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

This manuscript was automatically generated on June 11, 2020.

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 & 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 organism. Perturbations to any of component of a reaction can have rippling downstream effects, which can be difficult to trace across the global reaction network, particularly when the effects occur between canonical representations of pathways. Researchers have primarily utilized reductionist approaches to understand metabolic reaction systems; however, customary methods often limit the scope of the analysis. Even the power of systems-centric omics approaches can be limited when only a handful of high magnitude signals in the data are prioritized. 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 global reaction networks. This framework will be foundational in increasing our ability to holistically understand static and temporal metabolic events and shifts and gene-metabolite intra-cooperativity. Metaboverse is freely available under a GPL-3.0 license at https://github.com/Metaboverse/.

Graphical Abstract (displayed as Figure 1)
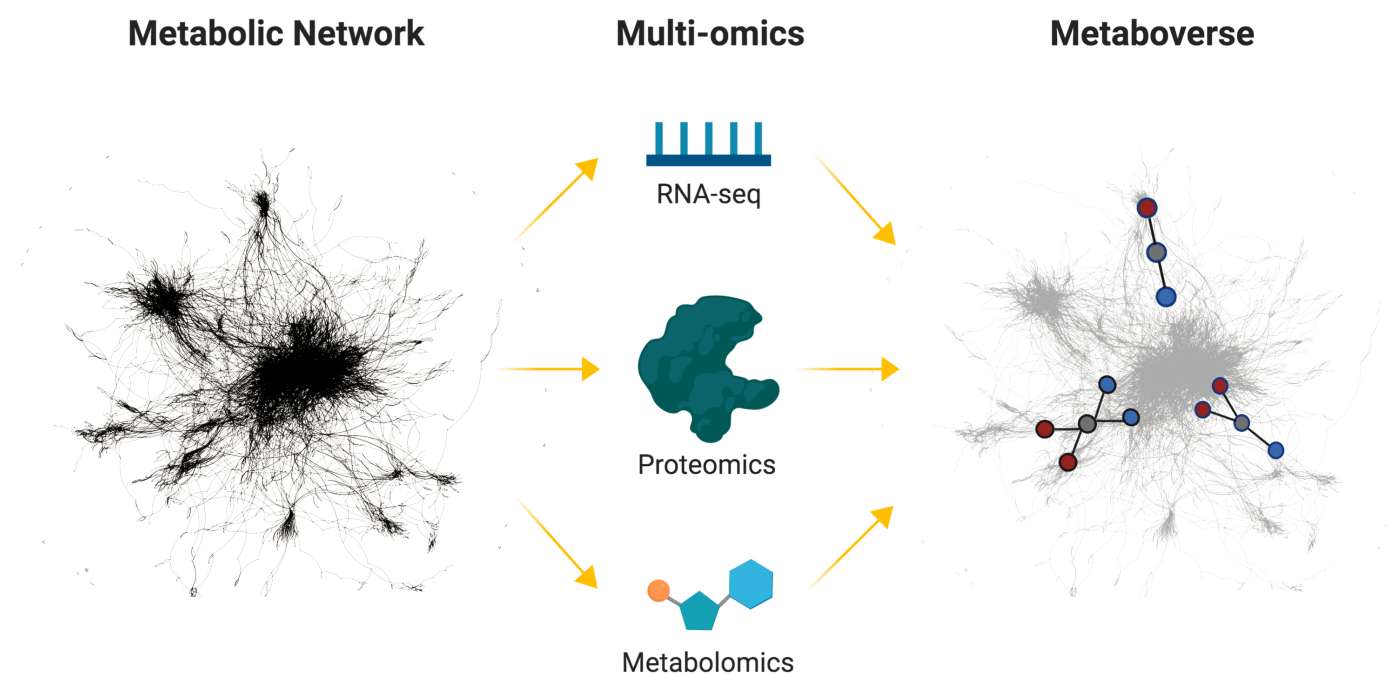
Figure 1: Metaboverse conceptual overview. Illustration summarizing the usage of Metaboverse to model biological data on the global reaction network to rapidly identify regulatory hotspots. Traditionally, when a scientist performs an omics experiment, they tend to focus on a few features that are differentially regulated. Metaboverse allows a user to input multiple omics data types which it then layers across the metabolic network. Metaboverse then uses this integrated model to identify interesting regulatory patterns in the data. Users can also dynamically explore metabolic pathways and other network representations.

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 a reductionist approach to tease apart the characteristics and mechanics of these processes and how they fit into the larger picture of biology and disease. While such approaches are a vital component in the scientific process, a reductionist may miss many interesting 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 rippling effects across the network [2,3]. The current approach is analogous to telling the story of Little Red Riding Hood, but only by reading the 20 most frequent words used in the study. Indeed, doing so efficiently highlights keywords like "wolf" and "little red riding hood," but also prevents a coherent story from being told and would make it difficult for someone who had never heard the story of Little Red Riding Hood to comprehend the actual plot (Figure 2). In short, biological perturbations involve complex, cooperative effects, many of which seem negligible in isolation.

Figure 2: Word cloud of 20 most frequent words of phrases in the story of Little Red Riding Hood. Text taken from the original telling of Little Red Riding Hood by Charles Perrault. Words and phrases derived using [4]. Outline derived from the piece, "Little Red Riding Hood" by J.W. Smith.

Over the past decade, several computational tools have risen to prominence for their attempts to resolve these issues in data contextualization. We will highlight four such tools that are popular and representative for their respective properties. First is MetaboAnalyst, which relies heavily on set enrichment methods for analysis of data, or looking at the belongingness of sets of significantly changed analytes (i.e., metabolite, protein, or gene measurements), for extracting interesting information. While network visualization is available, it focuses primarily on interaction networks, and its ability to extract regulatory information is limited, particularly in an automated fashion [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 develop customized

analyses; however, comprehensive and metabolism-specific regulatory identification methods are unavailable [7]. One plug-in for Cytoscape that is focused on metabolic data is MetScape, but again, this tool is limited to pathway enrichment, correlation networks, and simple 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 curations. It additionally can layer experimental data on the network for visualization [11,12]. Reactome, which Metaboverse uses for the curation of biological networks, also offers analytical tools for user data, but again relies on set enrichment or manual methods for identifying patterns. While all have their respective utility, there is still a need for tools that integrate these features and automate pattern and trend detection, especially when data is sparse, across metabolic networks in order to extract regulatory and other features from data [13,14, 15]. This need is particularly pronounced in common cases where experimental data have sparse coverage of the network, a situation common in modern metabolomics.

To address these limitations in current conventions of metabolic data analysis, contextualization, and interpretation, we created Metaboverse, an operating-system-independent application, to aid users in filling in and expanding upon the details of their model's metabolic story. Metaboverse is an interactive tool for exploratory data analysis that searches user data in the context of the metabolic network to identify unusual patterns and trends within the data. By identifying these trends and patterns across the global reaction network, Metaboverse will aid scientists in formulating new hypotheses from their data and in designing follow-up experiments for a deeper understanding of their model. In addition to a pre-designed pattern search library, additional patterns will be included in future releases of Metaboverse, along with the ability for a user to interactively design patterns of interest to be included in the pattern search library.

