Gazing into the Metaboverse: Automated exploration and contextualization of metabolic data

Authors

Jordan A. Berg[1], Youjia Zhou[2,3], T. Cameron Waller[4], Yeyun Ouyang[1], Sara M. Nowinski[1], Tyler Van Ry[1,5], Ian George[1], James E. Cox[1,5,6], Bei Wang[†2,3], Jared Rutter[†1,6,7]

1. Department of Biochemistry, University of Utah
2. School of Computing, University of Utah
3. Scientific Computing and 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: jordan.berg@biochem.utah.edu; rutter@biochem.utah.edu

† Senior Author

Abstract

Metabolism and its component 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 an organism. Perturbations to any reaction or set of reactions can have rippling downstream effects, which can be challenging to trace across the total reaction network, particularly when the effects occur between complex interconnected pathways. Researchers have primarily utilized reductionist approaches to understand metabolic reaction systems; however, these simplistic methods often limit the scope of the analysis. Even systems-centric omics approaches can be limited when only a handful of high magnitude signals in the data are prioritized for interpretation. To address these challenges, we developed Metaboverse, an interactive tool for the exploration and automated extraction of potential regulatory events, patterns, and trends from multi-omic data within the context of the metabolic network and other reaction networks. This framework will be foundational in increasing our ability to holistically understand static, temporal, and multi-condition metabolic events and perturbations as well as 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, protein complexes, and metabolites [1]. To understand these complex components, researchers frequently adopt reductionist approaches to tease apart the characteristics and mechanics of these processes and determine how they fit into the larger picture of biology and disease. Such approaches are a vital component in the scientific process, but a reductionist approach may miss many essential properties of metabolism. 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 coordinated or rippling effects across the network [2,3]. Biological perturbations involve complex, cooperative effects, many of which may seem negligible in isolation. The current approach is analogous to telling the story of Little Red Riding Hood by reading only the 20 most frequent words. Indeed, doing so efficiently highlights keywords like “wolf” and “little red riding hood,” but also prevents a coherent story from being told. This approach makes it difficult for someone who has never heard the story of Little Red Riding Hood to comprehend the plot and especially the higher-level interpretations of the story.

Several computational tools have risen to prominence over the past decade to try to resolve these issues in data contextualization. We will highlight four such representative and popular tools for their respective properties, though many more exist [4]. First is MetaboAnalyst, which relies heavily on set enrichment methods for the analysis of data, or examining the belongingness of sets of significantly changed analytes (i.e., metabolite, protein, or gene measurements), for extracting interesting information. Network visualization is available, but it focuses primarily on interaction networks, and its ability to extract regulatory information is limited, particularly in an automated fashion [5,6]. Second is Cytoscape, which serves as a general go-to platform for representing biological or other networks. One strength of Cytoscape is the ability to design apps or plug-ins to conduct customized analyses; however, comprehensive and metabolism-specific regulatory identification methods are unavailable [7]. One plug-in for Cytoscape that focuses on metabolic data is MetScape, but again, this tool is generally limited to pathway enrichment, correlation networks, and data visualization and does not integrate approaches to identify regulatory mechanisms within the data [8,9,10]. MetExplore focuses on the curation of networks and is particularly useful for collaborative annotation of emerging

models of organisms with incomplete metabolic network curations. It additionally can layer experimental data on the network for visualization [11,12]. A companion tool to MetExplore is MetExploreViz, which enables interactive and flexible visualization of omics data on metabolic networks [13]. Reactome, which our tool 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 a pressing need for tools that integrate these features and automate pattern and trend detection across metabolic networks to extract regulatory and other features from data [14,15,16]. This need is particularly pronounced in common cases where experimental data have sparse coverage of the metabolic network, a common situation in current metabolomics datasets. To address these limitations in metabolic data analysis, contextualization, and interpretation, we created Metaboverse, a cross-platform, interactive application to aid users in exploring and interpreting their model's metabolic story and generating new hypotheses from the available information (Figure 1).

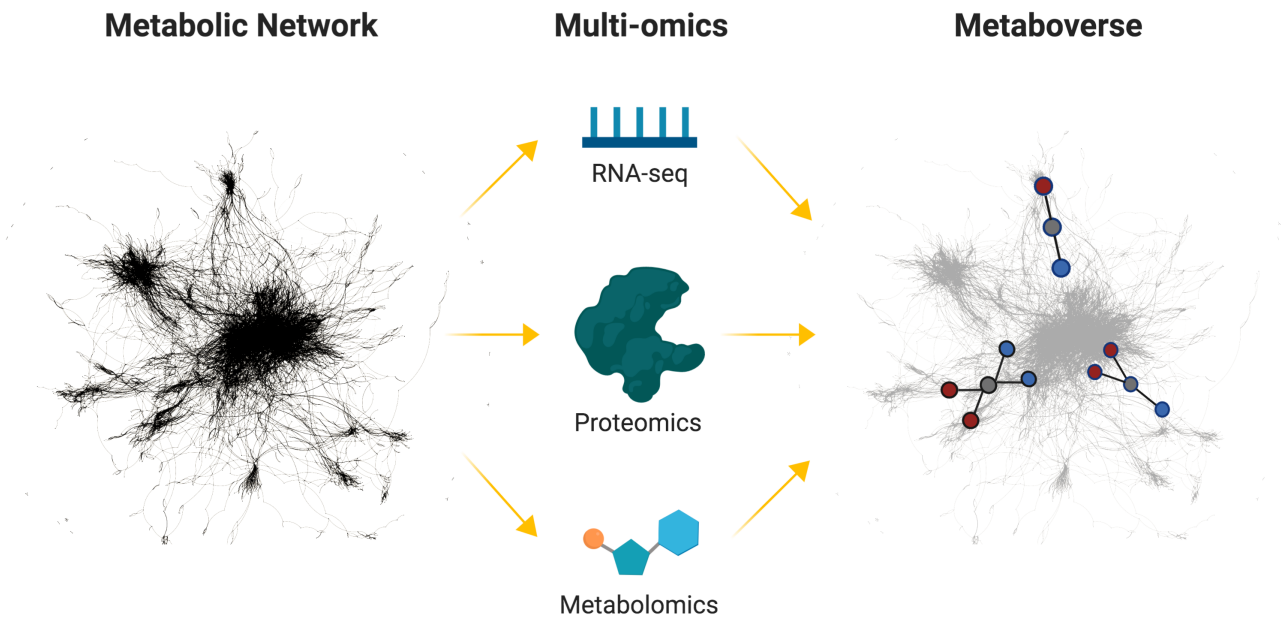


Figure 1: Metaboverse conceptual overview. Illustration summarizing the usage of Metaboverse to model biological data on the reaction network to rapidly identify regulatory patterns. Traditionally, when scientists perform an omics experiment, they tend to focus on a few features that are differentially regulated. Metaboverse allows a user to input multiple omics data types, which it layers upon the metabolic network. Metaboverse then uses this integrated model to identify patterns of putative regulatory potential within the data. The user can also dynamically explore metabolic pathways and other network representations.

To provide a robust platform for the exploration of single or multi-omic metabolic data in two-condition, multi-condition, and time-course contexts, we developed several computational features to address the aims discussed above. We developed a pattern search engine for the rapid and automated identification of different patterns and trends in omics data on the metabolic network. Metaboverse allows for the interactive exploration of specific reactions or reaction entities with on-the-fly pattern search analysis. The user can explore canonical pathways of interest and look for other interesting patterns and trends within these pathways. Perturbations in the abundance or behavior of a particular metabolic component can lead to downstream gene expression changes and other phenotypic modulations in a biological system. Users can therefore select a given component and explore the upstream and downstream reaction neighbors across the total reaction network. In addition, users can also use a drop-down menu to select any component and immediately view its reaction neighborhood. We also include an interactive perturbation exploration module, which displays all the perturbed reactions from the selected super-pathway. Finally, we tackle the challenge of sparsity, particularly in metabolomics datasets. We implement a reaction collapsing feature that

can help the user identify more inconspicuous patterns across multiple reactions within the total reaction network.

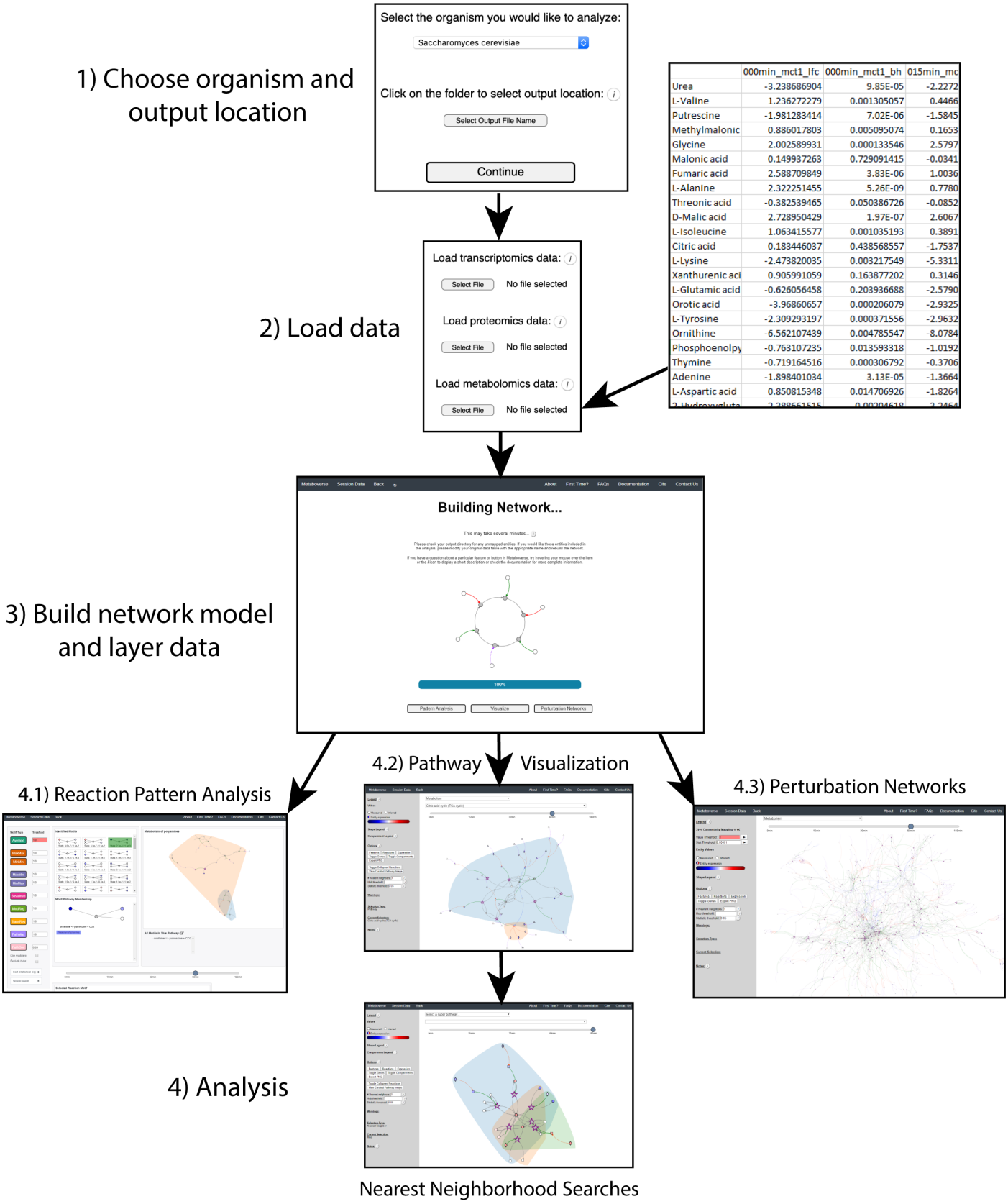
To validate the features and methods available in Metaboverse, we analyzed a variety of datasets to provide representative vignettes that highlight the reliability of Metaboverse for extracting canonical features, and its power for identifying novel features and patterns and generate new hypotheses. We expect that Metaboverse will become a foundational tool in the analytical toolkit and augment the user's ability to more holistically explore and contextualize metabolic reaction networks.

