold option prevents plotting of nodes with more than the specified number of edges. This is handling during graphing by excluding any entity nodes that meet this criteria as the neighborhood is propagated. This is particularly useful in removing hub nodes, such as water or protons, which may be less relevant to the user experience and can quickly clutter the graph. This feature can also help plot more extensive neighborhoods, as often neighborhoods quickly link to high-degree nodes, such as water, and limit graphing ability.

### 7.6h Metadata display

To help inform the user of selection information and relevant metadata, a space in the legend bar during visualization is reserved for spaces where this information can be displayed, which is updated based on the user's input as it is provided.

## 8. Packaging

The Metaboverse app is packaged using Electron. Back-end network curation and data processing is performed using Python and the NetworkX library. Front-end visualization is performed using Javascript and relies on the D3 and JQuery packages. Saving network representations to a PNG file is performed using the d3-save-svg and string-pixel-width packages (Table 1). Documentation for Metaboverse is found at [metaboverse.readthedocs.io](https://metaboverse.readthedocs.io). Continuous integration services are performed by Travis CI to routinely run test cases for each change made to the Metaboverse architecture. The Metaboverse source code can be accessed at <https://github.com/Metaboverse/metaboverse>. The code used to draft and revise this manuscript, as well as all associated scripts used to generate and visualize the data presented in this manuscript can be accessed at <https://github.com/Metaboverse/manuscript>.

**Table 1:** Dependencies table.

Name	Reference
HTML	N/A
CSS	N/A
Javascript	N/A
Electron	[28]
JQuery	[29]
D3	[30]
string-pixel-width	[31]
d3-save-svg	[32]
Python	[33]
pandas	[34,35]
numpy	[36]
scipy	[37]
matplotlib	[38]
NetworkX	[39]

## 9. Validation using biological data

### 9.1 *mct1* perturbation in *Saccharomyces cerevisiae*

#### Yeast Strains

*Saccharomyces cerevisiae* BY4743 (MATa/A, his3/his3, leu2/leu2, ura3/ura3, met15/MET15, lys2/LYS2) was used to generate the *mct1*Δ strain as described in [40].

#### RNA-sequencing sample preparation

RNA sequencing data were generated by growing *Saccharomyces cerevisiae* biological replicates for strains *mct1*Δ (n=4) and wild-type (n=4). Briefly, cells were grown in glucose and switched to raffinose-supplemented growth medium for 0, 3, and 12 hours such that at time of harvest, cultures were at OD<sub>600</sub>=1. Cultures were flash frozen and later total RNA was isolated using the Direct-zol kit (Zymo Research) with on-column DNase digestion and water elution. Sequencing libraries were prepared by purifying intact poly(A) RNA from total RNA samples (100-500 ng) with oligo(dT) magnetic beads and stranded mRNA sequencing libraries were prepared as described using the Illumina TruSeq Stranded mRNA Library Preparation Kit (RS-122-2101, RS-122-2102). Purified libraries were qualified on an Agilent Technologies 2200 TapeStation using a D1000 ScreenTape assay (cat# 5067-5582 and 5067-5583). The molarity of adapter-modified molecules was defined by quantitative PCR using the Kapa Biosystems Kapa Library Quant Kit (cat#KK4824). Individual libraries were normalized to 5 nM and equal volumes were pooled in preparation for Illumina sequence analysis. Sequencing libraries (25 pM) were chemically denatured and applied to an Illumina HiSeq v4 single read flow cell using an Illumina cBot. Hybridized molecules were clonally amplified and annealed to sequencing primers with reagents from an Illumina HiSeq SR Cluster Kit v4-cBot (GD-401-4001). Following transfer of the

flowcell to an Illumina HiSeq 2500 instrument (HCSv2.2.38 and RTA v1.18.61), a 50 cycle single-read sequence run was performed using HiSeq SBS Kit v4 sequencing reagents (FC-401-4002).