To provide a platform for the exploration of single or multi-omic metabolic data, we developed several computational features to aid in the aims discussed above. We developed a pattern search engine for the rapid and automated identification of patterns and trends in omics data on the metabolic network. Conceptually, this search engine borrows principles of topological motif searching from graph theory [16,17] and builds upon the concept of "activity motifs" [18]. We will use "search pattern" and "motif" interchangeably. Another feature introduced in Metaboverse allows for the interactive exploration of specific reactions or reaction entities with on-the-fly pattern search analysis. Users can explore specific pathways of interest and look for other interesting patterns and trends within these pathways. Perturbations in the abundance or behavior of a particular metabolic component can lead to downstream genotypic and phenotypic modulations in a biological system. We include an interactive perturbation connectivity module to allow users to explore the effects of perturbations on their system.

Finally, we tackle the challenge of sparsity, particularly in metabolomics datasets. We designed a reaction collapsing feature that allows for searching for motifs across multiple reactions, where intermediate steps of the reaction trajectory may not have measured values for their respective input or output components. This is particularly important in allowing for pattern detection across missing components of a pathway. For example, the human metabolic network alone consists of over 1000 metabolite nodes, but an average gas chromatography mass spectrometry (GC-MS) experiment may only quantify 100-200 metabolites by name. By bridging proximal reactions, one can detect patterns that would require the comparison of the terminal ends of that sub-pathway. This can help users identify more inconspicuous patterns within the global reaction network.

Metaboverse is currently designed to handle standard two-condition experiments and time course or multi-condition experiments. Users can provide any combination of datasets from each omics category (values mapping to gene, protein, and metabolite identifiers). Time course inputs can consist

of matched samples for every time point or can provide a combination of steady-state levels and time points. Users input fold change and appropriate statistical measures with proper multiple-testing adjustments from their respective omics datasets, which Metaboverse then layers onto the metabolic network. Metaboverse can handle data from a variety of model organisms, including humans, mice, yeast, zebrafish, and more. The foundational curation of Metaboverse is built on Reactome reaction curations, providing 90+ species which the Metaboverse environment can process. To validate these methodologies available in Metaboverse, we analyzed a variety of datasets to provide representative vignettes that highlight Metaboverse's reliability in extracting canonical features, as well as utility in identifying novel features and patterns. We expect that Metaboverse will become a foundational piece of the analytical toolkit and augment one's ability to more deeply and holistically explore metabolism and other reaction networks. It will importantly aid in our ability to provide clearer context within these models.

Results

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

Overview.

We designed Metaboverse as a light-weight, self-contained app for the dynamic exploration of high-throughput biological data. The pathway curations are derived from Reactome [19], coupled with metabolite name cross-referencing from the ChEBI [20] and the Human Metabolome (HMDB) [21] databases. As of the time of writing this manuscript, Metaboverse is capable of analyzing data for over ninety species. A user begins by providing the desired output location for a new curation and organism of interest. If the user has previously curated data onto a reaction network for their 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 layering onto the network and downstream analysis. For each omics type, users provide 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 wild-type conditions, as well as the appropriate statistical values. As transcriptomics count data follows a negative binomial distribution, the appropriate adjusted p-values based on multiple hypothesis testing should be used [22]. 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. Following these user inputs, the organism's network, data processing, and motif analysis (discussed further below) are curated, and a curation file is output for future analysis (see Figure 3).

We chose to format network visualizations in a reaction-centric format. This means that a sub-network unit is comprised of a reaction node, linked with its reaction input and reaction output nodes, along with any modifier 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 users quickly distinguish the details of a reaction sub-network structure. Reactions with matching input or output sides are then naturally linked, allowing a multi-reaction pathway to emerge from this network structure. This formatting differs from other formatting styles users may be familiar with, such as KEGG pathway maps, for various of reasons. While this methodology dynamic visualization differs from more classical

methods, each pathway in the visualization module contains a button linking to a more classically formatted representation of the pathway for reference. And while classical pathway definitions are used to group sets of reactions, analytical methods in Metaboverse search the global reaction network. This approach aids in making these approaches less subject to biases arising from constraining analysis to a pathway that is not directly contextualized with its place in the global network.
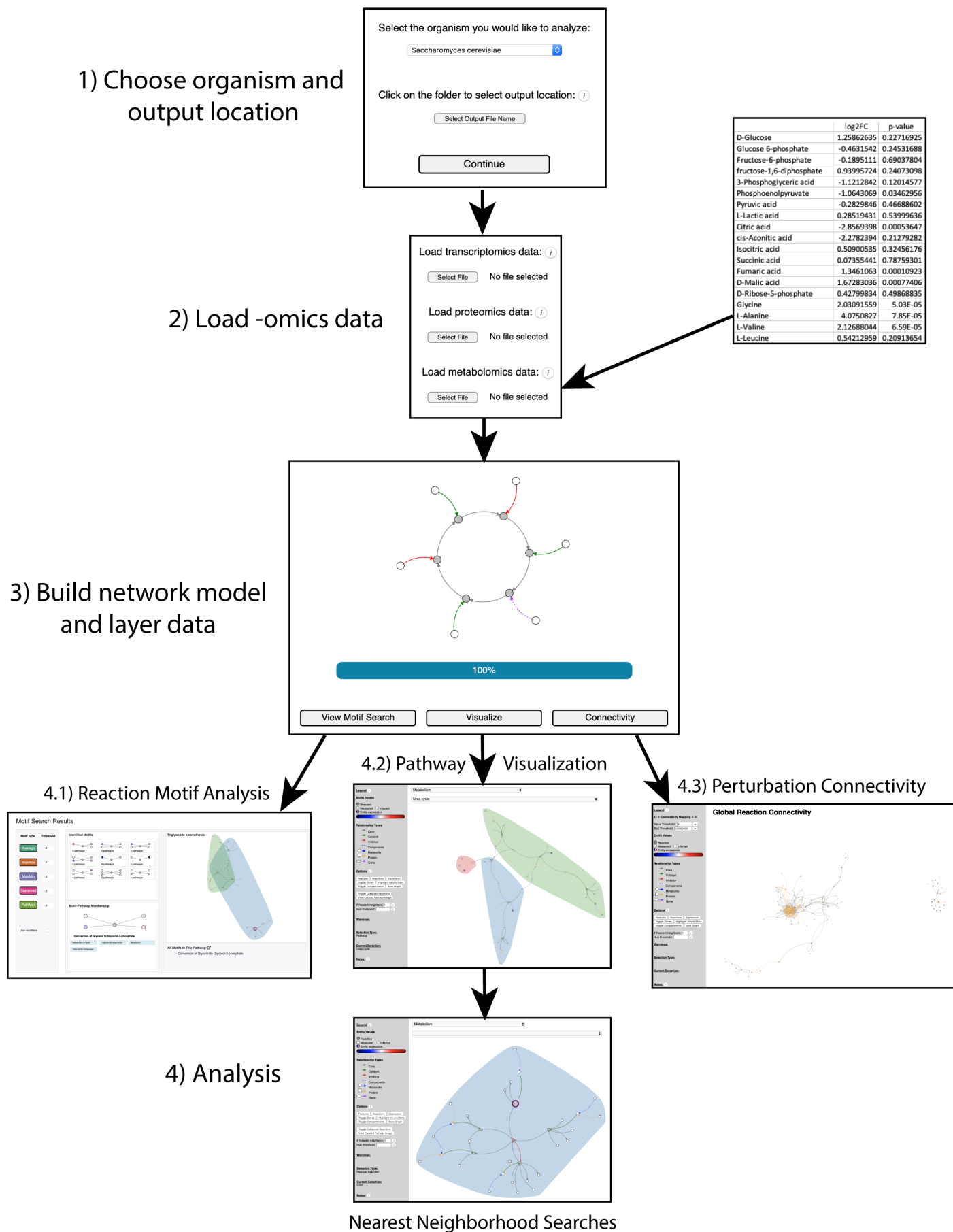
Figure 3: Metaboverse package overview. Illustration outlining the data input, processing, and exploration steps of the Metaboverse package. Users provide the organism of interest's name from a drop-down menu along with an output location. Users then have the option to provide transcriptomics, proteomics, 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. During this step, users can also provide sample labels and other modifiers that will 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 reaction motifs identified, explore pathway or global perturbation networks, or perform general pathway exploration of the data.