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 lightweight, self-contained, cross-platform app for the dynamic exploration of high-throughput biological data. The pathway curations are derived from Reactome [17], coupled with additional metabolite synonym cross-referencing and mapping from the ChEBI [18] and Human Metabolome (HMDB) [19] databases. As of the time of writing, Metaboverse is capable of analyzing data for over 90 species. A user begins by providing the desired output location for a new data model and specifying the organism of interest. If the user has previously curated data onto a reaction network for the selected organism of interest, they can alternatively provide this intermediate file. Next, the user provides the relevant transcriptomic, proteomic, and/or metabolomic datasets to layer onto the global reaction network of their organism of interest. Input data types 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 analysis upon the network. For each omics type, the user provides fold change and statistical values for each measurement. For example, if a user-provided data for a two-condition comparison transcriptomics dataset, these values would consist of $\log_2$(fold change) values between the experimental and control conditions, and the appropriate statistical values. As transcriptomics count data follows a negative binomial distribution, the appropriately adjusted p-values based on multiple hypothesis testing for this distribution should be used [20]. For other omics types, these multiple hypothesis testing methods will vary based on the data distribution and other conventions. During the data input step, the user will also specify a few experimental parameters for consideration during downstream analysis and visualization. For example, users with gene expression data can choose to "broadcast" their measured gene values to the protein nodes where the appropriate protein values are missing. Following these user inputs, the organism's network is curated and the data is overlaid upon the network. A database file for the dataset is output for future analysis (see Figure 2).



	000min_mct1_lfc	000min_mct1_bh	015min_mc
Urea	-3.238686904	9.85E-05	-2.2272
L-Valine	1.236272279	0.001305057	0.4466
Putrescine	-1.981283414	7.02E-06	-1.5845
Methylmalonic	0.886017803	0.005095074	0.1653
Glycine	2.002589931	0.000133546	2.5797
Malonic acid	0.149937263	0.729091415	-0.0341
Fumaric acid	2.588709849	3.83E-06	1.0036
L-Alanine	2.322251455	5.26E-09	0.7780
Threonic acid	-0.382539465	0.050386726	-0.0852
D-Malic acid	2.728950429	1.97E-07	2.6067
L-Isoleucine	1.063415577	0.001035193	0.3891
Citric acid	0.183446037	0.438568557	-1.7537
L-Lysine	-2.473820035	0.003217549	-5.3311
Xanthurenic aci	0.905991059	0.163877202	0.3146
L-Glutamic acid	-0.626056458	0.203936688	-2.5790
Orotic acid	-3.96860657	0.000206079	-2.9325
L-Tyrosine	-2.309293197	0.000371556	-2.9632
Ornithine	-6.562107439	0.004785547	-8.0784
Phosphoenolpy	-0.763107235	0.013593318	-1.0192
Thymine	-0.719164516	0.000306792	-0.3706
Adenine	-1.898401034	3.13E-05	-1.3664
L-Aspartic acid	0.850815348	0.014706926	-1.8264

Figure 2: Metaboverse software overview. An illustration outlining the data input, processing, and exploration steps of the Metaboverse package. The user provides the name of the organism of interest from a drop-down menu along with an output location. The user then has the option to provide transcriptomics, proteomics, and/or metabolomics datasets. These datasets can be single- or multi-condition or time-course experiments. Data is formatted as follows: row names are the measured entity names or IDs, the first column is $\log_2$(fold change) or other measurement values, and the second column is statistical measurements. For time-course and multi-condition datasets, this pattern is repeated for each subsequent sample. During this step, the user can also provide sample labels and other modifiers to customize the curation of the data on the reaction network. Metaboverse will then build the model. Once the model is complete,

the user will be able to visualize the patterns identified within reactions, explore pathway-specific or general perturbation networks, and perform general pathway and nearest reaction neighborhood exploration of the data.

We chose to format network visualizations in a reaction-centric format, which means that a sub-network unit is comprised of a reaction node, linked with its reaction input (substrate) and reaction output (product) nodes, along with any modifier (catalyst or inhibitor) nodes. If modifier nodes are a protein complex, the component proteins of that complex are linked to the complex node. Any genes available in the database are linked to their related proteins. Each node and link type are color-coded to help the user quickly distinguish the details of a reaction sub-network structure. Reactions with matching input or output sides are naturally linked when the network is constructed, allowing a multi-reaction pathway to emerge from this network structure. This formatting differs from other formatting styles with which the user may be familiar, such as KEGG pathway maps, for various reasons. While classical pathway definitions are used to group sets of reactions during visualization, analytical methods such as the pattern search in Metaboverse explore the global reaction network and display the pathways more accurately within their network context. This approach aids in making pattern analysis within the reaction network less subject to biases arising from constraining analysis to only familiar pathways and enables the exploration of patterns that may not neatly fit into canonical pathway representations. As this dynamic visualization approach differs from more classical imagery, each pathway within the visualization module contains a button linking to a more traditional representation of the pathway for reference, which is sourced from Reactome [17].

Rapid identification of interesting regulatory patterns in the reaction network

Following network curation, the user can visualize the available patterns identified across all reactions in the network. In a computer science context, a motif is a recurring pattern within a network structure or the organization of network entities and their relationships. However, with omics data, we are more interested in identifying patterns involving expression or abundance of genes, proteins, and metabolites. Previous work by Checkik, et al. introduced the concept of “activity motifs.” Instead of identifying motifs based on sub-network structure, they identified motifs using the expression characteristics of nodes in transcription factor binding signaling networks [22]. We adapted this methodology to identify and interactively display interesting regulatory hotspots across all reactions in the network. For example, a reaction’s inputs may exhibit increased levels and its outputs decreased levels, suggesting a possible regulatory event occurring at the reaction (see Supplementary Figure 6 for examples of reaction patterns and Supplementary Figure 7 for an example of the visual output of this module in Metaboverse).

Magnitude change and statistical values associated with each measurement for a measured entity are provided by the user. Once the data is layered upon the network, Metaboverse will consider all the reactions in the network and compare the expression or abundance patterns to those from a pre-defined library of regulatory patterns (please refer to the documentation for more details on available patterns and details on how they each operate [23]). Metaboverse then returns a graphical stamp view of identified patterns and sorts these reactions based on the significance values associated with those components that determined the regulatory feature. We use a three-tiered sorting strategy when sorting by the associated statistical values. The highest prioritized reaction patterns are those where the relevant components that determined the presence of a regulatory pattern (inputs vs. outputs) are statistically significant. Of these reactions from the first tier, reactions are sorted by lowest to highest cumulative p-value or other relevant statistical value. In the second sorting tier, reactions with at least one side of the reaction with a statistically significant reaction component are sorted by statistical strength. Finally, all other reactions are sorted by the cumulative statistical value for the reaction components. The reaction’s stamp shading indicates the sorting tier of each reaction, where green symbolizes that the components on both sides of the reaction passed significance thresholds, yellow symbolizes components on one side of the reaction passed significance thresholds,

and gray symbolizes neither side of the reaction contained significant components. Reactions can also be ordered based on the magnitude of change.

During dynamic pathway visualization, reaction pattern nodes are visually distinct to quickly draw the user's attention to these interesting patterns. For a given regulatory pattern, the user can then explore each pathway where the reaction was found. Pattern analysis of the all reactions within the network will allow the user to rapidly identify interesting features within the data, particularly patterns between canonical pathway representations or within pathways that may seem tangential to their research. In the data vignettes below, we demonstrate this utility further.

Handling data sparsity within the global biological reaction network

Missing data points, particularly in metabolomics experiments, can be frequent and make the analysis of pathways and identification of regulatory patterns in the network challenging. Thousands of metabolites are known to participate in human metabolism, but the current state of the technology often results in the quantification of a limited number of these metabolites. This limitation can lead to gaps in the measured metabolic network and can confound pattern recognition across reactions. Therefore, we developed a reaction compression algorithm (described 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 series. Similar concepts have been used in metabolomic analysis before to identify amino acid-related metabolites [24]. However, we introduce the first computational implementation to our knowledge and adopt a slightly more conservative collapsing scheme. These bridged reactions, or collapsed reactions, are visually distinct to show the user where reactions were collapsed and understand what that collapsed reaction summarizes. We found that these collapsed reaction representations aided in identifying regulatory mechanisms across multiple reactions that would have been missed by pattern searching across traditional, uncollapsed representations of human metabolic reactions (see Supplementary Figure 8 and accompanying text for a brief discussion of the benefits of this feature).

Dynamic visualization of organism-specific reaction pathways

Following the curation of the global network as described above, the user can manually search individual 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 visualized, along with their core input and output 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. For nodes representing protein complexes, simulated values are calculated by taking the average of all measured component entities in that complex. If gene values are available, but protein values are not, the user has the option to "broadcast" the gene values to the protein nodes. Assuming all statistical values are between 0 and 1, simulated statistics are calculated by taking the maximum statistical value of all of the measured component entities in that complex. In cases where a gene value is known, but its protein value is unmeasured, the protein value is optionally inferred using aggregated gene component values. Relevant pathway and analytical metadata are also displayed. All reaction pattern types are displayed using default thresholds if found within the selected pathway. Additional aids for visualization are also available. These aids include the ability to remove nodes from the visualization that contain a high number of relationships to other network features so that these nodes, which act as hubs in the network, do not lead to cluttered representations. These hub nodes are often ubiquitous features such as water and protons, which may be of limited interest to the user during data visualization [25]. Compartment domains, typically defined as organelles, are also visualized to show the cellular locations of reactions and their

components (see Supplementary Figure 9 for an example of the visual output of this module in Metaboverse).

Visualization of downstream effects of network perturbations using reaction neighborhoods

The user may be interested in a particular metabolite or protein and the downstream effects their perturbation has on other reactions across the entire network. By double-clicking a node of interest or selecting the entity name from the drop-down menu, the user can explore all upstream and downstream effects across reactions connected to that entity from all pathways. The user can also define how many neighborhoods to display such that two or more reaction steps downstream or upstream of the selected entity are visualized [26,27]. This functionality moves the analysis past traditional, pathway-centric approaches and contextualizes the far-reaching effects that the disruption of metabolism can have across classical pathways. However, as metabolism is a network with high connectivity [25], expanding past one reaction neighborhood without defining a hub threshold can detrimentally effect performance. In these instances, Metaboverse will display a warning and suggest thresholds to use for the number of neighbors and hubs.

Exploring perturbation networks within the data

At times, a user may be interested in exploring all of the perturbed reactions within a network. Metaboverse will display the total perturbation network for a given super-pathway selected by the user. For example, a user could select the *fatty acid metabolism* super-pathway, then explore all of the perturbed reactions within this context for interesting features. During perturbation network display, all reactions are included and plotted if at least one component of that reaction passes the indicated magnitude or statistical threshold. Naturally, sequentially perturbed reactions will be stitched together in the resulting perturbation network (see Supplementary Figure 10 for an example of the visual output of this module in Metaboverse).

Data vignettes

To demonstrate the utility of Metaboverse, we analyzed both public and our previously unpublished metabolically-relevant datasets. From the vignettes provided below, we show that Metaboverse identifies points of interest previously described or expected and rapidly identifies unexpected and systematic regulatory patterns in a reaction network context. For an example of the reaction collapse algorithm's utility, we refer the reader to the associated supplement (Supplementary Figure 8). Data from Vignette 1 is available in distributed versions of Metaboverse (as `test_data.zip`) to act as a test dataset for users to familiarize themselves with the input data format. Input and processed data from these vignettes can be found at [28].

Vignette 1: Acute adaptations of mitochondrial fatty acid synthesis-impaired yeast during respiratory growth

Mitochondrial fatty acid synthesis (mtFAS), an evolutionarily conserved pathway, has long been recognized 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 coordinates its fatty acid synthesis with the regulation of iron-sulfur (Fe-S) cluster biogenesis and assembly of oxidative phosphorylation complexes [29,30,31]. Additionally, this pathway has received increased physiological focus with the discovery of patients with mutations in key mtFAS enzymes [32].