### Sequence analysis

Sequence FASTQ files were processed using XPRESSpipe [41]. Batch and log files are available at [42]. Notably, reads were trimmed of adapters (Read1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA, Read2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). Based on library complexity quality control, de-duplicated alignments were used for read quantification due to the high number of duplicated sequences in each library. Differential expression analysis was performed using DESeq2 [43] by comparing *mct1Δ* samples with wild-type samples at the 12-hour time-point to match the steady state proteomics data.  $\log_2$ (fold change) and false discovery rate ("p-adj") values were extracted from the DESeq2 output.

### Proteomics analysis

Steady state quantitative proteomics data were generated as described in [40]. Briefly, cells were grown in glucose and switched to raffinose-supplemented growth medium overnight, and harvested at mid-log phase. For this analysis, we compared the *mct1Δ* (n=3) with the wild-type (n=3) cell populations.  $\log_2$ (fold change) values and Benjamini-Hochberg corrected p-values were generated by comparing *mct1Δ* with the wild-type cells. P-values were generated using a 2-tailed, homoscedastic Student's T-test.

### Metabolomics sample preparation

Metabolomics data were generated by growing the appropriate yeast strains in synthetic minimal media (S-min) supplemented with 2% glucose until they reached OD<sub>600</sub> between 0.6-0.8. Cells were then transferred to S-min media containing 2% raffinose and harvested after 0, 15, 30, 60, and 180 minutes (n=6/time-point/strain).

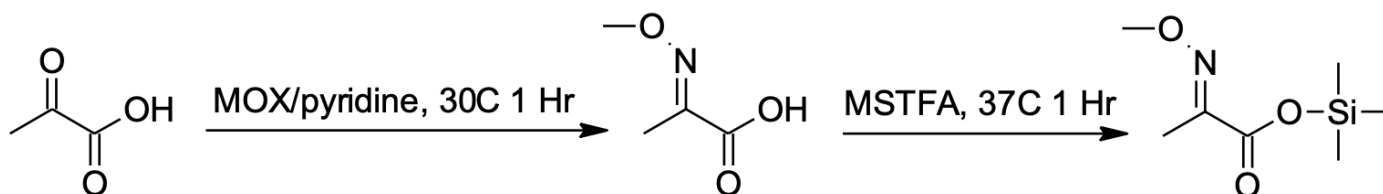
### Metabolite extraction

A 75% boiling ethanol (EtOH) solution containing the internal standard d4-succinic acid (Sigma 293075) was then added to each sample. Boiling samples were vortexed and incubated at 90°C for 5 min. Samples were then incubated at -20 °C for 1 hr. After incubation, samples were centrifuged at 5,000 x g for 10 minutes at 4 °C. The supernatant was then transferred from each sample tube into a labeled, fresh 13x100mm glass culture tube. A second standard was then added (d27-myristic acid CDN Isotopes: D-1711). Pooled quality control samples were made by removing a fraction of collected supernatant from each sample and process blanks were made using only extraction solvent and no cell culture. The samples were then dried en vacuo. This process was completed in three separate batches.

### Mass Spectrometry Analysis of Samples

All GC-MS analysis was performed with an Agilent 5977b GC-MS MSD-HES and an Agilent 7693A automatic liquid sampler. Dried samples were suspended in 40 μL of a 40 mg/mL O-methoxylamine hydrochloride (MOX) (MP Bio #155405) in dry pyridine (EMD Millipore #PX2012-7) and incubated for one hour at 37 °C in a sand bath. 25 μL of this solution was added to auto sampler vials. 60 μL of N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA with 1%TMCS, Thermo #TS48913) was added automatically via the auto sampler and incubated for 30 minutes at 37 °C. After incubation, samples were vortexed and 1 μL of the prepared sample was injected into the gas chromatograph inlet in the split mode with the inlet temperature held at 250°C. A 10:1 split ratio was used for analysis of the

majority of metabolites. For those metabolites that saturated the instrument at the 10:1 split concentration, a split of 50:1 was used for analysis. The gas chromatograph had an initial temperature of 60°C for one minute followed by a 10°C/min ramp to 325°C and a hold time of 5 minutes. A 30-meter Phenomenex Zebron AB-5HT with 5m inert Guardian capillary column was employed for chromatographic separation. Helium was used as the carrier gas at a rate of 1 mL/min. Figure 5 demonstrates an example of the two-step derivatization process used to convert non-volatile metabolites to a volatile form amenable to GC-MS.