Handling data sparsity within the biological global network.

Missing data points, particularly in metabolomics experiments, are frequent and can make the analysis of pathways and identification of regulatory patterns in the network challenging. While thousands of metabolites are known to participate in human metabolism, the current state of the technology for determining which mass spectra belong to which metabolite can be challenging and often results in a limited number of data points being quantified. These can lead to gaps in the metabolic network, which can be challenging to explore and analyze. For example, the human metabolic network is comprised of over 1000 metabolites, while the average GC-MS experiment may only measure 100-200 metabolites by name, leading to measurement gaps between reactions in a pathway. We, therefore, developed a reaction compression algorithm (detailed more in the Methods section) that collapses up to three reactions with missing data points if they can be bridged with known data on the distal ends of the reaction path. Methods similar to this have been used in metabolomic analysis before to identify amino acid-related metabolites [24], but we introduce the first computational implementation to our knowledge and adopt a slightly more conservative collapsing scheme. These reactions, or pseudo-reactions, are visually distinct to show users where reactions were collapsed and understand what that pseudo-reaction summarizes.

Rapid identification of interesting regulatory patterns in the reaction network.

Following network curation, the user can visualize available reaction motifs identified across the global reaction network. In a computational science context, a motif is a recurring pattern in network structure or the organization of network entities and their relationships to one another. However, with omic data, we are more interested in identifying patterns in expression or abundance of genes, proteins, and metabolites. Previous work by Checkik, et al. introduced the concept of "activity motifs," where instead of identifying motifs based on network structure, they were identified using the expression characteristics of nodes in transcription factor binding signaling networks [18]. We adapted this methodology to identify and interactively display interesting regulatory hotspots within the global reaction network. For example, a reaction's inputs may exhibit increased levels and its outputs decreased levels, indicating a regulatory event occurring at the reaction.

In Metaboverse, we define a motif as a regulatory pattern identified across a reaction or pseudo-reaction. Metaboverse contains a library of default motifs to search for within the global reaction network, and users can define custom motifs they would like to identify across the global network (please refer to the documentation for more details on available motifs [25]). Metaboverse will search the global network from a pre-defined library of regulatory patterns (regulatory hotspots, or motifs) and return a graphical stamp view of conserved patterns. These motifs by default are ordered by the statistical values associated with the entity components of the motif, but can be ordered by the strength of the motif based on the magnitude of change. Statistical values associated with each measurement for a measured entity as provided by the user are used to weight these motifs and prioritize returned results. We use a three-tiered sorting strategy when sorting by the associated statistical values. The highest prioritized motifs are those where the relevant components on either side of the reaction (inputs vs. outputs) that determined the reaction motif are statistically significant. Of these motifs from the first tier, motifs are sorted by lowest to highest cumulative p-value or other relevant statistical value. In the second sorting tier, reactions with at least one side of the reaction with a statistically significant motif component are sorted by statistical strength. Finally, all other motifs are sorted by the cumulative statistical value for the motif components.

Reaction motif nodes are visually distinct to quickly draw the user's attention to these interesting patterns. For a given pattern, the user can then explore each pathway this particular motif is found in. Motif analysis of the global regulatory network will allow users to rapidly identify interesting features within the data, particularly patterns between canonical pathways or in other pathways that may seem tangential in their research, but are instead rather striking based on these contextualized analytical methods. In the data vignettes below, we demonstrate this utility further. Users will also be able to design their own patterns through an interactive pattern drawing tool and even design specific scenarios that are cognizant of feature type (in progress). For example, one might be interested in a pattern where a protein displays higher expression, but the resulting metabolite is decreased. This could then be coded interactively into the Metaboverse framework by the user.

Dynamic visualization of organism-specific reaction pathways.

Following the curation of the global network as described above, the user can manually search individual canonical pathways or individual entities and their reaction neighborhoods. For a given selection, all relevant reactions that are annotated as a part of that pathway will be visualized, along with their core input (substrate) and output (product) components. In addition to these core elements, known catalysts and inhibitors are included, as well as the component proteins, genes, and metabolites that form functional complexes involved in a particular reaction. For complex nodes, simulated values are calculated by taking the average of all measured component entities in that complex. Assuming a statistical value between 0 and 1, simulated statistics are calculated by taking the maximum statistic value of all measured component entities in that complex. In cases where a gene value is known, but its protein value is unmeasured, the protein value can be inferred using aggregated gene component values. Relevant pathway and analytical metadata are also displayed. All reaction motif types are displayed using default thresholds if present in a selected pathway. Additional aids for visualization are also available, such as 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. Often, these hub nodes are ubiquitous features such as water and protons, which may be of limited interest to the user during data visualization [26]. Compartment domains are also visualized to include relevant reactions and their components that occur in a given cellular compartment.

Visualization of downstream effects of network perturbations using reaction neighborhoods.

Users may be interested in a particular metabolite or protein and the downstream effects their perturbation has on related pathways. By double-clicking a node of interest, or by selecting the entity name from the drop-down menu, the user can explore all downstream effects across all pathways in the global network. The user can also define how many neighborhoods to display such that one can visualize two or more reaction steps downstream or upstream of the selected entity [27,28]. This functionality moves the analysis away from traditional, pathway-centric approaches, and helps contextualize the far-reaching effects changes in metabolism or other biological systems can have across classical pathways.

Exploring perturbation connectivity within the global network

Abundance or behavioral changes of a metabolic network component can lead to downstream genotypic and phenotypic modulations in a biological system. One important measure of robustness of a network is "connectivity" [26,29]. In a biological context, an example of connectivity, or the loss of network connectivity, is easily grasped when considering a transport reaction from the cytosol into the mitochondria of a critical metabolite [26,30,31,32]. When the transport hub is abolished, and

network connectivity is lost, the critical metabolite cannot participate in the required downstream reactions. For example, our laboratory previously established the consequences of the ablation of the mitochondrial pyruvate carrier on downstream citric acid cycle processes and the corresponding increase in lactate production [33] and serious cellular remodeling and dysfunction [34,35].