The *Saccharomyces cerevisiae* Mct1 protein is an acyltransferase responsible for the transfer of a malonyl group from malonyl-CoA to the mitochondrial acyl carrier protein (ACP). By removing this gene, the activity of the mtFAS pathway is abolished. 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 an *mct1*Δ mutant *S. cerevisiae* strain. We previously performed steady-state comparative proteomics in this *mct1*Δ model 12 hours after shifting to a non-fermentable carbon source [30]. To complement these previous studies, we performed RNA-sequencing of transcript abundances at 0, 3, and 12 hours after the shift from a fermentable, glucose carbon source to a non-fermentable, raffinose carbon source and steady-state metabolomics at 0, 15, 30, 60, and 180 minutes after the shift in carbon source. By transitioning between these two growth states, the yeast move from a glucose-repressed state to a glucose-derepressed state accompanied by an increase in respiratory metabolism. By layering these data upon the *S. cerevisiae* metabolic network in Metaboverse, we observed interesting respiratory signatures as expected based on previous work [30,33]. For example, Metaboverse identified a strong pattern in the electron transfer from ubiquinol to cytochrome C via complex III of the electron transport chain (ETC) (Figure 3 A). At the protein level, the cytochrome C isoforms, CYC1 and CYC7, are both significantly reduced in concentration compared to wild-type cells (log_2(fold change): -1.57 & -0.88; adjusted p-value: 1.21E-03 & 5.84E-03; Cohen's *d*: -7.99 & -3.94; respectively). This reduction in cytochrome C concentration is paired with a marked reduction in the concentration of the protein components of Complex III. These components catalyze the transfer of electrons from ubiquinol to cytochrome C (average log_2(fold change) of all measured protein components: -2.03, where 9/11 component proteins were measured; 8/11 passed statistical significance thresholds of adjusted p-value <= 0.05. Range of significant log_2(fold change) values: -0.34 to -3.21).

The second expected pattern of interest identified by Metaboverse was the marked reduction of TCA cycle-related enzymes. Reaction pattern analysis within the TCA cycle identified several putative regulatory hotspots between metabolites and metabolite-protein interactions throughout the TCA cycle (Figure 3 B). However, visualizing these identified reactions and data across the time-course revealed interesting patterns. For this discussion, we will provide metabolite measurements at the 15-minute time point because the statistical strength for the discussed metabolites was generally strongest. For example, citrate levels decrease across the metabolomics time-course (at 15 min; log_2(fold change): -1.75, adjusted p-value: 6.61E-03; Cohen's *d*: -1.86), which is coincident with reduced steady-state levels of Ctp1 (log_2(fold change): -0.64, adjusted p-value: 6.06E-03; Cohen's *d*: -4.73), a protein that catalyzes the transfer of citrate from the mitochondrial matrix to the cytosol [34]. Citrate is a key metabolite and the first step in the TCA cycle. We hypothesize that due to central carbon metabolite reductions, Ctp1 may be down-regulated in response to the decrease in citrate to maintain citrate concentrations within the mitochondrial matrix, where it is perhaps most physiologically important in respiring cells.

Another point of interest is the up-regulation of Dic1 (log_2(fold change): 2.15, adjusted p-value: 5.95E-06; Cohen's *d*: 28.97), which catalyzes the exchange of dicarboxylates like malate (at 15 min; log_2(fold change): 2.61, adjusted p-value: 3.23E-06; Cohen's *d*: 5.12) between the mitochondrial matrix and cytosol. Interestingly, Dic1 is essential for growth on non-fermentable carbon source media [35]. When yeast, especially those with deficits in TCA cycle flux due to loss of Mct1, are switched to a non-fermentable carbon, as was done in this study, they might adapt by up-regulating Dic1, which we observed in this data.

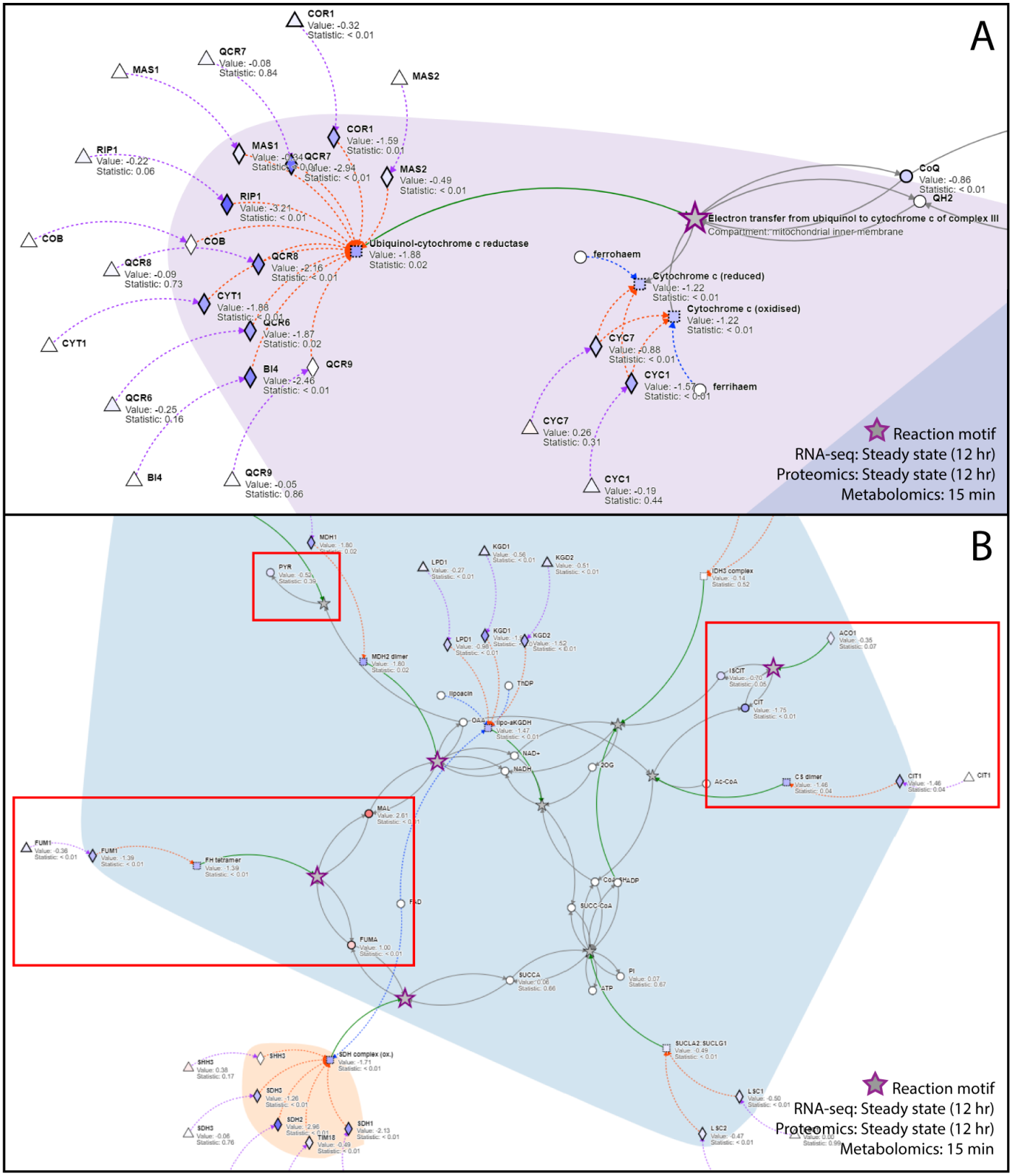


Figure 3: Metaboverse identifies several reaction patterns of interest in *mct1*Δ cells. (A) Steady-state proteomics (12 hours) overlaid on the reaction, "Electron transfer from ubiquinol to cytochrome c of complex III." (B) Steady-state proteomics (12 hours) and metabolomics at 15 minutes overlaid on TCA-related reactions. Time stamps for each data type are displayed in the lower-right hand corner of each subplot. Measured values are shown as node shading, where an increasingly blue shade indicates down-regulation, and an increasingly red shade indicates up-regulation. Measured $\log_2$(fold change) and statistical values for each entity are displayed below the node name. A gray node indicates a reaction. A bold gray node with a purple border indicates a potential regulatory pattern at this reaction for the given data type time points. Circles indicate metabolites, squares indicate complexes, diamonds indicate proteins, and triangles indicate gene components. Gray edges are core relationships between reaction inputs and outputs. Green edges indicate a catalyst. Dashed blue edges point from a metabolite component to the complex in which it is involved. Dashed orange edges point from a protein component to the complex in which it is involved. Dashed purple edges point

from a gene component to its protein product. Protein complexes with dashed borders indicate that the values displayed on that node were inferred from the constituent protein, metabolite, and gene measurements. The background shading demonstrates Metaboverse's ability to show cellular compartmentalization, although users may opt to toggle compartment shading off at any time.

We also noticed that although the TCA cycle enzymes are all reduced in concentration, several related metabolites are up-regulated across multiple time points in the dataset. It is possible, for example, that the increased fumarate levels (at 15 min; $\log_2$(fold change): 1.00, adjusted p-value: < 0.01; Cohen's *d*: 3.89) are related to the reduction in fumarate hydratase (FH tetramer, FUM1; $\log_2$(fold change): -1.39, adjusted p-value: < 0.01; Cohen's *d*: -4.91), as has been demonstrated previously and shown to associate with hereditary leiomyomatosis and renal cell cancer in humans [36]. These patterns would be an interesting point of further study to understand the adaptations these cells make to disruptions to the TCA cycle and the downstream consequences of these disruptions and adaptations.

By analyzing this multi-omics dataset using reaction pattern analysis and other interactive visualizations (or tools) available within Metaboverse, interesting questions arise. We see several reaction patterns that are expected based on prior knowledge of this biological model and other puzzling behaviors worthy of further follow-up. These identified patterns demonstrate the potential Metaboverse has to act as a valuable hypothesis-generation tool, particularly with multi-omics and time-course datasets, and how this platform allows for better visualization and analysis of a user's dataset in the context of metabolic reaction networks.

Vignette 2: Metabolic signatures in malignant human lung adenocarcinomas compared to nonmalignant lung tissue

We next turned to published human lung adenocarcinoma steady-state metabolomics data [37] to assess the utility of Metaboverse when analyzing data. Lung cancer remains a leading cause of death worldwide, and an improved understanding of these tumors' metabolism is essential in understanding how to treat this disease more effectively.

Consistent with the original study [37] and our recent re-analysis of the data [25], nucleotide metabolism was broadly up-regulated in adenocarcinomas based on the reaction pattern and perturbation network analyses. In particular, the highest-scoring reaction pattern identified by Metaboverse involved the transformation of dc-adenosyl methionine and spermidine to form spermine and 5'-methylthioadenosine. This reaction is a step in polyamine metabolism, but also connects with nucleotide metabolism (Figure 4 A). We also noticed similar perturbations in xanthine ($\log_2$(fold change): 1.35; adjusted p-value: 1.52E-05; Cohen's *d*: 0.86) (Figure 4 B) and other metabolites, which are highlighted to the user quickly in the reaction pattern and perturbation network analysis modules. However, as we previously emphasized [25], by approaching these perturbations in a reaction-centric approach, we can identify regulatory behavior that further contextualizes the data. For example, using a more manual approach, we previously highlighted the up-regulation of glyceric acid ($\log_2$(fold change): 0.50; adjusted p-value: 1.91E-04; Cohen's *d*: 0.73) coupled with the proximal down-regulation of 3-Phosphoglyceric acid ($\log_2$(fold change): -1.09; adjusted p-value: 4.74E-04; Cohen's *d*: -0.68) which could indicate the decreased activity of glycerate kinase [25]. This connection was missed in the original study but was highlighted by Metaboverse's reaction pattern analysis module. This sub-pathway activity is consequential as it has connections to serine metabolism that contributes to generating an ideal environment for tumorigenesis (Figure 4 C). Interestingly, of the measured and significant metabolites involved in the TCA cycle, we see a moderate up-regulation of malate ($\log_2$(fold change): 0.39, adjusted p-value: 3.85E-03; Cohen's *d*: 0.56) and down-regulation of citrate ($\log_2$(fold change): -0.64, adjusted p-value: 1.20E-03; Cohen's *d*: -0.63). This pattern is consistent with the hypothesis that the TCA cycle is starved of pyruvate-derived

carbons and is compensating by shunting carbon into the TCA cycle via glutaminolysis. This compensatory pathway, a feature in many cancers, consists of glutamine transport into mitochondria and its conversion to alpha-ketoglutarate. Alpha-ketoglutarate then enters the second half of the TCA cycle to generate ATP, reducing equivalents, and biosynthetic intermediates [38,39]. Pyruvate is likely shunted towards the production of lactate and other biosynthetic products (no measurements for lactate were available in the dataset). This pattern might explain the lower citrate levels, while the second half of the TCA cycle could be fed through glutamate ($\log_2$(fold change): 0.47, adjusted p-value: 1.99E-06; Cohen's *d*: 0.98), providing the increased malate concentration (Figure 4 D) [40].