**Figure 5: Mass Spectrometry derivatization process.** This figure demonstrates an example of the two-step derivatization process used to convert non-volatile metabolites to a volatile form amenable to GC-MS. Pyruvic acid is used here as an example.

### Analysis of Mass Spectrometry Data

Data was collected using MassHunter software (Agilent). Metabolites were identified and their peak area was recorded using MassHunter Quant. This data was transferred to an Excel spreadsheet (Microsoft, Redmond WA). Metabolite identity was established using a combination of an in-house metabolite library developed using pure purchased standards, the NIST library and the Fiehn library. There are a few reasons a specific metabolite may not be observable through GC-MS. The metabolite may not be amenable to GC-MS due to its size, or a quaternary amine such as carnitine, or simply because it does not ionize well. Metabolites that do not ionize well include oxaloacetate, histidine and arginine. Cysteine can be observed depending on cellular conditions. It often forms disulfide bonds with proteins and is generally at a low concentration. Metabolites may not be quantifiable if they are only present in very low concentrations.

Resulting data from all samples were normalized to the internal standard d4-succinate. Samples highlighted in yellow were swapped with normalized data from the unsaturated run. The identity of each peak was ensured by visualization in Mass Hunter Qual and Quant. The data has also been roughly reordered by metabolic process. False positives were removed. All but the best representative of duplicate metabolites, which were created through variability in the derivatization process, were also removed.

## 10. Data availability.

Raw data produced for this study are in the process of being uploaded to the appropriate repositories. In the meantime, the raw data can be accessed at [42].

The curated networks for these data are available at [42]. Networks were generated by taking the 12-hour transcriptomics and proteomics datasets with their appropriate  $\log_2$ (fold change) and statistical values, along with the 0, 15, 30, 60, and 180 minute metabolomics datasets with their respective  $\log_2$ (fold change) and statistical values and layering these data on the *Saccharomyces cerevisiae* global reaction network as curated by Metaboverse from the Reactome database. Reaction motifs and global connectivity analyses were performed within the Metaboverse platform.

The Metaboverse source code is available at [44] and archived versions can be accessed at Zenodo [1].

The source code and data for this manuscript and subsequent analyses is available at [45] and archived versions can be accessed at Zenodo [].

## Acknowledgements

---

We thank Alex J. Bott, Ahmad A. Cluntun, Kevin Hicks, and other members of the Rutter lab for their thoughtful insights and suggestions. We thank the Brian Dalley and the University of Utah High-Throughput Genomics Core for help with RNA-sequencing library preparation. The support and resources from the Center for High-Performance Computing at the University of Utah are gratefully acknowledged.

## Funding

---

J.A.B. received support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Inter-disciplinary Training Grant T32 Program in Computational Approaches to Diabetes and Metabolism Research, 1T32DK11096601 to Wendy W. Chapman and Simon J. Fisher (<https://www.niddk.nih.gov/>). This work was supported by NIDDK fellowship 1T32DK11096601 (to J.A.B.) (<https://www.niddk.nih.gov/>) and NIH grant R35GM13185 (to J.R.) (<https://www.nih.gov/>). The computational resources used were partially funded by the NIH Shared Instrumentation Grant 1S10OD021644-01A1 (<https://www.nih.gov/>). Mass spectrometry equipment was obtained through NCR Shared Instrumentation Grant 1S10OD016232-01, 1S10OD018210-01A1, and 1S10OD021505-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Contributions

---

- Conceptualization: J.A.B., T.C.W., B.W., J.R.
- Supervision: J.A.B., B.W., J.R.
- Project Administration: J.A.B.
- Investigation: J.A.B., T.C.W., Y.O., S.M.N., T.V.
- Formal Analysis: J.A.B.
- Software: J.A.B., Y.Z.
- Methodology: J.A.B., Y.Z.
- Validation: J.A.B.
- Data Curation: J.A.B., T.C.W.
- Resources: J.A.B., J.C., B.W., J.R.
- Funding Acquisition: J.A.B., J.C., B.W., J.R.
- Writing - Original Draft Preparation: J.A.B.
- Writing - Review & Editing: J.A.B., Y.Z., T.C.W., Y.O., S.N.M., T.V., J.C., B.W., J.R.
- Visualization: J.A.B.