The importance of connectivity could also be considered in a clinical context. For example, a druggable and critical metabolite may be perturbed in a particular disease context. However, if a metabolite participating in a neighboring, downstream reaction is also perturbed in a way not related to the perturbation of the first metabolite, the efficacy of the drug treatment could be severely impaired [29,36,37,38]. To aid in the exploration of the connectivity of the biological network, we developed a connectivity module where users can display all global reactions or reactions of a specific super pathway that have at least one involved component perturbed based on either an abundance or statistical level. By doing so, proximal reactions that were perturbed will be "sewn" together to reconstruct a perturbation connectivity map in the network visualization. Perturbed, but un-connected reactions are displayed as single reactions in the network.

Data vignettes

In order to demonstrate the utility of Metaboverse to the community, we used Metaboverse to analyze public and new datasets. From the vignettes provided below, we show that Metaboverse not only identifies points of interest previously described or expected, but also rapidly identifies unexpected and systematic regulatory patterns in a reaction network context.

1. Perturbation of mitochondrial fatty acid synthesis across time.

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 can perform metabolic nutrient-sensing roles to coordinate lipoic acid synthesis with iron-sulfur (Fe-S) cluster biogenesis and assembly of oxidative phosphorylation complexes, and more [39,40,41]. Additionally, this pathway has received increased physiological focus with the discovery of patients with mutations in key mtFAS enzymes [42].

Mct1 in *Saccharomyces cerevisiae* is an acyltransferase responsible for the transfer of a malonyl group from malonyl-CoA to the mitochondrial acyl carrier protein (ACP). By knocking out this gene, one can simulate the effects of a disruption in mtFAS. In order to probe the relationship between mtFAS-related protein concentration and the effects of its perturbation on downstream metabolic processes from a systematic perspective, we used a *mct1Δ* model in *Saccharomyces cerevisiae*. In this model, we previously measured steady-state proteomics in this *mct1Δ* yeast model 12 hours after shifting to a non-fermentable carbon source [40]. For this study, we additionally measured transcriptomics at 0, 3, and 12 hours after a shift to a non-fermentable carbon source, as well as steady-state metabolomics at 0, 15, 30, 60, and 180 minutes after the shift in carbon source. By layering these data onto the *Saccharomyces cerevisiae* metabolic network using Metaboverse, we observed interesting respiratory signatures as expected based on previous work [40,43]. For example, we noticed a strong pattern in the electron transfer from ubiquinol to cytochrome C of complex III of the electron transport chain (ETC) (Figure 4 A). At the proteomics level, cytochrome C component proteins, CYC1 and CYC7 are both significantly reduced in concentration compared to wild-type cells ($\log_2$(fold change): -1.57 & -0.88; p-value: < 0.01 & < 0.01; 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 cytochrome C reductase, which is a catalyst of this electron transfer reaction (average $\log_2$(fold change) of all measured protein components = -2.03,

where 9/11 component proteins were measured; 7/11 passed statistical significance thresholds of BH-corrected p-value <= 0.05. Range of significant $\log_2$(fold change) values: -0.34 to -3.21).
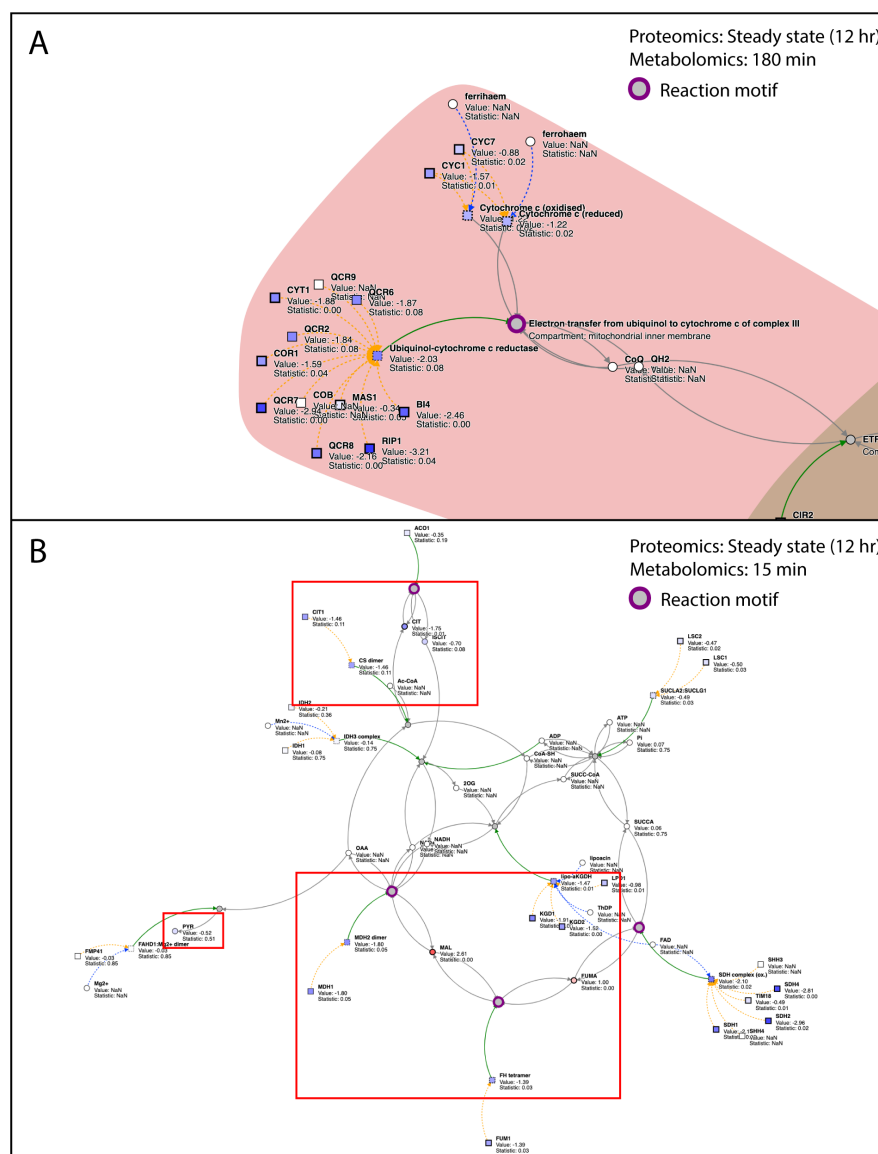
Figure 4: Metaboverse identifies several reaction motifs of interest in *mct1Δ* cells. (A) Steady-state proteomics (12 hours) overlaid on the reaction, "Electron transfer from ubiquinol to cytochrome c of complex III". (B) Steady-state proteomics (12 hours) and metabolomics at 15 minutes overlaid on TCA-related reactions. Appropriate time stamps for the data type are displayed in the upper-right hand corner of each subplot. Metabolomics and proteomics 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 Changes and statistical values for each entity are displayed below the node name. A grey node indicates a reaction. A bold grey node with purple border indicates a motif was found at this reaction for the given data type time points. Non-reaction circles indicate metabolites and squares indicate proteins or protein complexes. Grey 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 it is involved in. Dashed orange edges point from a protein component to the complex it is involved in. Protein complexes with dashed borders indicate that the values displayed on that node were inferred from the constituent protein and metabolite measurements. The shading in sub-plot A demonstrates Metaboverse's ability to show compartmentalization, while sub-plot B has compartmentalization shading turned off.