Metaboverse simplifies the analysis process of metabolic data and allows the user more flexibility via an interactive platform that augments the identification of both canonical and novel regulatory patterns within the data. We were able to identify the regulatory pattern identified in Figure 4 C previously using more manual approaches, but Metaboverse offers an automated platform for the discovery of such events not identified in the original study. We anticipate that the use of Metaboverse will enable the user to extract new and exciting hypotheses that can drive their fields forward.

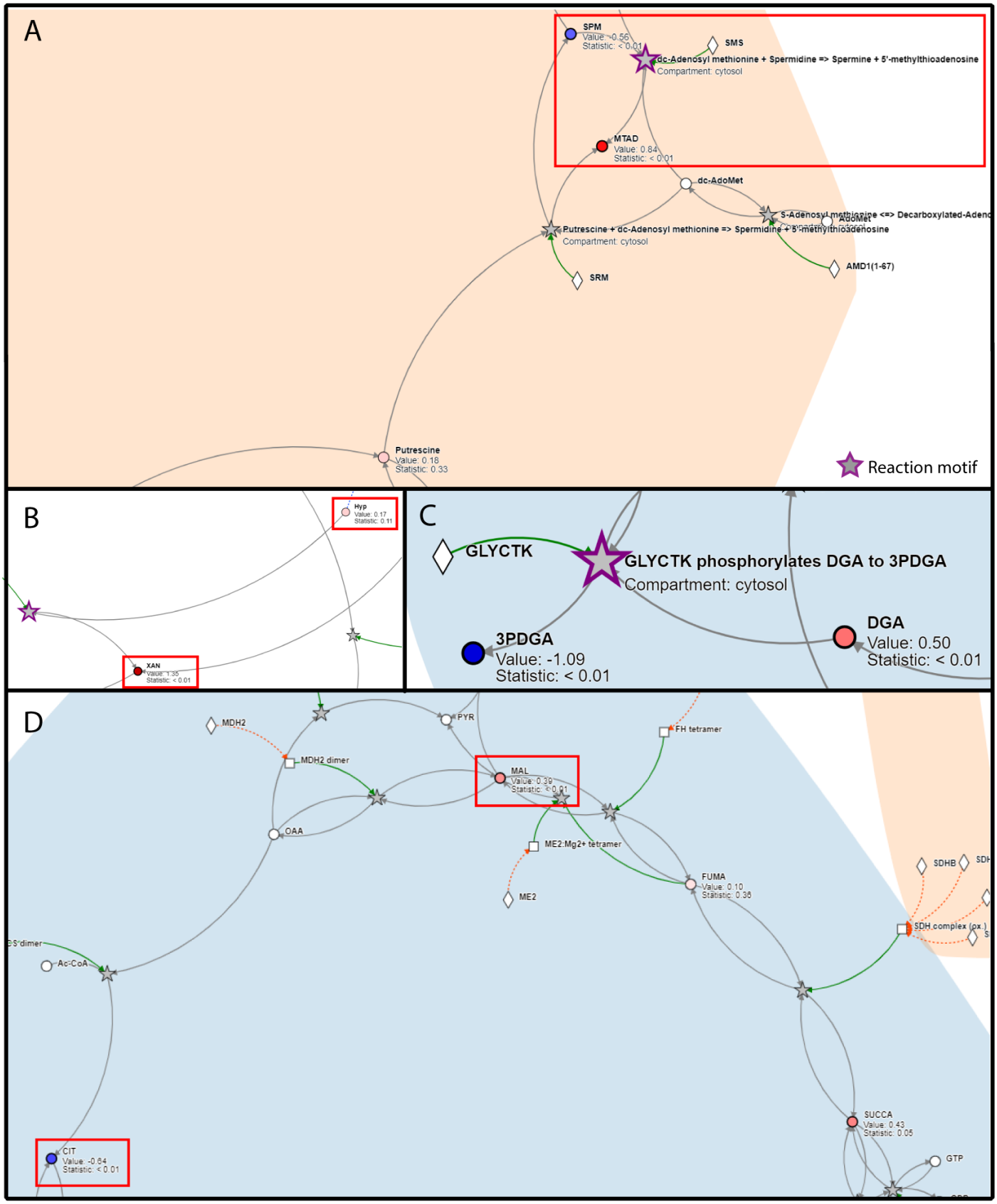


Figure 4: Metaboverse identifies nucleotide metabolism and other signatures in lung adenocarcinoma metabolomics data. (A) Up-regulation of polyamine and nucleotide metabolites. (B) Identification of xanthine regulation by both the pattern recognition and perturbation analysis modules. (C) Regulatory activity between glyceric acid and 3-Phosphoglyceric acid identified by Metaboverse's pattern recognition analysis module. (D) Disruptions of TCA metabolism support canonical disruptions during adenocarcinoma development. Metabolomics values are shown as node shading, where an increasingly blue shade indicates down-regulation and an increasingly red shade indicates up-regulation. Measured $\log_2$(fold change) and statistical values for each entity are displayed below the node name. A gray node indicates a reaction. A bold gray node with a purple border indicates a motif at this reaction. Circles indicate metabolites, squares indicate complexes, and diamonds indicate proteins. Gray edges indicate core relationships

between reaction inputs and outputs. Green edges indicate a catalyst, and red edges indicate inhibitors. Dashed blue edges point from a metabolite component to the protein complex in which it is involved. Dashed orange edges point from a protein component to the protein complex in which it is involved. Protein complexes with dashed borders indicate that the values displayed on that node were inferred from the constituent protein and metabolite measurements. Hub limit was set at 30 during generation of the network visualization as shown in sub-panel (D).

Discussion

In this manuscript, we introduce a new software tool for the analysis and exploration of user data layered on the metabolic reaction network. To improve on software with similar visualization capabilities, we introduced several new analytical tools and methods to aid the user in the automated identification and discovery of regulatory patterns within their data. These tools and methods include the automated ability to identify regulatory events across reactions, such as a reaction where an input has a high measured abundance and an output has a low measured abundance. Metaboverse also provides dynamic and interactive visualization capabilities to search for patterns and features within the user data manually within classical pathway representations. If a user is interested in how a reaction pattern is propagating across the entire reaction network and not just a single pathway, they can explore an entity's nearest reaction neighborhood. The user can also explore the total extent of perturbations across the network and begin to explore hypotheses around the role of connected or disconnected reactions within a particular biological model.

To address the challenge of data sparsity, particularly regarding metabolomics data and the metabolic reaction network, we introduce a reaction collapsing feature. This feature summarizes a series of connected reactions in which values may be missing between the reactions, but where the terminal ends of the reaction path have measured values. Importantly, the ability to collapse reactions augments the capabilities available within Metaboverse, especially in identifying disease-relevant reaction patterns that may be of interest to the user but are otherwise obfuscated.

We demonstrated the utility of Metaboverse in exploring single- and multi-omic datasets. We analyzed previously published studies and generated new datasets to highlight the time-course and multi-omic capabilities of this tool. We demonstrated that Metaboverse was able to identify regulatory patterns that were expected in the models based on the current literature and identify intriguing patterns that led us to form new hypotheses. We expect these features to be a powerful tool in researchers' toolkits as they analyze their data and plan their next experimental steps. We plan to continue to implement additional features that will expand the applicability of Metaboverse, and new features requested by users of Metaboverse have already been added and will continue to be added.

Metaboverse aims to enhance the computational toolkit for data analysis and hypothesis generation in metabolic and other experiments. However, numerous challenges remain, which we intend to address in future versions of this software. For example, the reaction collapsing features of Metaboverse aid in identifying patterns across several reactions where data may be missing, but various biological and technical edge cases will need to be accounted for in future implementations of this feature. As more technical limitations in metabolomics data generation and analysis are overcome, we hope that more complete snapshots of metabolism will be visible. Additionally, we currently take a more straightforward and somewhat rudimentary approach to statistical significance integration in the reaction pattern searches. However, more holistic platforms for multi-omics integration are needed and remain a significant challenge within the broader field of multi-omics data analysis.

We hope that Metaboverse will bring a new perspective to users' data. We envision Metaboverse will become a staple in the metabolic research toolkit that will help researchers critically and holistically

consider their data in the context of biological network interactions and draw the connections needed to extract new and exciting hypotheses that might be challenging without these features.

Methods

A tutorial for how to use Metaboverse can be found at https://metaboverse.readthedocs.io/en/latest/content/general-usage.html. For the most up-to-date instructions and details of available features, we refer the user to this documentation.

1. Network curation

Biological networks are curated using the current version of the Reactome database. In particular, the pathway records and Ensembl- and UniProt-Reactome mapping tables are integrated into the network database for Metaboverse. Additionally, the ChEBI and The Human Metabolome databases are also referenced for metabolite synonym mapping to accept more flexible metabolite input nomenclature from the user [18,19]. These data are used to generate a series of mapping dictionaries for entities to reactions and reactions to pathways for the curation of the total reaction network.

After the relevant information is parsed from each table or record, the global network is propagated using the NetworkX framework [41] 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 downstream visualization; however, user data for the given entity would be 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 visual network entanglement and cluttering and a decrease in computational performance [25].

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 table is 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 are available, but protein abundance values are missing, Metaboverse will take the average of the available gene expression values to broadcast to the protein node. For complexes, all available component values (metabolites, proteins, etc.) are averaged. 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. Statistical values are derived from the highest value of the components (assuming a scale of 1 denotes no statistical significance and 0 denotes high statistical significance).

3. Collapsing reactions with missing expression or abundance values in user data

After data mapping is complete, Metaboverse will generate a collapsed network representation for optional viewing during later visualization. Metaboverse enforces a limit of up to three reactions that

can be collapsed as data down a pathway should be inferred only so far. Additional parameters for the reaction-collapse are as follows:

1. If a reaction has at least one known or inferred value for inputs (substrates) and one known or inferred value for outputs (products), the reaction will be left as is. During the entire reaction collapse step, known catalysts can be included when assessing whether a reaction has measured output values (increased catalyst should lead to more output in most cases), and inhibitors can be included when assessing whether the reaction has measured input values (increased inhibitor should lead to an 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 inputs with the first reaction's outputs is determined whether it has a measured output. If the neighbor reaction does not contain a known output value, the reaction is left as is. If the neighboring reaction does contain a measured output, the first reaction's inputs and the neighboring reaction's 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 components identical to the reaction's 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.
4. All other reactions are maintained in the network.

For collapsed reactions, appropriate notes are included to describe the collapse. During visualization, these collapsed reactions are marked by black dashed edges and dashed node borders. A graphical representation of how this reaction collapse scheme is executed can be found in Figure 5.