## References

---

**1. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community**

Farhana R. Pinu, David J. Beale, Amy M. Paten, Konstantinos Kouremenos, Sanjay Swarup, Horst J. Schirra, David Wishart

*Metabolites* (2019-04-18) <https://doi.org/ggt2fd>

DOI: [10.3390/metabo9040076](https://doi.org/10.3390/metabo9040076) · PMID: [31003499](https://pubmed.ncbi.nlm.nih.gov/31003499/) · PMCID: [PMC6523452](https://pubmed.ncbi.nlm.nih.gov/PMC6523452/)

**2. MetaboAnalyst: a web server for metabolomic data analysis and interpretation**

J. Xia, N. Psychogios, N. Young, D. S. Wishart

*Nucleic Acids Research* (2009-05-08) <https://doi.org/c4hzc2>

DOI: [10.1093/nar/gkp356](https://doi.org/10.1093/nar/gkp356) · PMID: [19429898](https://pubmed.ncbi.nlm.nih.gov/19429898/) · PMCID: [PMC2703878](https://pubmed.ncbi.nlm.nih.gov/PMC2703878/)

**3. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis**

Jasmine Chong, Othman Soufan, Carin Li, Iurie Caraus, Shuzhao Li, Guillaume Bourque, David S Wishart, Jianguo Xia

*Nucleic Acids Research* (2018-05-14) <https://doi.org/gdwvwd>

DOI: [10.1093/nar/gky310](https://doi.org/10.1093/nar/gky310) · PMID: [29762782](https://pubmed.ncbi.nlm.nih.gov/29762782/) · PMCID: [PMC6030889](https://pubmed.ncbi.nlm.nih.gov/PMC6030889/)

**4. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks**

P. Shannon

*Genome Research* (2003-11-01) <https://doi.org/b7kgpg>

DOI: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303) · PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/) · PMCID: [PMC403769](https://pubmed.ncbi.nlm.nih.gov/PMC403769/)

**5. MetExplore: collaborative edition and exploration of metabolic networks**

Ludovic Cottret, Clément Frainay, Maxime Chazalviel, Floréal Cabanettes, Yoann Gloaguen, Etienne Camenen, Benjamin Merlet, Stéphanie Heux, Jean-Charles Portais, Nathalie Poupin, ... Fabien Jourdan

*Nucleic Acids Research* (2018-04-30) <https://doi.org/gdwx39>

DOI: [10.1093/nar/gky301](https://doi.org/10.1093/nar/gky301) · PMID: [29718355](https://pubmed.ncbi.nlm.nih.gov/29718355/) · PMCID: [PMC6030842](https://pubmed.ncbi.nlm.nih.gov/PMC6030842/)

**6. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks**

L. Cottret, D. Wildridge, F. Vinson, M. P. Barrett, H. Charles, M.-F. Sagot, F. Jourdan

*Nucleic Acids Research* (2010-05-05) <https://doi.org/dkx7z7>

DOI: [10.1093/nar/gkq312](https://doi.org/10.1093/nar/gkq312) · PMID: [20444866](https://pubmed.ncbi.nlm.nih.gov/20444866/) · PMCID: [PMC2896158](https://pubmed.ncbi.nlm.nih.gov/PMC2896158/)

**7. Reactome: a knowledgebase of biological pathways**

G. Joshi-Tope

*Nucleic Acids Research* (2004-12-17) <https://doi.org/cjhf62>

DOI: [10.1093/nar/gki072](https://doi.org/10.1093/nar/gki072) · PMID: [15608231](https://pubmed.ncbi.nlm.nih.gov/15608231/) · PMCID: [PMC540026](https://pubmed.ncbi.nlm.nih.gov/PMC540026/)