The second pattern of interest that was expected was the marked reduction of TCA-related enzymes. Motif analysis within the TCA cycle found several regulatory hotspots between metabolites and metabolite-protein interactions throughout the steps of the TCA cycle (Figure 4 B). However, visualizing these motifs and data across the time course reveals interesting patterns. For the following metabolite values, we will provide those for the 15 minute time point as the statistical strength for the

discussed metabolites is strongest at this time. First, citrate levels decrease across the metabolomics time course (at 15 min; $\log_2$FC: -0.-1.75, adjusted-p: 0.01; Cohen's d: -1.86), which is paired by reduced steady-state levels of Ctp1 ($\log_2$FC: -0.64, adjusted-p: 0.03; Cohen's d: -4.73), a protein that catalyzes the transfer of citrate from the mitochondrial matrix to the cytosol [44]. Citrate is a key metabolite and the first step in the TCA cycle. We can hypothesize that due to central carbon metabolite reductions, Ctp1 may be down-regulated as a form of regulation to maintain citrate in the mitochondria where it is most physiologically important from a metabolism point of view.

Another point of interest is the up-regulation of Dic1 ($\log_2$FC: 2.15, adjusted-p: < 0.01; Cohen's d: 28.97), which catalyzes the exchange of malate (at 15 min; $\log_2$FC: 2.61, adjusted-p: < 0.01; Cohen's d: 5.12) between the mitochondrial matrix and cytosol [45]. Interestingly, Dic1 is essential for growth in non-fermentable carbon source media. When yeast, especially those with deficits in TCA metabolism due to ablation of Mct1, are switched to a non-fermentable carbon, as was done in this experiment, they must adapt by up-regulating Dic1, as we see in this situation.

We also see that while the catalytic enzymes in the TCA cycle are all reduced in concentration, several related metabolites are up-regulated across multiple time points in the dataset. It possible, for example, that the increased fumarate levels (at 15 min; $\log_2$FC: 1.00, adjusted-p: < 0.01; Cohen's d: 3.89) are related to the reduction in active fumarate hydratase (FH tetramer, FUM1; $\log_2$FC: -1.39, adjusted-p: 0.03; Cohen's d: -4.91) as has been demonstrated previously and shown to associate with hereditary leiomyomatosis and renal cell cancer in humans [46]. Therefore, this would be an interesting point of further study to understand the adaptations these cells make to disruptions to the TCA cycle and their consequences.

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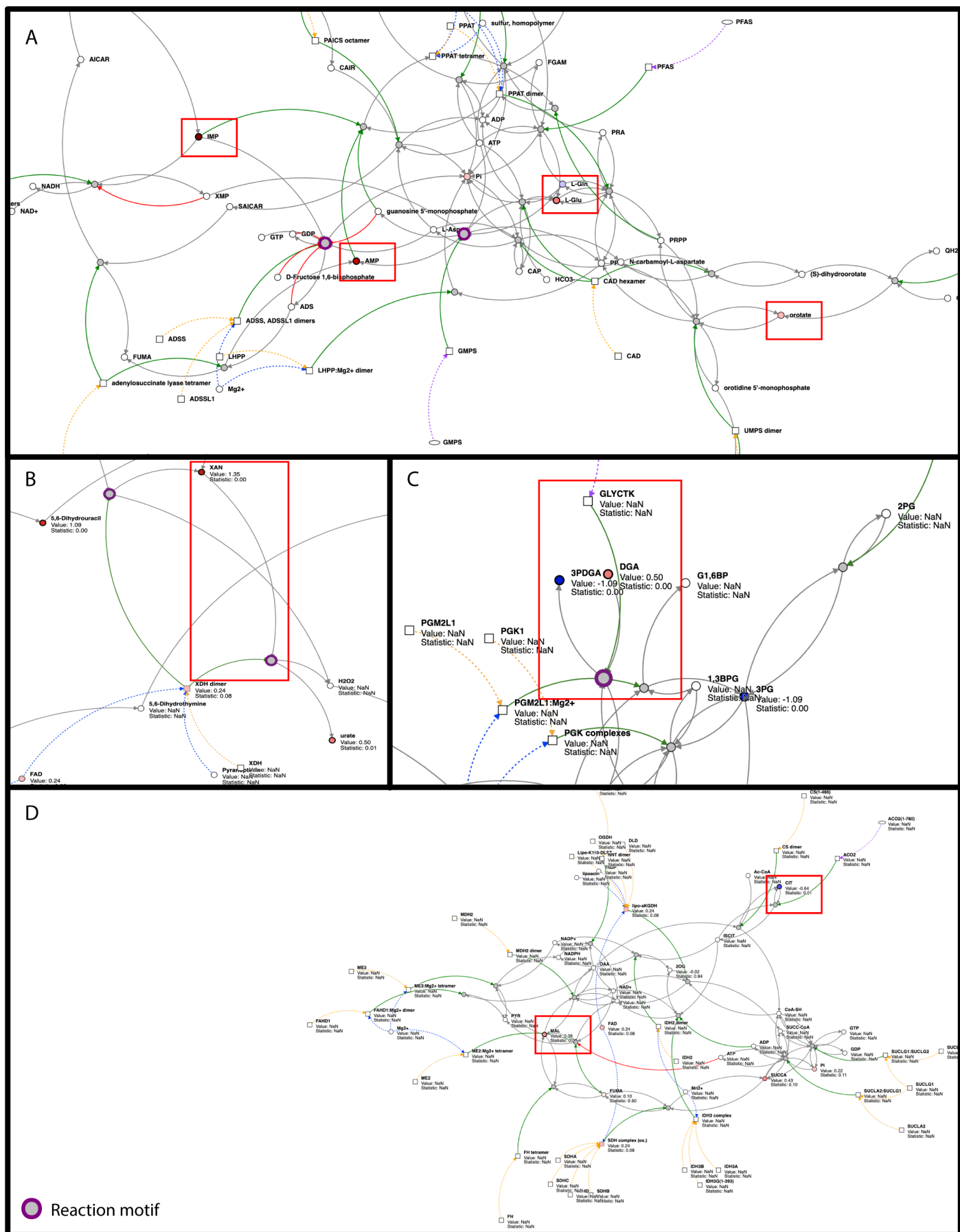
2. Metabolic signatures in human lung adenocarcinomas compared to normal lung tissue.

We next turned to published human lung adenocarcinoma steady-state metabolomics data [47] to assess the utility of Metaboverse when analyzing this data. Lung cancer remains a leading cause of death worldwide, and a more complete understanding of the metabolism of these tumors is essential in understanding how to better treat the condition.