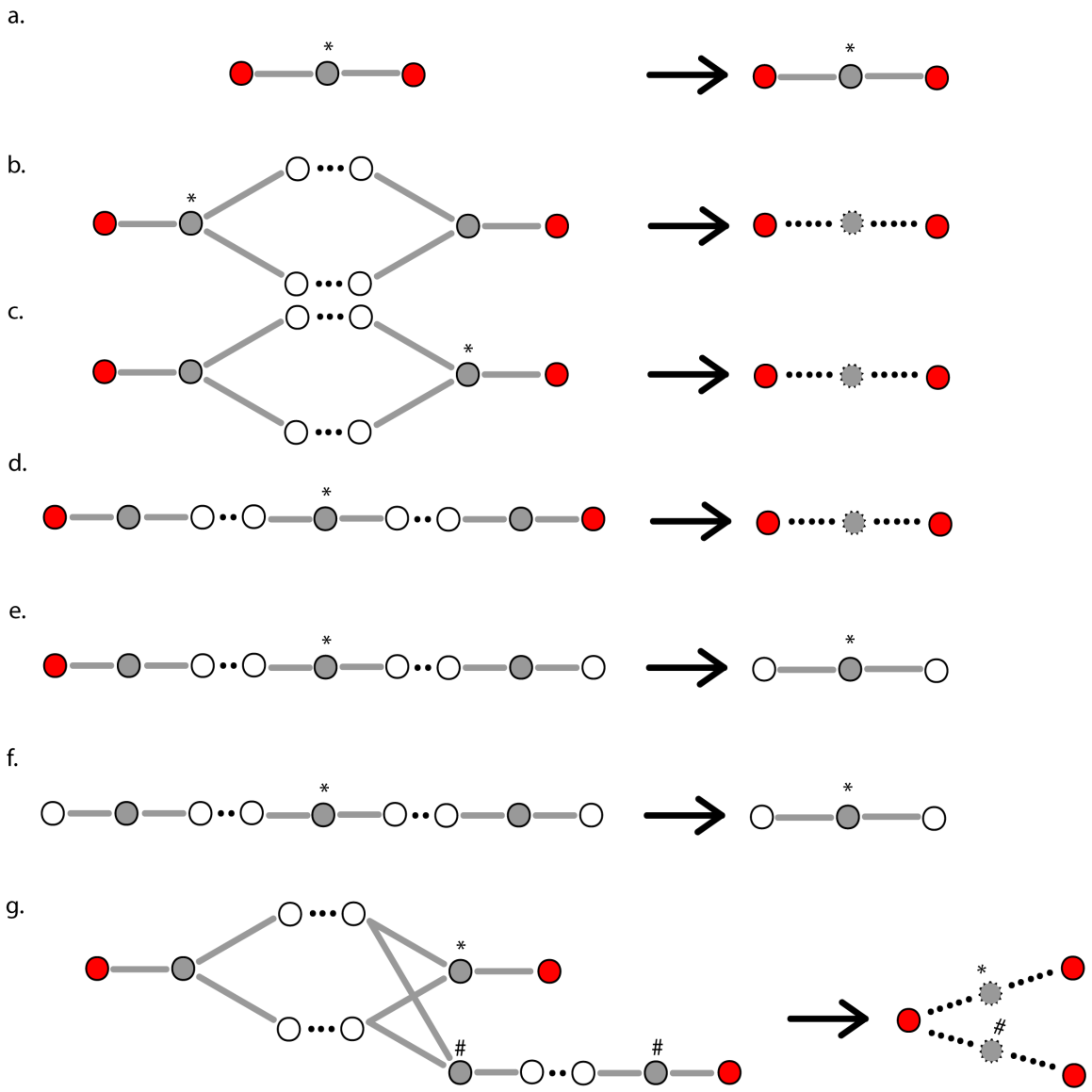


Figure 5: Reaction node collapse schematic. (a) For reactions for which at least one input and at least one output component contain a measured value from the user data, the reaction will be maintained as is. (b) When an input of a reaction is known, but no output has a known value, Metaboverse will 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) When 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) is performed as in (d), but if one neighbor does not contain a value and the other 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. Example (g) demonstrates how reactions with multiple paths are collapsed to a representative reaction. 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 gray node indicates a reaction. A gray 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 searches and sorting

Metaboverse provides a variety of different regulatory patterns for the user to explore. To identify a reaction-pattern is to compare some value that is computed from a reaction or a pathway with a user-specified threshold.

The identified reaction-patterns will be listed in a stamp view. Each stamp represents a reaction, with a glyph of the reaction, or the name of the pathway on it. In this stamp view, the identified patterns can be sorted according to three criteria: the number of pathways containing the reaction (not applicable for pathway pattern identification), the magnitude of the change of the computed value, and the statistical significance. When sorting by the number of pathways or the magnitude of the change, the identified reactions are arranged in order from the largest to the smallest. When sorting by the statistical significance, reactions with statistical significance on both the input side (substrates) and the output side (products) are listed first by the product of their maximum statistics, followed by the reactions with statistical significance on one of the two sides, and finally the reactions with no statistical significance on both sides. Within each tier, the reactions are sorted from lowest to highest p-values. For all values or statistics used in sorting, only those that determined the reaction-pattern are used.

When a reaction is selected from the stamp view, all the pathways containing the corresponding reactions will be listed below the stamp. Clicking on a pathway ID will draw the selected pathway in which the reaction-pattern was found, with all other reaction-patterns within this pathway also highlighted. For time-course and multi-condition datasets, the selected reaction-pattern's total behavior is displayed below these windows as line-plots showing the reaction components' behavior across all time-points or conditions.

5. Nearest neighborhood searches and prioritization

To visualize all connections to a given network component, a user can select an entity (a gene, protein, or metabolite) and visualize all reactions in which the component is involved. By doing so, the user can visualize other downstream effects the change of one entity might have across the total network, which consequently aids in bridging and identifying any reaction 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 software.

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

6. Perturbation networks

Perturbation networks are generated by searching each reaction in the total reaction network for any reaction where at least one component is significantly perturbed. The user 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, these networks could represent reactions where a component was perturbed to a significant degree on a $\log_2$ fold change scale, z-score scale, or other appropriate unit for that biological context.

Once a list of perturbed reactions is collected, the network is constructed, including each of these reactions and their components. Perturbed neighboring reactions that share components are thus

connected within the network, and perturbed reactions that are not next to other perturbed reactions are shown as disconnected sub-networks.

7. Network visualization and exploration

7.1 Dynamic network plotting

The user interacts with Metaboverse through an interactive app interface. Metaboverse comes packaged as a single executable app with all necessary dependencies included for running on Linux, macOS, and Windows.

Force-directed layouts of networks are constructed by taking a user-selected pathway or entity and querying the reactions that are components of the selected pathway or entity. All inputs, outputs, modifiers, and other components of these reactions, along with edges where both source and target are found in the sub-network as nodes, are included and displayed. Relevant metadata, such as user-provided data and reaction descriptions, can be accessed by the user in real-time.

Some performance optimization features are included by default to prevent computational overload. For example, nearest neighbor sub-networks with more than 1,500 nodes, or nodes with more than 500 edges, will not be plotted because the plotting of this information in real-time can be quite slow.

7.2 Visualizing pathways and super-pathways

To visualize a pathway, a user selects a pathway, and all component reactions and their substrates, products, modifiers, and metadata are queried from the total reaction database. Super-pathways help categorize these pathways and are defined as any pathway containing more than 200 nodes.

7.3 Visualizing time-course and multiple condition experiments

Time-course and multiple condition experiments are automatically detected from the user's input data. When users provide these data and specifies the appropriate experimental parameters on the variable input page, they will have the option to provide time point or condition labels. Provided data should be listed in the data table in the same order that the labels are provided. Within all visualization modules, the data for each time point or condition can then be displayed using a slider bar, which will allow the user to cycle between time points or conditions.

7.4 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 substrates, products, and modifiers for that given compartment.

7.5 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 differentiate them from measured values.

7.6 Additional features

Metaboverse will continue to undergo development, and new features will be added, but we will briefly highlight some additional features available at the time of publication. We encourage the user to check the documentation for more current updates and information regarding the use of Metaboverse [23].

7.6a Toggling genes

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

7.6b Toggling values

The user can switch between coloring nodes based on the value provided by toggling the appropriate button. Color bar information for the dataset is saved in the network metadata during curation and used to generate a color bar. Statistically significant nodes, as determined by the user's input threshold, are displayed with bold node borders.

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.

7.6d Toggling collapsed reactions

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

7.6e Viewing curated pathway image

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

7.6f Saving network visualizations

The user 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 visualize, or the limit to the number of edges a visualized node can have, can be modulated by the user using the appropriate input spaces. When visualizing a nearest neighbors network, Metaboverse will recursively fetch related reactions and their neighbors until the critical node display threshold is reached. This searching capability allows the user to visualize downstream effects of a change that may propagate across several reactions. The hub

threshold option prevents the plotting of nodes with more than the specified number of edges. This thresholding is handled during the visualization process by excluding any entity nodes that meet these criteria as the neighborhood is propagated. This thresholding 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 network visualization. This feature can also help plot more extensive neighborhoods, as often neighborhoods quickly link to high-degree nodes, such as water, and limit visualization clarity.

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 displaying this information, which is updated based on the user's node selections. Compartment and node legends are available for the user's reference by hovering over the appropriate icons.

8. Packaging

The Metaboverse app is packaged using Electron [42] . Back-end network curation and data processing are performed using Python [43] and the NetworkX library [41]. This back-end functionality is packaged as a single, operating system-specific executable using the PyInstaller library [44] and is available to the app's visual interface for data processing. Front-end visualization is performed using Javascript and relies on the D3 [45] and JQuery packages. Saving network representations to a PNG file is performed using the d3-save-svg [46] and string-pixel-width [47] packages (Table 1). Documentation for Metaboverse is available at https://metaboverse.readthedocs.io. Continuous integration services are performed by Travis CI [48] 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.come/Metaboverse/manuscript.

Table 1: Dependencies table.

Name	Reference
HTML	N/A
CSS	N/A
Javascript	N/A
Electron	[42]
JQuery	[49]
D3	[45]
string-pixel-width	[47]
d3-save-svg	[46]
Python	[43]
pandas	[50,51]
numpy	[52]
scipy	[53]
matplotlib	[54]

Name	Reference
NetworkX	[41]

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 [30].

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 the time of harvest, cultures were at OD_{600}=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 (v0.6.0) [55]. Batch and log files are available at [56]. Notably, reads were trimmed of adapters (AGATCGGAAGAGCACACGTCTGAACTCCAGTCA). 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 [20] by comparing *mct1*Δ 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 [30]. 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 *mct1*Δ (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 *mct1*Δ with the wild-type cells. P-values were generated before correction using a 2-tailed, homoscedastic Student's T-test.

Metabolomics sample preparation

Metabolomics data were generated by growing the appropriate yeast strains in synthetic complete media supplemented with 2% glucose until they reached saturation (n=6; except in one 3-hour wild-type sample, where n=5). Cells were then transferred to S-min media containing 2% raffinose and leucine and harvested after 0, 15, 30, 60, and 180 minutes (n=6/time-point/strain, except for the 3-hour wild-type samples, where n=5) at OD_{600}=0.6-0.8.

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 minutes. Samples were then incubated at -20 ˚C for 1 hour. 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 1 hour at 37 °C in a sand bath. 25 µL of this solution were added to auto sampler vials. 60 µL of N-methyl-N-trimethylsilyltrifluoracetamide (MSTFA with 1%TMCS, Thermo #TS48913) were 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 the 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 the analysis. The gas chromatograph had an initial temperature of 60 °C for 1 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.

Analysis of mass spectrometry data

Data were collected using MassHunter software (Agilent). Metabolites were identified, and their peak area was recorded using MassHunter Quant. These data were 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 and Fiehn libraries. Resulting data from all samples were normalized to the internal standard d4-succinate. P-values were derived using a homoscedastic, two-tailed Student's T-test and adjusted using the Benjamini-Hochberg correction procedure.

9.2 Human lung adenocarcinoma metabolomics

Data were accessed from Metabolomics Workbench [57] and processed as in our previous re-study of these data [25]. P-values were derived using a two-tailed, homoscedastic Student's T-test and adjusted using the Benjamini-Hochberg correction procedure.

10. Data availability

*mct1*Δ and accompanying wild-type transcriptomics time-course data are deposited at the GEO repository under identifier GSE151606 [58]. Metabolomics data are deposited at the Metabolomics Workbench repository under project identifier PR000961, study identifier ST001401 [59].

The curated networks for these data are available at [60]. 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 cereviseae* global reaction network as curated by Metaboverse from the Reactome database. reaction-patterns and perturbation network analyses were performed within the Metaboverse platform.

The Metaboverse source code is available at https://github.com/Metaboverse/Metaboverse. The latest version of the software can be found at https://github.com/Metaboverse/Metaboverse/releases/latest.

The source code and data for this manuscript and the subsequent analyses are available at [28].

Acknowledgements

We thank Thomas O'Connell (Indiana University School of Medicine) for valuable feedback and feature suggestions during beta-testing of the software. We thank Alex J. Bott (University of Utah), Ahmad A. Cluntun (University of Utah), Kevin G. Hicks (University of Utah), Jeffrey T. Morgan (University of Utah), and other members of the Rutter lab for their thoughtful insights and suggestions. We thank Brian Dalley and the University of Utah High-Throughput Genomics Core for help with RNA library preparation and sequencing. We also thank Christine Pickett (University of Utah) for providing editing services for this manuscript. The support and resources from the Center for High-Performance Computing at the University of Utah are gratefully acknowledged.

Funding

J.A.B. is supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award number F99CA253744 (https://www.cancer.gov/) and received additional 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/). S.M.N. received support from The United Mitochondrial Disease Foundation PF-15-046 (https://www.umdf.org/) and the American Cancer Society PF-18-106-01 (https://www.cancer.org/) postdoctoral fellowships, along with T32HL007576. J.E.C. is funded by S10OD016232, S10OD021505, and U54DK110858. This work was supported by funds from the NIDDK fellowship 1T32DK11096601 and the NCI fellowship 1F99CA253744 (to J.A.B.) (https://www.niddk.nih.gov/, https://www.cancer.gov/), the NSF grants DBI-1661375 and IIS-1513616 (to B.W., Y.Z.) (https://www.nsf.gov/), and the NIH grant R35GM131854 (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 NCRR Shared Instrumentation Grant 1S10OD016232-01, 1S10OD018210-01A1, and 1S10OD021505-01 (to J.E.C.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the

manuscript. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH).