**8. The reactome pathway knowledgebase.**

Bijay Jassal, Lisa Matthews, Guilherme Viteri, Chuqiao Gong, Pascual Lorente, Antonio Fabregat, Konstantinos Sidiropoulos, Justin Cook, Marc Gillespie, Robin Haw, ... Peter D'Eustachio

*Nucleic acids research* (2020-01-08) <https://www.ncbi.nlm.nih.gov/pubmed/31691815>

DOI: [10.1093/nar/gkz1031](https://doi.org/10.1093/nar/gkz1031) · PMID: [31691815](https://pubmed.ncbi.nlm.nih.gov/31691815/) · PMCID: [PMC7145712](https://pubmed.ncbi.nlm.nih.gov/PMC7145712/)

## 9. The Reactome Pathway Knowledgebase.

Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korninger, Bruce May, ... Peter D'Eustachio  
*Nucleic acids research* (2018-01-04) <https://www.ncbi.nlm.nih.gov/pubmed/29145629>  
DOI: [10.1093/nar/gkx1132](https://doi.org/10.1093/nar/gkx1132) · PMID: [29145629](https://pubmed.ncbi.nlm.nih.gov/29145629/) · PMCID: [PMC5753187](https://pubmed.ncbi.nlm.nih.gov/PMC5753187/)

## 10. Superfamilies of Evolved and Designed Networks

R. Milo  
*Science* (2004-03-05) <https://doi.org/10.1126/science.1089167>  
DOI: [10.1126/science.1089167](https://doi.org/10.1126/science.1089167) · PMID: [15001784](https://pubmed.ncbi.nlm.nih.gov/15001784/)

## 11. Network motifs in the transcriptional regulation network of Escherichia coli

Shai S. Shen-Orr, Ron Milo, Shmoolik Mangan, Uri Alon  
*Nature Genetics* (2002-04-22) <https://doi.org/10.1038/ng881>  
DOI: [10.1038/ng881](https://doi.org/10.1038/ng881) · PMID: [11967538](https://pubmed.ncbi.nlm.nih.gov/11967538/)

## 12. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network

Gal Chechik, Eugene Oh, Oliver Rando, Jonathan Weissman, Aviv Regev, Daphne Koller  
*Nature Biotechnology* (2008-10-26) <https://doi.org/10.1038/nbt.1499>  
DOI: [10.1038/nbt.1499](https://doi.org/10.1038/nbt.1499) · PMID: [18953355](https://pubmed.ncbi.nlm.nih.gov/18953355/) · PMCID: [PMC2651818](https://pubmed.ncbi.nlm.nih.gov/PMC2651818/)

## 13. Metaboverse — Metaboverse 0.0.0-alpha documentation <https://metaboverse.readthedocs.io/en/latest/>

## 14. Using graph theory to analyze biological networks

Georgios A Pavlopoulos, Maria Secrier, Charalampos N Moschopoulos, Theodoros G Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider, Pantelis G Bagos  
*BioData Mining* (2011-04-28) <https://doi.org/10.1186/1756-0381-4-10>  
DOI: [10.1186/1756-0381-4-10](https://doi.org/10.1186/1756-0381-4-10) · PMID: [21527005](https://pubmed.ncbi.nlm.nih.gov/21527005/) · PMCID: [PMC3101653](https://pubmed.ncbi.nlm.nih.gov/PMC3101653/)

## 15. Compartment and hub definitions tune metabolic networks for metabolomic interpretations

T Cameron Waller, Jordan A Berg, Alexander Lex, Brian E Chapman, Jared Rutter  
*GigaScience* (2020-01-01) <https://doi.org/10.1093/gigascience/giz137>  
DOI: [10.1093/gigascience/giz137](https://doi.org/10.1093/gigascience/giz137) · PMID: [31972021](https://pubmed.ncbi.nlm.nih.gov/31972021/) · PMCID: [PMC6977586](https://pubmed.ncbi.nlm.nih.gov/PMC6977586/)