Consistent with the original study [47], and in our recent re-study of the data [26], nucleotide metabolism was broadly up-regulated in adenocarcinoma samples in the motif analysis and perturbation connectivity analysis (Figure 5 A). We also notice similar perturbations in xanthine ($\log_2$ Fold Change: 1.35; p-value: 0.0003; Cohen's d: 0.86) (Figure 5 B) and others, which are highlighted to the user quickly in the motif and perturbation connectivity analyses. However, as we previously emphasized [26], by approaching these perturbations in a reaction-centric approach, we can identify regulatory behavior that further contextualizes the data. For example, we previously highlighted through a more manual approach the up-regulation of glyceric acid ($\log_2$ Fold Change: 0.50; p-value: 0.002; Cohen's d: 0.73) coupled with the proximal down-regulation of 3-Phosphoglyceric acid ($\log_2$ Fold Change: -1.09; p-value: 0.004; Cohen's d: -0.68) which could indicate the activity of glycerate kinase [26]. This connection was missed in the original study, but was highlighted by the motif analysis available within Metaboverse. Activity of this sub-pathway is consequential as it has

connections to serine metabolism which may be fostering an ideal environment for tumorigenesis (Figure 5 C). Interestingly, within 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: 0.02; Cohen's d: 0.56) and down-regulation of citrate ($\log_2$ Fold Change: -0.64, adjusted-p: 0.01; Cohen's d: -0.63). This is consistent with the hypothesis that the TCA cycle is being starved of carbon via pyruvate and compensating for this deficiency by shunting carbon into the TCA cycle via glutaminolysis. This compensatory pathway, a feature in many cancers, consists of glutamine being transported into mitochondria where they can be converted to alpha-ketoglutarate and enter the second half of the TCA cycle to generate ATP [48,49]. The pyruvate that would have normally been processed within the TCA cycle is probably being shunted to lactate production (no measurements were available in the dataset), thus the lower citrate levels, while part of the TCA cycle can be fed through glutamate ($\log_2$ Fold Change: 0.47, adjusted-p: < 0.01; Cohen's d: 0.98), thus the increased malate levels (Figure 5 D) [50].

Metaboverse simplifies the analysis process and allows users more flexibility via its interactive platform to identify canonical and interesting regulatory patterns within their data with ease. While we previously identified the regulatory pattern identified in Figure 5 C, Metaboverse offers an automated platform for the discovery of such events, which were missed in the original study. We hope that by using Metaboverse, users will be able to extract new and exciting hypotheses that can drive their field forward.

Figure 5: Metaboverse identifies nucleotide metabolism signatures in lung adenomcarcinoma metabolomics data. (A) Up-regulation of several nucleotide metabolites. (B) Identification of xanthine regulation by both motif and perturbation connectivity analysis. (C) Regulatory activity between glyceric acid and 3-Phosphoglyceric acid identified by motif analysis. (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 Changes and statistical values for each entity are displayed below the node name. A grey node indicates a reaction. A bold grey node with purple border indicates a motif was found at this reaction for the given data type time points. Non-reaction circles indicate metabolites, squares indicate proteins or protein complexes, and ellipses indicate gene components. Grey edges are core relationships between reaction inputs and outputs. Green edges indicate a catalyst, while red edges indicate inhibitors. Dashed blue edges point from a metabolite component to the complex it is involved in. Dashed orange edges point from a protein component to the complex it is involved in. 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 and metabolite measurements.

Discussion

In this manuscript, we introduce a new software tool for analysis and exploration of user data layered on the metabolic and global reaction networks. To improve on tools with similar capabilities, we introduced several new analytical tools and methods to aid users in the automated identification and discovery of regulatory patterns within their data in a reaction network context. These include the automated ability to identify reaction regulatory events across the global reaction network, such as a reaction where an input has a low measured abundance and an output has a high 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 motif is propagating across the global reaction network and not just a single pathway, they can explore a reaction component's nearest reaction neighborhood. The user can also explore the connectedness of perturbations across the global network and begin to explore hypotheses around the role of connected or disconnected reactions within a particular biological model.

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

We demonstrated the utility of Metaboverse in exploring single- and multi-omic datasets. We analyzed previously published studies and generated a novel dataset that highlights the time course, multi-omic capabilities of this framework. We demonstrated that Metaboverse was able to identify regulatory motifs that were expected in the models based on the current literature, as well as identify intriguing patterns that lead us to form new hypotheses. We expect these features to be a powerful tool in the researcher's toolkit as they analyze their data and plan their next steps. In the near future, we will implement several additional features that will expand the applicability of Metaboverse. For example, we will integrate additional common data quality control features for each data type, the ability to visualize and analyze flux metabolomics data, additional data pre-processing modules, other more targeted analysis approaches to aid users in following up on motif reaction search results, and an interactive motif builder.

While Metaboverse aims to enhance the computational toolkit for data analysis and hypothesis generation in metabolic and other experiments, a number of challenges remain, which we intend to address in subsequent versions of this software. For example, while the reaction collapsing features of Metaboverse aid in identifying patterns across several reactions where data may be missing, there are various biological and technical edge cases that need to be considered in future implementations of this feature. This is particularly challenging in datasets where few metabolites were measured. Hopefully, as technical limitations in metabolomics are also overcome, more complete snapshots of

metabolism will be visible. Additionally, while we take a more straightforward and somewhat rudimentary approach to statistical significance integration in the reaction motif searches, more holistic platforms for cross-omics integration are needed and remain a significant challenge across multi-omics research.

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

Methods

A tutorial for how to use Metaboverse can be found at metaboverse.readthedocs.io/getting-started.

1. Network Curation

Biological networks are curated using the current version of the Reactome database. In particular, the pathway records for each species, complex component and interaction data, Ensembl, and UniProt Reactome mapping tables are integrated into the network database for Metaboverse. Additionally, the ChEBI and The Human Metabolome databases are also referenced for metabolite synonym mapping [20,21]. These data are used to generate a series of mapping dictionaries for entities to reactions and reactions to pathways for the curation of the global network.

After the relevant information is parsed from each table or record, the global network is propagated using the NetworkX framework [51] 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 properly mapped to both nodes.

After the network is curated for the user-specified organism, each node's degree (or magnitude of edges or connections) is determined to aid in the user's downstream ability to avoid visualizing high-degree components, such as a proton or water, on the metabolic network, which can lead to visual network entanglement and cluttering and a decrease in computational performance [26].

2. Data overlay and broadcasting for missing entities

In order to overlay user data on the global network, first, user-provided gene expression, protein abundance, and/or metabolite abundances' names are mapped to Metaboverse compatible identifiers. For components that Metaboverse is unable to map, a list will be returned to the user so they can provide alternative names to aid in mapping. Second, provided data values are mapped to the appropriate nodes in the network. In cases where gene expression data is available, but protein abundance data is missing, Metaboverse will take the average 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 user data

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

1. If a reaction has at least one known or inferred value for inputs (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 are included when assessing whether a reaction has measured output values (more of a catalyst should lead to more output in most cases) and inhibitors are included when assessing whether the reaction has measured input values (more inhibitor should lead to accumulation of input in most cases). Catalysts and inhibitors are not included when determining reaction neighbors, as described below.