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., Y.O., I.G.
- Data Curation: J.A.B., T.C.W.
- Resources: J.A.B., J.E.C., B.W., J.R.
- Funding Acquisition: J.A.B., J.E.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., J.E.C., B.W., J.R.
- Visualization: J.A.B., Y.Z., Y.O.

Supplementary Material

Examples of reaction patterns in Metaboverse

The following figure provides examples of reaction patterns that Metaboverse will search for in the network based on the user's data (Supplementary Figure 6).

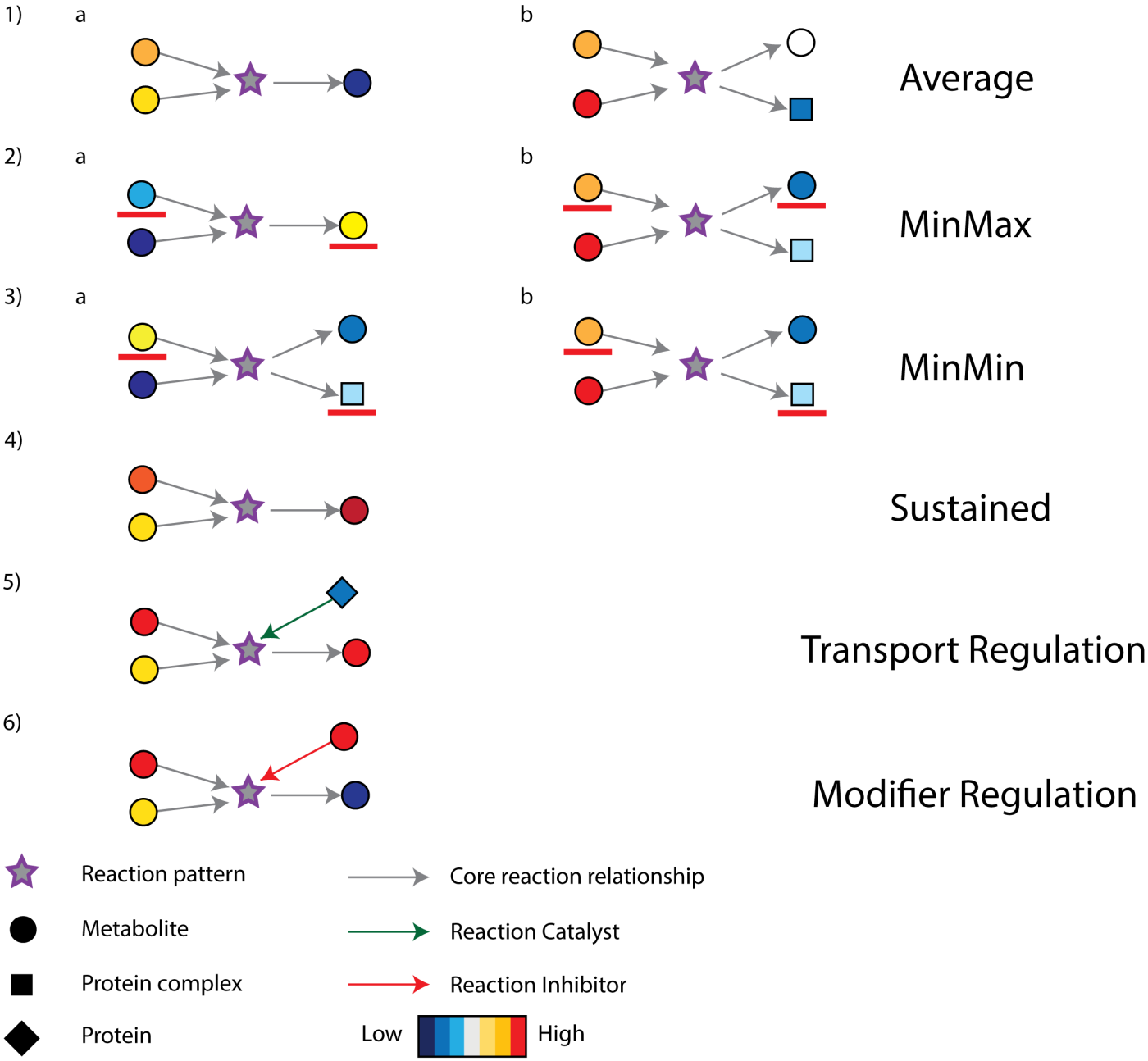


Figure 6: Reaction pattern examples. (1a-b) Examples where the measured values on both sides of a reaction are averaged. The difference between the two averages is then compared and, if greater than the specified threshold, identified as a reaction pattern. Where the metabolite product is unmeasured (a white node), that entity is ignored during averaging and comparing the two sides of the reaction. (2a-b) Examples where the minimum measured substrate value and the maximum measured product, or vice versa, are compared and identified as a reaction pattern if the difference passes the specified threshold. (3a-b) Examples where the minimum measured substrate and product values are compared and identified as a reaction pattern if the difference passes the given threshold. Alternatively, an option exists to compare the maximum values on both sides of the reaction. (4) An example where at least two different entities on both sides of the reaction are both upregulated or downregulated past the specified threshold. (5) An example where the same reaction entity on both sides is upregulated or downregulated past the specified threshold, with a modifier component also being upregulated or downregulated. (6) An example where the difference between at least one core reaction component and one reaction modifier is greater than the specified threshold. A star node indicates a reaction node. A purple border on a reaction node indicates a reaction pattern was identified. Circles

indicate metabolites, squares indicate protein complexes, and diamonds indicate proteins. Grey arrows indicate the core relationships of a reaction. Green arrows indicate a catalyst component of a reaction, and red arrows indicate an inhibitor component of a reaction. A representative colormap is provided, where increasingly red shades indicate increased abundance, while increasingly blue shades indicate decreased abundance. Entities underlined in red are those that are relevant to the particular reaction pattern.

Example for the Pattern Analysis module

The following figure provides an example of the `Pattern Analysis` module within Metaboverse, with relevant points of interest highlighted and explained in the figure legend (Supplementary Figure 7).

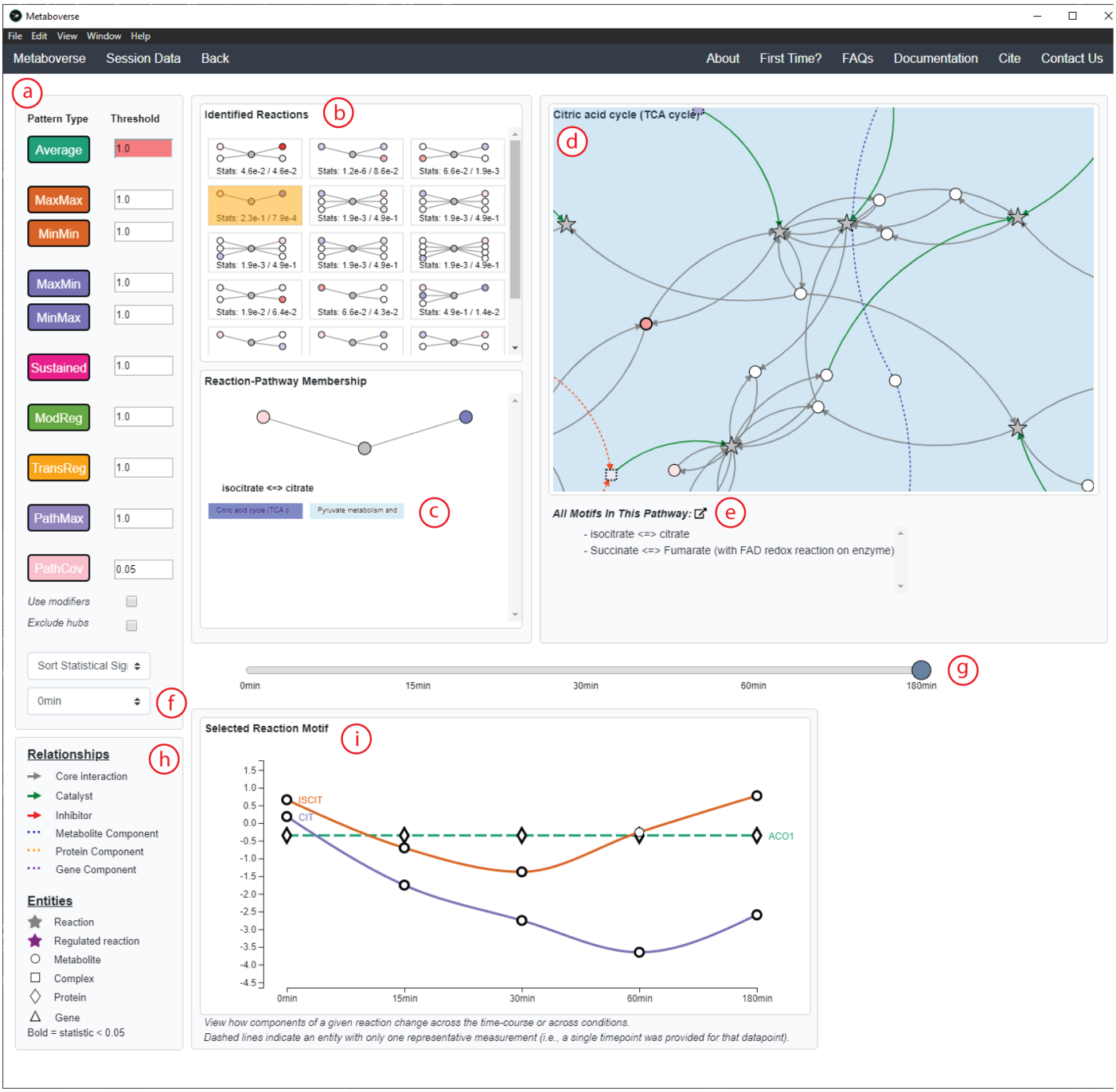


Figure 7: Pattern Analysis module. (a) The patterns currently available to search in Metaboverse are displayed in this panel. Users can adjust the thresholds for the given pattern using the appropriate fields. (b) For the selected pattern type, all available reaction-patterns within the organism's network will appear here. Reactions will be sorted by statistical values, as provided by the user, that were relevant to the identified pattern. Green stamps indicate both sides of the reaction contained significant values, yellow stamps indicate one side of the reaction contained significant values, and gray stamps indicate neither side of the reaction contained significant values. (c) This panel displays all of the pathways

in which the reaction-pattern can be found. The graphic provides a simplified view of the reaction's primary substrates and products, with the nodes shaded by associated value. A user can select one of these pathways for viewing. (d) The selected pathway is displayed here, with all reaction-patterns in the pathway highlighted with a bold purple border. (e) If a user would like to view the selected pathway within the `Explore` module, they can click the icon to open the pathway in a new window. All reaction-patterns within the displayed pathway are listed below. (f) For time-course or multi-condition datasets, the user can choose to display reaction-patterns for the selected time-point or condition that are not present in another time-point or condition. (g) For time-course or multi-condition datasets, a slider bar will appear with the time-points or conditions for the user to select. (h) This panel provides a shape legend for the different shape types used in Metaboverse. (i) For time-course or multi-condition datasets, a line plot will appear for the selected reaction. While panels above only display the behavior of that reaction-pattern at a single time-point or condition, the line plot will display that reaction's behavior across all of the available time-points or conditions.

Reaction collapsing reveals multi-reaction putative regulatory patterns

Yeast time-course metabolomics processed data were obtained as measured and discussed in this study in Vignette 1. Metaboverse reaction-pattern searches identified a pattern across the two-step transformation of glutathione and methylglyoxal into lactate at 60 minutes after the carbon source shift. Since the intermediate metabolite was not measured in the dataset, but the inputs to the first reaction and the outputs to the second reaction were measured, Metaboverse collapsed the two reactions into one, thus enabling pattern recognition across two reactions (Figure 8). Conventional methods might have missed such a reaction without summarizing the intermediate steps where the intermediate metabolite(s) are not measured. This transformation is particularly important in this model as lactate is closely tied to TCA metabolism and may provide explanatory mechanisms for the changes in TCA metabolism observed in this model.