## 16. Functional and transcriptional connectivity of communities in breast cancer co-expression networks

Guillermo de Anda-Jáuregui, Sergio Antonio Alcalá-Corona, Jesús Espinal-Enríquez, Enrique Hernández-Lemus  
*Applied Network Science* (2019-05-20) <https://doi.org/10.1007/s41109-019-0129-0>  
DOI: [10.1007/s41109-019-0129-0](https://doi.org/10.1007/s41109-019-0129-0)

## 17. Loss of Connectivity in Cancer Co-Expression Networks

Roberto Anglani, Teresa M. Creanza, Vania C. Liuzzi, Ada Piepoli, Anna Panza, Angelo Andriulli, Nicola Ancona  
*PLoS ONE* (2014-01-28) <https://doi.org/10.1371/journal.pone.0087075>  
DOI: [10.1371/journal.pone.0087075](https://doi.org/10.1371/journal.pone.0087075) · PMID: [24489837](https://pubmed.ncbi.nlm.nih.gov/24489837/) · PMCID: [PMC3904972](https://pubmed.ncbi.nlm.nih.gov/PMC3904972/)

**18. A Mitochondrial Pyruvate Carrier Required for Pyruvate Uptake in Yeast, Drosophila, and Humans**

Daniel K. Bricker, Eric B. Taylor, John C. Schell, Thomas Orsak, Audrey Boutron, Yu-Chan Chen, James E. Cox, Caleb M. Cardon, Jonathan G. Van Vranken, Noah Dephore, ... Jared Rutter  
*Science* (2012-05-24) <https://doi.org/f33x88>  
DOI: [10.1126/science.1218099](https://doi.org/10.1126/science.1218099) · PMID: [22628558](https://pubmed.ncbi.nlm.nih.gov/22628558/) · PMCID: [PMC3690818](https://pubmed.ncbi.nlm.nih.gov/PMC3690818/)

**19. Control of intestinal stem cell function and proliferation by mitochondrial pyruvate metabolism**

John C. Schell, Dona R. Wisidagama, Claire Bensard, Helong Zhao, Peng Wei, Jason Tanner, Aimee Flores, Jeffrey Mohlman, Lise K. Sorensen, Christian S. Earl, ... Jared Rutter  
*Nature Cell Biology* (2017-08-14) <https://doi.org/gbrf73>  
DOI: [10.1038/ncb3593](https://doi.org/10.1038/ncb3593) · PMID: [28812582](https://pubmed.ncbi.nlm.nih.gov/28812582/) · PMCID: [PMC6137334](https://pubmed.ncbi.nlm.nih.gov/PMC6137334/)

**20. Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier**

Claire L. Bensard, Dona R. Wisidagama, Kristofor A. Olson, Jordan A. Berg, Nathan M. Krah, John C. Schell, Sara M. Nowinski, Sarah Fogarty, Alex J. Bott, Peng Wei, ... Jared Rutter  
*Cell Metabolism* (2020-02) <https://doi.org/ggt7w8>  
DOI: [10.1016/j.cmet.2019.11.002](https://doi.org/10.1016/j.cmet.2019.11.002) · PMID: [31813825](https://pubmed.ncbi.nlm.nih.gov/31813825/) · PMCID: [PMC7004878](https://pubmed.ncbi.nlm.nih.gov/PMC7004878/)

**21. Redundancy: A Critical Obstacle to Improving Cancer Therapy**

O. Lavi  
*Cancer Research* (2015-01-09) <https://doi.org/ggt7xd>  
DOI: [10.1158/0008-5472.can-14-3256](https://doi.org/10.1158/0008-5472.can-14-3256) · PMID: [25576083](https://pubmed.ncbi.nlm.nih.gov/25576083/) · PMCID: [PMC6250436](https://pubmed.ncbi.nlm.nih.gov/PMC6250436/)

**22. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease**

J. Lamb  
*Science* (2006-09-29) <https://doi.org/c92ptt>  
DOI: [10.1126/science.1132939](https://doi.org/10.1126/science.1132939) · PMID: [17008526](https://pubmed.ncbi.nlm.nih.gov/17008526/)