2. If a reaction has at least one known input, the input is left as is, and each reaction that shares the same input with the assessed reaction inputs are determined whether they have a measured output. If the neighbor reaction does not contain a known output value, the reaction is left as is. If the neighboring reaction does contain a measured output, the original 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 identical components as 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.

For pseudo-reactions, appropriate notes are included to describe the collapse. During visualization, these pseudo-reactions are marked by black dashed edges and dashed node borders. A graphical representation of how this reaction collapse scheme is executed can be found in Figure 6.
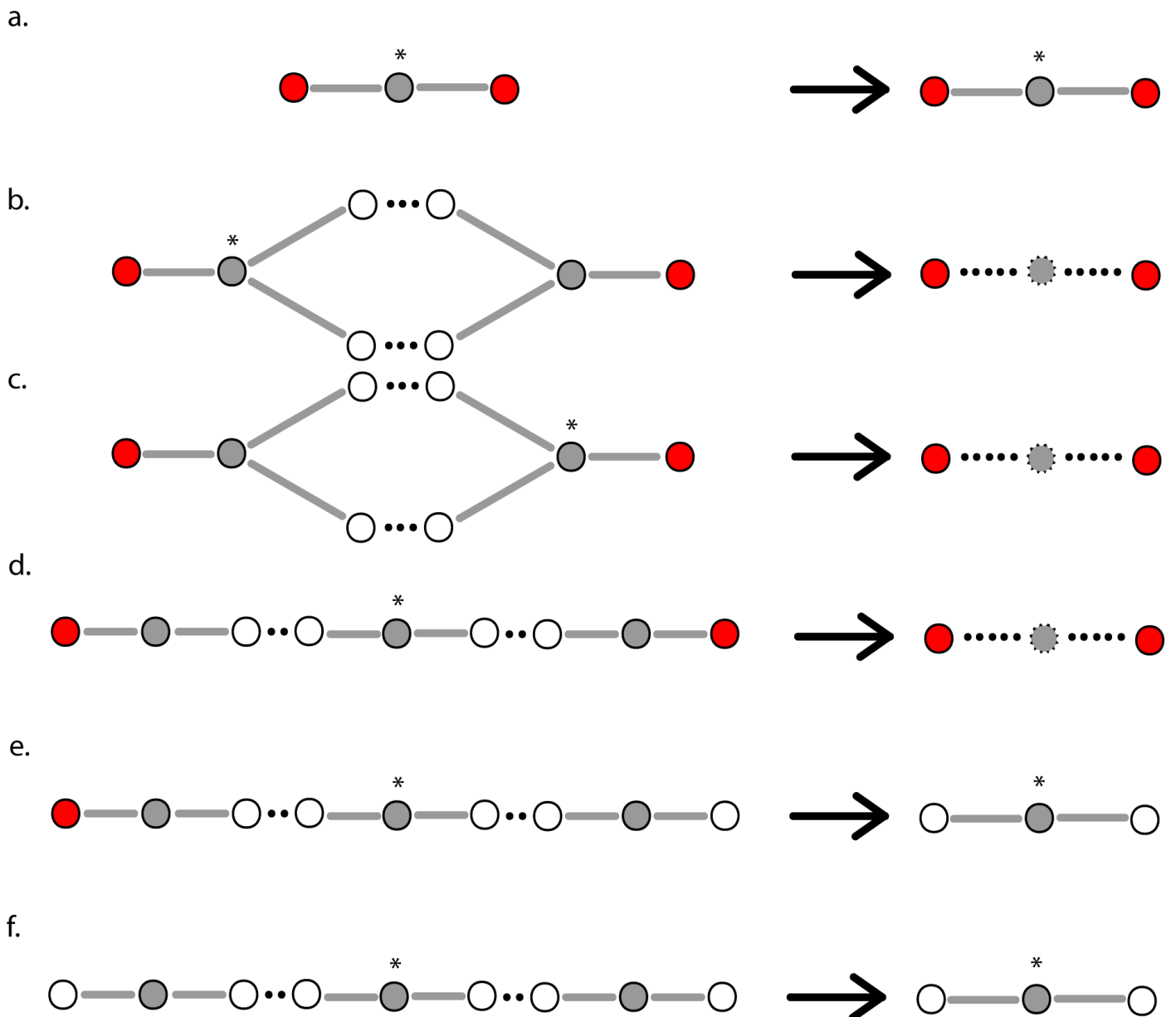
Figure 6: Reaction node collapse schematic. (a) For reactions where at least one input and at least one output component contain a measured value from the user data, the reaction will be maintained as is. (b) Where an input of a reaction is known, but no output has a known value, Metaboverse will search for all neighboring reactions that contain identical inputs. If the neighboring reaction has a known output value, the two reactions will be merged into one pseudo-reaction. (c) Where an output of a reaction is known, but no input has a known value, Metaboverse will search for all neighboring reactions that contain identical outputs. If the neighboring reaction has a known input value, the two reactions will be merged into one pseudo-reaction. (d) For reactions with no known values, neighbor pairs that match the inputs and outputs of the considered reaction will be evaluated for whether their respective outputs and inputs both have known values. If values are known for both neighbors, the three reactions will be merged into one pseudo-reaction. (e) As in (d), but if one neighbor does not contain a value 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. An asterisk (*) indicates the target reaction being considered for a given reaction collapse. A red node indicates a reaction input or output with a measured value. A white node indicates a reaction input or output with no measured value. A grey node indicates a reaction. A grey node with a dashed border indicates a pseudo-reaction. A solid edge indicates a known relationship. A dashed edge indicates a relationship inferred via reaction merging.

4. Regulatory pattern (motif) searches and sorting

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

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

When the stamp view demonstrates draws reaction motifs, all the pathways containing the corresponding motif will be listed below the stamp view by clicking on the appropriate stamp. For pathway motifs, clicking on a pathway ID will draw the selected pathway as a network, with all the motifs in this pathway highlighted.

5. Nearest neighborhood searches and prioritization

To visualize all global connections, 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 a change of one entity might have across the global network, which consequently aids in bridging and identifying any motifs that may occur between canonically annotated pathways. These neighborhoods can be expanded to view multiple downstream reaction steps and their accompanying genes, proteins, and metabolites by modulating the appropriate user option in the app.

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

6. Perturbation connectivity

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

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

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

Interactive visualization is handled using the D3 and JQuery JavaScript libraries. Force-directed layouts of networks are constructed by taking an 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 plotted. Relevant metadata, such as user-provided data and reaction descriptions, can be accessed by the user in real-time. Metadata for categorical displays, such as edge or node type, are extracted from the metadata during visualization of the sub-network.