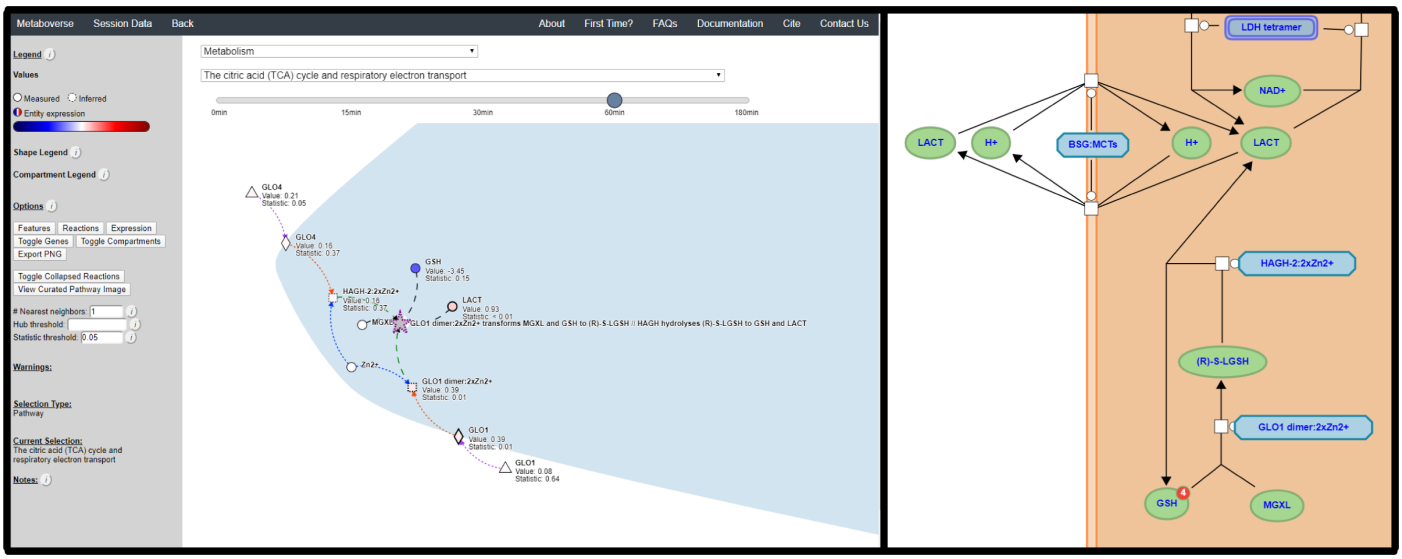


Figure 8: Reaction collapsing reveals putative regulatory patterns across TCA metabolism-proximal reactions in mtFAS-deficient yeast. (Left) Metabolomics values are shown as node shading, where an increasingly blue shade indicates down-regulation, and an increasingly red shade indicates up-regulation. Measured $\log_2$(fold change) and statistical values for each entity are displayed below the node name. A gray node indicates a reaction. A bold gray node with a purple border indicates a pattern was found at this reaction for the given data-type time points. A dashed purple border indicates a pattern was discovered across collapsed reactions. Circles indicate metabolites, squares indicate protein complexes, diamonds indicate proteins, and triangles indicate genes. Gray edges are core relationships between reaction inputs and outputs. Green edges indicate a catalyst. Dashed blue edges point from a metabolite component to the complex in which it is involved. Dashed orange edges point from a protein component to the complex in which it is involved. The background shading indicates different cellular compartments for each color. The displayed values are for the 60 minute time point. (Right) A screenshot from the Reactome (v73; https://reactome.org/PathwayBrowser/#/R-HSA-1428517&PATH=R-HSA-1430728) pathway explorer of this two-step reaction set.

Example for the General Exploration module

The following figure provides an example of the `Explore` module within Metaboverse, with relevant points of interest highlighted and explained in the figure legend (Supplementary Figure 9).

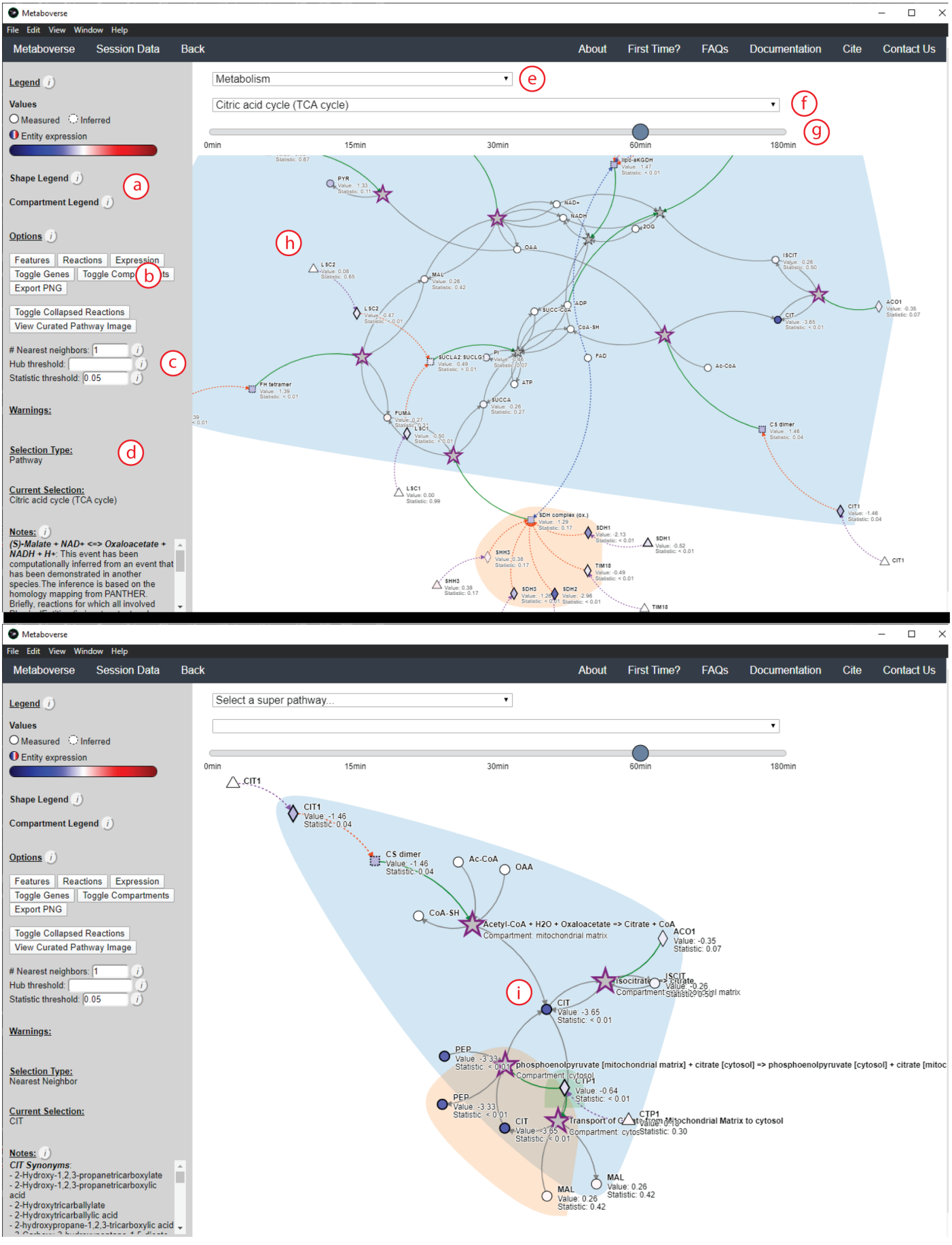


Figure 9: Explore module. (a) By hovering on the icon, users can view the shape legend. Below the shape legend is the compartment legend, where all currently displayed compartments in the selected pathway will be shown. (b) Users can toggle features on or off using the appropriate buttons. For example, users could choose to display all reaction names, not just when a reaction is selected. (c) Users can modulate graphing parameters, such as the number of reaction neighborhoods to graph when selecting a network component for nearest neighborhood analysis. (d) Notes about the selected reaction or synonyms for the selected reaction component appear here, as well as other information. (e) The drop-down menu for the selection of a super-pathway. (f) The drop-down menu for the selection of a specific pathway within the selected super-pathway. (g) For time-course and multi-condition datasets, a slider bar will appear that users can use to change value shading of nodes and highlighted reaction-patterns. (h) Viewing area for the selected pathway. (i) When users double-click on a reaction component, a nearest neighborhood graph will be displayed. In this instance, citrate was selected and all reactions citrate is involved in across all pathways are plotted.

Example for the Perturbation Networks module

The following figure provides an example of the `Perturbation Networks` module within Metaboverse, with relevant points of interest highlighted and explained in the figure legend (Supplementary Figure 10).

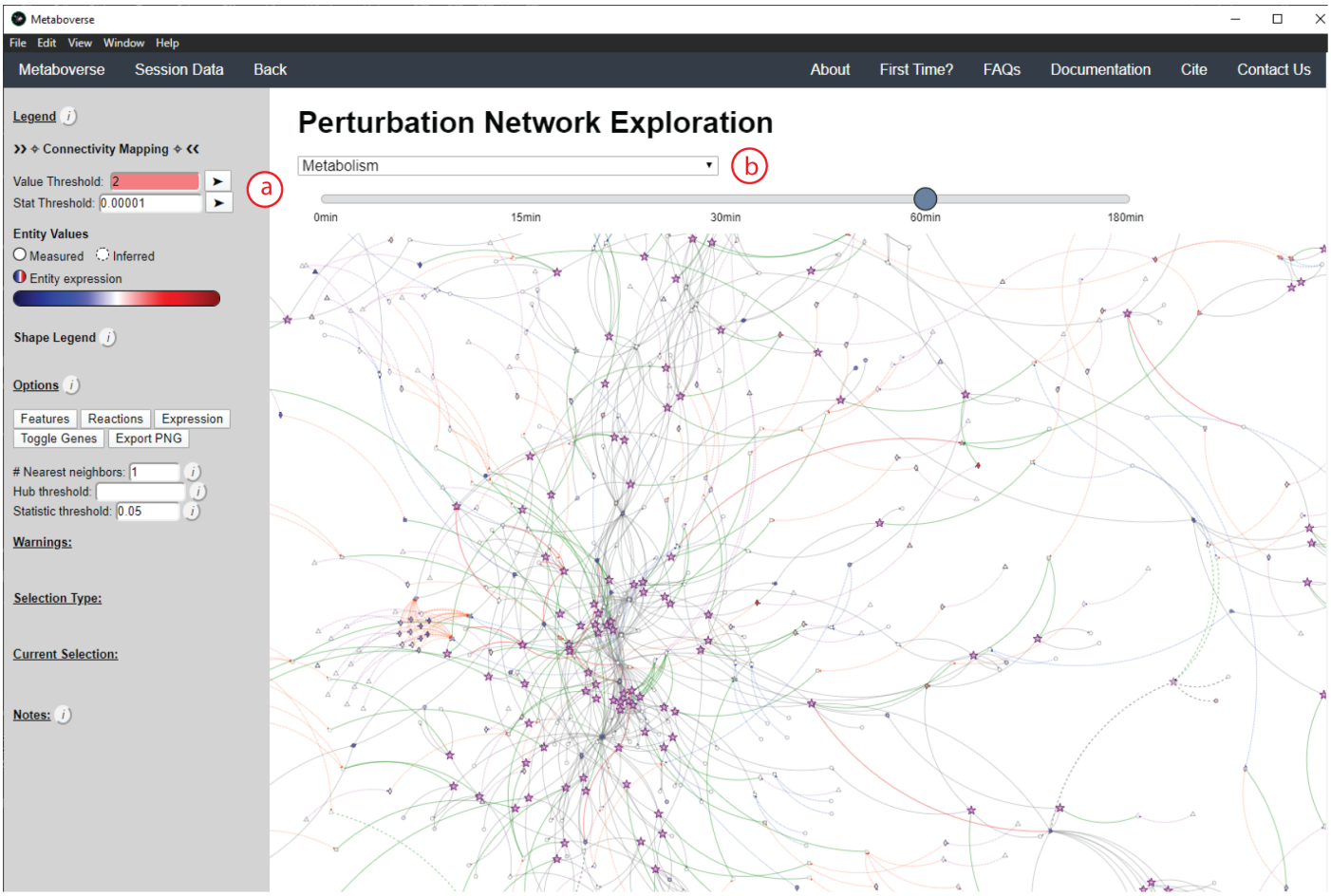


Figure 10: Perturbation Networks module. (a) Users can select what constitutes a perturbed reaction by modulating the appropriate thresholds. Only one type of perturbation (by magnitude or by statistical value) will be used. (b) Users select a super-pathway, for which all perturbed reactions belonging to that super-pathway are selected and displayed. Reactions that are neighbors and perturbed will be shown stitched together within the network, even if these reactions are connected across different pathways.