**23. Pathway crosstalk perturbation network modeling for identification of connectivity changes induced by diabetic neuropathy and pioglitazone**

Guillermo de Anda-Jáuregui, Kai Guo, Brett A. McGregor, Eva L. Feldman, Junguk Hur  
*BMC Systems Biology* (2019-01-07) <https://doi.org/ggt7zx>  
DOI: [10.1186/s12918-018-0674-7](https://doi.org/10.1186/s12918-018-0674-7) · PMID: [30616626](https://pubmed.ncbi.nlm.nih.gov/30616626/) · PMCID: [PMC6322225](https://pubmed.ncbi.nlm.nih.gov/PMC6322225/)

**24. Impact of Mitochondrial Fatty Acid Synthesis on Mitochondrial Biogenesis**

Sara M. Nowinski, Jonathan G. Van Vranken, Katja K. Dove, Jared Rutter  
*Current Biology* (2018-10) <https://doi.org/gffn55>  
DOI: [10.1016/j.cub.2018.08.022](https://doi.org/10.1016/j.cub.2018.08.022) · PMID: [30352195](https://pubmed.ncbi.nlm.nih.gov/30352195/) · PMCID: [PMC6258005](https://pubmed.ncbi.nlm.nih.gov/PMC6258005/)

**25. The mitochondrial acyl carrier protein (ACP) coordinates mitochondrial fatty acid synthesis with iron sulfur cluster biogenesis**

Jonathan G Van Vranken, Mi-Young Jeong, Peng Wei, Yu-Chan Chen, Steven P Gygi, Dennis R Winge, Jared Rutter  
*eLife* (2016-08-19) <https://doi.org/ggvrxc>  
DOI: [10.7554/elife.17828](https://doi.org/10.7554/elife.17828) · PMID: [27540631](https://pubmed.ncbi.nlm.nih.gov/27540631/) · PMCID: [PMC4991935](https://pubmed.ncbi.nlm.nih.gov/PMC4991935/)

26. **MECR Mutations Cause Childhood-Onset Dystonia and Optic Atrophy, a Mitochondrial Fatty Acid Synthesis Disorder**  
Gali Heimer, Juha M. Kerätär, Lisa G. Riley, Shanti Balasubramaniam, Eran Eyal, Laura P. Pietikäinen, J. Kalervo Hiltunen, Dina Marek-Yagel, Jeffrey Hamada, Allison Gregory, ... Qian Yi  
*The American Journal of Human Genetics* (2016-12) <https://doi.org/f9gbs7>  
DOI: [10.1016/j.ajhg.2016.09.021](https://doi.org/10.1016/j.ajhg.2016.09.021) · PMID: [27817865](https://pubmed.ncbi.nlm.nih.gov/27817865/) · PMCID: [PMC5142118](https://pubmed.ncbi.nlm.nih.gov/PMC5142118/)
27. **Mitochondrial fatty acid synthesis coordinates mitochondrial oxidative metabolism**  
Sara M Nowinski, Ashley Solmonson, Scott F Rusin, J. Alan Maschek, Claire L. Bensard, Sarah Fogarty, Mi-Young Jeong, Sandra Lettlova, Jordan A Berg, Jeffrey T Morgan, ... Jared Rutter  
*Cold Spring Harbor Laboratory* (2020-05-10) <https://doi.org/ggvq48>  
DOI: [10.1101/2020.05.09.086199](https://doi.org/10.1101/2020.05.09.086199)
28. **Electron | Build cross-platform desktop apps with JavaScript, HTML, and CSS.** <https://www.electronjs.org/>
29. **jQuery**  
JS Foundation- js.foundation  
<https://jquery.com/>
30. **D3.js - Data-Driven Documents**  
Mike Bostock  
<https://d3js.org/>
31. **adambisek/string-pixel-width**  
GitHub  
<https://github.com/adambisek/string-pixel-width>
32. **edeno/d3-save-svg**  
GitHub  
<https://github.com/edeno/d3-save-svg>
33. **Welcome to Python.org**  
Python.org  
<https://www.python.org/>
34. **pandas-dev/pandas: Pandas 1.0.3**  
Jeff Reback, Wes McKinney, Jbrockmendel, Joris Van Den Bossche, Tom Augspurger, Phillip Cloud, Gfyoung, Sinhrks, Adam Klein, Matthew Roeschke, ... Mortada Mehyar  
*Zenodo* (2020-03-18) <https://doi.org/ggt8bh>  
DOI: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134)
35. **Data Structures for Statistical Computing in Python**  
Wes McKinney  
*Proceedings of the 9th Python in Science Conference* (2010) <https://doi.org/ggr6q3>  
DOI: [10.25080/majora-92bf1922-00a](https://doi.org/10.25080/majora-92bf1922-00a)
36. (2007-01-16) <https://web.mit.edu/dvp/Public/numpybook.pdf>

37. **SciPy 1.0: fundamental algorithms for scientific computing in Python**

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, ... Paul van Mulbregt  
*Nature Methods* (2020-02-03) <https://doi.org/ggj45f>  
DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2) · PMID: [32015543](https://pubmed.ncbi.nlm.nih.gov/32015543/) · PMCID: [PMC7056644](https://pubmed.ncbi.nlm.nih.gov/PMC7056644/)

38. **Matplotlib: A 2D Graphics Environment**

John D. Hunter  
*Computing in Science & Engineering* (2007) <https://doi.org/drjbhg>  
DOI: [10.1109/mcse.2007.55](https://doi.org/10.1109/mcse.2007.55)

39. (2014-01-28) [http://conference.scipy.org/proceedings/SciPy2008/paper\\_2/full\\_text.pdf](http://conference.scipy.org/proceedings/SciPy2008/paper_2/full_text.pdf)

40. **ACP Acylation Is an Acetyl-CoA-Dependent Modification Required for Electron Transport Chain Assembly**

Jonathan G. Van Vranken, Sara M. Nowinski, Katie J. Clowers, Mi-Young Jeong, Yeyun Ouyang, Jordan A. Berg, Jeremy P. Gygi, Steven P. Gygi, Dennis R. Winge, Jared Rutter  
*Molecular Cell* (2018-08) <https://doi.org/gd5dhm>  
DOI: [10.1016/j.molcel.2018.06.039](https://doi.org/10.1016/j.molcel.2018.06.039) · PMID: [30118679](https://pubmed.ncbi.nlm.nih.gov/30118679/) · PMCID: [PMC6104058](https://pubmed.ncbi.nlm.nih.gov/PMC6104058/)

41. **XPRESSyourself: Enhancing, standardizing, and automating ribosome profiling computational analyses yields improved insight into data**

Jordan A. Berg, Jonathan R. Belyeu, Jeffrey T. Morgan, Yeyun Ouyang, Alex J. Bott, Aaron R. Quinlan, Jason Gertz, Jared Rutter  
*PLOS Computational Biology* (2020-01-31) <https://doi.org/ggvg4j>  
DOI: [10.1371/journal.pcbi.1007625](https://doi.org/10.1371/journal.pcbi.1007625) · PMID: [32004313](https://pubmed.ncbi.nlm.nih.gov/32004313/) · PMCID: [PMC7015430](https://pubmed.ncbi.nlm.nih.gov/PMC7015430/)

42. **Metaboverse/manuscript**

GitHub  
<https://github.com/Metaboverse/manuscript>

43. **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**

Michael I Love, Wolfgang Huber, Simon Anders  
*Genome Biology* (2014-12) <https://doi.org/gd3zvn>  
DOI: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8) · PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/) · PMCID: [PMC4302049](https://pubmed.ncbi.nlm.nih.gov/PMC4302049/)

44. **Metaboverse/Metaboverse**

GitHub  
<https://github.com/Metaboverse/Metaboverse>

45. **Metaboverse/manuscript**

GitHub  
<https://github.com/Metaboverse/manuscript>