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 · PMID: 31003499 · PMCID: PMC6523452

2. **Interrogating metabolism as an electron flow system**
Christian Zerfaß, Munehiro Asally, Orkun S. Soyer
Current Opinion in Systems Biology (2019-02) https://doi.org/ggzxv9
DOI: 10.1016/j.coisb.2018.10.001 · PMID: 31008413 · PMCID: PMC6472609

3. **Flux-dependent graphs for metabolic networks**
Mariano Beguerisse-Díaz, Gabriel Bosque, Diego Oyarzún, Jesús Picó, Mauricio Barahona
npj Systems Biology and Applications (2018-08-14) https://doi.org/ggzxwb
DOI: 10.1038/s41540-018-0067-y · PMID: 30131869 · PMCID: PMC6092364

4. **Integrated omics: tools, advances and future approaches**
Biswapriya B Misra, Carl Langefeld, Michael Olivier, Laura A Cox
Journal of Molecular Endocrinology (2019-01) https://doi.org/gfx8hg
DOI: 10.1530/jme-18-0055 · PMID: 30006342

5. **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 · PMID: 19429898 · PMCID: PMC2703878

6. **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 · PMID: 29762782 · PMCID: PMC6030889

7. **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 · PMID: 14597658 · PMCID: PMC403769

8. **Sparse network modeling and Metscape-based visualization methods for the analysis of large-scale metabolomics data**
Sumanta Basu, William Duren, Charles R. Evans, Charles F. Burant, George Michailidis, Alla Karnovsky
Bioinformatics (2017-01-30) https://doi.org/f97km2
DOI: 10.1093/bioinformatics/btx012 · PMID: 28137712 · PMCID: PMC5860222

9. **Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data**
Alla Karnovsky, Terry Weymouth, Tim Hull, V. Glenn Tarcea, Giovanni Scardoni, Carlo Laudanna,

Maureen A. Sartor, Kathleen A. Stringer, H. V. Jagadish, Charles Burant, … Gilbert S. Omenn
Bioinformatics (2011-11-30) https://doi.org/cggpz2
DOI: 10.1093/bioinformatics/btr661 · PMID: 22135418 · PMCID: PMC3268237

10. **Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks**
Jing Gao, V. Glenn Tarcea, Alla Karnovsky, Barbara R. Mirel, Terry E. Weymouth, Christopher W. Beecher, James D. Cavalcoli, Brian D. Athey, Gilbert S. Omenn, Charles F. Burant, H. V. Jagadish
Bioinformatics (2010-02-07) https://doi.org/d2qxr9
DOI: 10.1093/bioinformatics/btq048 · PMID: 20139469 · PMCID: PMC2844990

11. **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 · PMID: 29718355 · PMCID: PMC6030842

12. **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 · PMID: 20444866 · PMCID: PMC2896158

13. **MetExploreViz: web component for interactive metabolic network visualization**
Maxime Chazalviel, Clément Frainay, Nathalie Poupin, Florence Vinson, Benjamin Merlet, Yoann Gloaguen, Ludovic Cottret, Fabien Jourdan
Bioinformatics (2017-09-15) https://doi.org/ghbd7x
DOI: 10.1093/bioinformatics/btx588 · PMID: 28968733 · PMCID: PMC5860210

14. **Reactome: a knowledgebase of biological pathways**
G. Joshi-Tope
Nucleic Acids Research (2004-12-17) https://doi.org/cjhf62
DOI: 10.1093/nar/gki072 · PMID: 15608231 · PMCID: PMC540026

15. **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 · PMID: 31691815 · PMCID: PMC7145712

16. **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 · PMID: 29145629 · PMCID: PMC5753187

17. **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 (2019-11-06) https://doi.org/ggpx9m
DOI: 10.1093/nar/gkz1031 · PMID: 31691815 · PMCID: PMC7145712

18. **ChEBI in 2016: Improved services and an expanding collection of metabolites**
Janna Hastings, Gareth Owen, Adriano Dekker, Marcus Ennis, Namrata Kale, Venkatesh Muthukrishnan, Steve Turner, Neil Swainston, Pedro Mendes, Christoph Steinbeck
Nucleic Acids Research (2015-10-13) https://doi.org/f8cj93
DOI: 10.1093/nar/gkv1031 · PMID: 26467479 · PMCID: PMC4702775

19. **HMDB 4.0: the human metabolome database for 2018**
David S Wishart, Yannick Djoumbou Feunang, Ana Marcu, An Chi Guo, Kevin Liang, Rosa Vázquez-Fresno, Tanvir Sajed, Daniel Johnson, Carin Li, Naama Karu, … Augustin Scalbert
Nucleic Acids Research (2017-11-11) https://doi.org/gcwk5k
DOI: 10.1093/nar/gkx1089 · PMID: 29140435 · PMCID: PMC5753273

20. **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 · PMID: 25516281 · PMCID: PMC4302049

21. **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**
M. D. Robinson, D. J. McCarthy, G. K. Smyth
Bioinformatics (2009-11-11) https://doi.org/drxgw2
DOI: 10.1093/bioinformatics/btp616 · PMID: 19910308 · PMCID: PMC2796818

22. **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/ftng9m
DOI: 10.1038/nbt.1499 · PMID: 18953355 · PMCID: PMC2651818

23. **Metaboverse — Metaboverse 0.3.0-beta documentation** https://metaboverse.readthedocs.io/en/latest/

24. **Dietary methionine influences therapy in mouse cancer models and alters human metabolism**
Xia Gao, Sydney M. Sanderson, Ziwei Dai, Michael A. Reid, Daniel E. Cooper, Min Lu, John P. Richie Jr, Amy Ciccarella, Ana Calcagnotto, Peter G. Mikhael, … Jason W. Locasale
Nature (2019-07-31) https://doi.org/ggrj3d
DOI: 10.1038/s41586-019-1437-3 · PMID: 31367041 · PMCID: PMC6951023

25. **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/ggt7sk
DOI: 10.1093/gigascience/giz137 · PMID: 31972021 · PMCID: PMC6977586

26. **On Nearest-Neighbor Graphs**
D. Eppstein, M. S. Paterson, F. F. Yao

Discrete & Computational Geometry (1997-04) https://doi.org/brwn8g
DOI: 10.1007/pl00009293

27. **A vectorized "near neighbors" algorithm of order N using a monotonic logical grid**
Jay Boris
Journal of Computational Physics (1986-09) https://doi.org/bz69nr
DOI: 10.1016/0021-9991(86)90050-1

28. **Metaboverse/manuscript**
Metaboverse
(2020-09-27) https://github.com/Metaboverse/manuscript

29. **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 · PMID: 30352195 · PMCID: PMC6258005

30. **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 · PMID: 30118679 · PMCID: PMC6104058

31. **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/ggvrxg
DOI: 10.7554/elife.17828 · PMID: 27540631 · PMCID: PMC4991935

32. **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 · PMID: 27817865 · PMCID: PMC5142118

33. **Defects in mitochondrial fatty acid synthesis result in failure of multiple aspects of mitochondrial biogenesis inSaccharomyces cerevisiae**
V. A. Samuli Kursu, Laura P. Pietikäinen, Flavia Fontanesi, Mari J. Aaltonen, Fumi Suomi, Remya Raghavan Nair, Melissa S. Schonauer, Carol L. Dieckmann, Antoni Barrientos, J. Kalervo Hiltunen, Alexander J. Kastaniotis
Molecular Microbiology (2013-10-10) https://doi.org/ggxg4h
DOI: 10.1111/mmi.12402 · PMID: 24102902 · PMCID: PMC4153884

34. **High Level Expression and Characterization of the Mitochondrial Citrate Transport Protein from the YeastSaccharomyces cerevisiae**
Ronald S. Kaplan, June A. Mayor, David A. Gremse, David O. Wood
Journal of Biological Chemistry (1995-02-24) https://doi.org/c9h8k3
DOI: 10.1074/jbc.270.8.4108 · PMID: 7876161

35. **The mitochondrial dicarboxylate carrier is essential for the growth of Saccharomyces cerevisiae on ethanol or acetate as the sole carbon source**
Luigi Palmieri, Angelo Vozza, Angelika Honlinger, Klaus Dietmeier, Annamaria Palmisano, Vincenzo Zara, Ferdinando Palmieri
Molecular Microbiology (1999-01) https://doi.org/d6fj3j
DOI: 10.1046/j.1365-2958.1999.01197.x · PMID: 10027973

36. **Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition**
Marco Sciacovelli, Emanuel Gonçalves, Timothy Isaac Johnson, Vincent Roberto Zecchini, Ana Sofia Henriques da Costa, Edoardo Gaude, Alizee Vercauteren Drubbel, Sebastian Julian Theobald, Sandra Riekje Abbo, Maxine Gia Binh Tran, … Christian Frezza
Nature (2016-08-31) https://doi.org/f84vs5
DOI: 10.1038/nature19353 · PMID: 27580029 · PMCID: PMC5136292

37. **Metabolomic Markers of Altered Nucleotide Metabolism in Early Stage Adenocarcinoma**
W. R. Wikoff, D. Grapov, J. F. Fahrmann, B. DeFelice, W. N. Rom, H. I. Pass, K. Kim, U. Nguyen, S. L. Taylor, D. R. Gandara, … S. Miyamoto
Cancer Prevention Research (2015-02-05) https://doi.org/f69tn7
DOI: 10.1158/1940-6207.capr-14-0329 · PMID: 25657018 · PMCID: PMC4618700

38. **Evidence that glutamine, not sugar, is the major energy source for cultured HeLa cells.**
LJ Reitzer, BM Wice, D Kennell
The Journal of biological chemistry (1979-04-25) https://www.ncbi.nlm.nih.gov/pubmed/429309
PMID: 429309

39. **Glutaminolysis: A Hallmark of Cancer Metabolism**
Lifeng Yang, Sriram Venneti, Deepak Nagrath
Annual Review of Biomedical Engineering (2017-06-21) https://doi.org/ggzsfs
DOI: 10.1146/annurev-bioeng-071516-044546 · PMID: 28301735

40. **From Krebs to clinic: glutamine metabolism to cancer therapy**
Brian J. Altman, Zachary E. Stine, Chi V. Dang
Nature Reviews Cancer (2016-07-29) https://doi.org/gc95vq
DOI: 10.1038/nrc.2016.71 · PMID: 27492215 · PMCID: PMC5484415

41. (2014-01-28) http://conference.scipy.org/proceedings/SciPy2008/paper_2/full_text.pdf

42. **Electron | Build cross-platform desktop apps with JavaScript, HTML, and CSS.** https://www.electronjs.org/

43. **Welcome to Python.org**
Python.org
https://www.python.org/

44. **PyInstaller Quickstart — PyInstaller bundles Python applications** https://www.pyinstaller.org/

45. **D3.js - Data-Driven Documents**
Mike Bostock
https://d3js.org/

46. **edeno/d3-save-svg**
Eric Denovellis
(2020-08-19) https://github.com/edeno/d3-save-svg

47. **adambisek/string-pixel-width**
Adam Bisek
(2020-09-26) https://github.com/adambisek/string-pixel-width

48. **Travis CI - Test and Deploy Your Code with Confidence** https://travis-ci.org/

49. **jQuery**
JS Foundation- js.foundation
https://jquery.com/

50. **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

51. **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

52. (2007-01-16) https://web.mit.edu/dvp/Public/numpybook.pdf

53. **SciPy 1.0: fundamental algorithms for scientific computing in Python**
Pauli VirtanenRalf 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 · PMID: 32015543 · PMCID: PMC7056644

54. **Matplotlib: A 2D Graphics Environment**
John D. Hunter
Computing in Science & Engineering (2007) https://doi.org/drbjhg
DOI: 10.1109/mcse.2007.55

55. **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 · PMID: 32004313 · PMCID: PMC7015430

56. **Metaboverse/manuscript**
GitHub
https://github.com/Metaboverse/manuscript

57. **PR000305**
Metabolomics Workbench
Metabolomics Workbench (2016) https://doi.org/ggxfdh
DOI: 10.21228/m8pg66

58. **GEO Accession viewer** https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc

59. **PR000961**
Metabolomics Workbench
Metabolomics Workbench (2019) https://doi.org/gg2m2w
DOI: 10.21228/m8gd71

60. https://github.com/Metaboverse/manuscript/tree/master/data/sce_mct1