Some performance optimization features are included by default to prevent computational overload. For example, nearest neighbor sub-networks with more than 1500 nodes, or nodes with more than 500 edges will not be plotted as 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 their pathway of choice, and all component reactions and their substrates, products, modifiers, and metadata are queried from the global network. 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 specify 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

While Metaboverse will continue to undergo development and new features will be added, we will briefly highlight some additional features available at the time of publication. We encourage users to check the documentation for more current updates and information regarding the use of Metaboverse [25].

7.6a Toggle genes

As gene components can crowd the network visualization space, users 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

Users can switch between coloring nodes based on the value or statistic 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.

7.6c Toggling features/labels

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

7.6d Toggling collapsed reactions

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

7.6e View curated pathway image

While Metaboverse visualizes networks dynamically, 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

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

7.6g Nearest neighbor and hub thresholding

The number of nearest neighbors to 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 a node display threshold is reached. This allows the user to visualize downstream effects of a change that may propagate across several reactions. The hub threshold option prevents the plotting

of nodes with more than the specified number of edges. This is handled during the visualization process by excluding any entity nodes that meet these criteria as the neighborhood is propagated. This is particularly useful in removing hub nodes, such as water or protons, which may be less relevant to the user experience and can quickly clutter the 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 spaces where this information can be displayed, which is updated based on the user's input as it is provided.

8. Packaging

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

Table 1: Dependencies table.

Name	Reference
HTML	N/A
CSS	N/A
Javascript	N/A
Electron	[52]
JQuery	[53]
D3	[54]
string-pixel-width	[55]
d3-save-svg	[56]
Python	[57]
pandas	[58,59]
numpy	[60]
scipy	[61]
matplotlib	[62]
NetworkX	[51]

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 *mct1Δ* strain as described in [40].

RNA-sequencing sample preparation

RNA sequencing data were generated by growing *Saccharomyces cerevisiae* biological replicates for strains *mct1Δ* (n=4) and wild-type (n=4). Briefly, cells were grown in glucose and switched to raffinose-supplemented growth medium for 0, 3, and 12 hours such that at time of harvest, cultures were at OD_{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 [63]. Batch and log files are available at [64]. Notably, reads were trimmed of adapters (Read1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA, Read2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). Based on library complexity quality control, de-duplicated alignments were used for read quantification due to the high number of duplicated sequences in each library. Differential expression analysis was performed using DESeq2 [22] by comparing *mct1Δ* samples with wild-type samples at the 12-hour time-point to match the steady-state proteomics data. $\log_2$(fold change) and false discovery rate ("p-adj") values were extracted from the DESeq2 output.

Proteomics analysis

Steady-state quantitative proteomics data were generated as described in [40]. Briefly, cells were grown in glucose and switched to raffinose-supplemented growth medium overnight, and harvested at mid-log phase. For this analysis, we compared the *mct1Δ* (n=3) with the wild-type (n=3) cell populations. $\log_2$(fold change) values and Benjamini-Hochberg corrected p-values were generated by comparing *mct1Δ* with the wild-type cells. P-values were generated 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 minimal media (S-min) supplemented with 2% glucose until they reached OD_{600} between 0.6-0.8. Cells were then transferred to S-min media containing 2% raffinose and harvested after 0, 15, 30, 60, and 180 minutes (n=6/time-point/strain).

Metabolite extraction

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

Mass Spectrometry Analysis of Samples

All GC-MS analysis was performed with an Agilent 5977b GC-MS MSD-HES and an Agilent 7693A automatic liquid sampler. Dried samples were suspended in 40 µL of a 40 mg/mL O-methoxylamine hydrochloride (MOX) (MP Bio #155405) in dry pyridine (EMD Millipore #PX2012-7) and incubated for one hour at 37 °C in a sand bath. 25 µL of this solution was added to auto sampler vials. 60 µL of N-methyl-N-trimethylsilyltrifluoracetamide (MSTFA with 1%TMCS, Thermo #TS48913) was added automatically via the auto sampler and incubated for 30 minutes at 37 °C. After incubation, samples were vortexed and 1 µL of the prepared sample was injected into the gas chromatograph inlet in the split mode with the inlet temperature held at 250 °C. A 10:1 split ratio was used for analysis of the majority of metabolites. For those metabolites that saturated the instrument at the 10:1 split concentration, a split of 50:1 was used for analysis. The gas chromatograph had an initial temperature of 60 °C for one minute followed by a 10 °C/min ramp to 325 °C and a hold time of 5 minutes. A 30-meter Phenomenex Zebron AB-5HT with 5m inert Guardian capillary column was employed for chromatographic separation. Helium was used as the carrier gas at a rate of 1 mL/min.

Analysis of Mass Spectrometry Data

Data was collected using MassHunter software (Agilent). Metabolites were identified and their peak area was recorded using MassHunter Quant. This data was transferred to an Excel spread sheet (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.

9.2 Human lung adenocarcinoma metabolomics.

Data were accessed from Metabolomics Workbench [65] and processed as in our previous re-study of this data [26].

10. Data availability.

mct1Δ and accompanying wild-type transcriptomics time course data is deposited at the GEO database under identifier GSE151606 [66]. Metabolomics raw data is currently in submission to Metabolomics Workbench (https://www.metabolomicsworkbench.org). In the meantime, the raw data can be accessed at [64].

The curated networks for these data are available at [64]. 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 motifs and global connectivity analyses were performed within the Metaboverse platform.

The Metaboverse source code is available at [67] and archived versions can be accessed at Zenodo [].

The source code and data for this manuscript and subsequent analyses is available at [68] and archived versions can be accessed at Zenodo [].

Acknowledgements

Funding

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.

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. **amueller/word_cloud**
   GitHub
   https://github.com/amueller/word_cloud

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. **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

14. **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

15. **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

16. **Superfamilies of Evolved and Designed Networks**
R. Milo
Science (2004-03-05) https://doi.org/dm5h6p
DOI: 10.1126/science.1089167 · PMID: 15001784

17. **Network motifs in the transcriptional regulation network of Escherichia coli**
Shai S. Shen-Orr, Ron Milo, Shmoolik Mangan, Uri Alon
Nature Genetics (2002-04-22) https://doi.org/dr6vnt
DOI: 10.1038/ng881 · PMID: 11967538

18. **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

19. **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

20. **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

21. **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

22. **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

23. **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

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. **Metaboverse — Metaboverse 0.0.0-alpha documentation** https://metaboverse.readthedocs.io/en/latest/

26. **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

27. **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

28. **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

29. **Network biology concepts in complex disease comorbidities**
Jessica Xin Hu, Cecilia Engel Thomas, Søren Brunak
Nature Reviews Genetics (2016-08-08) https://doi.org/f84t9c
DOI: 10.1038/nrg.2016.87 · PMID: 27498692

30. **Using graph theory to analyze biological networks**
Georgios A Pavlopoulos, Maria Secrier, Charalampos N Moschopoulos, Theodoros G Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider, Pantelis G Bagos
BioData Mining (2011-04-28) https://doi.org/bggzw9
DOI: 10.1186/1756-0381-4-10 · PMID: 21527005 · PMCID: PMC3101653

31. **Functional and transcriptional connectivity of communities in breast cancer co-expression networks**
Guillermo de Anda-Jáuregui, Sergio Antonio Alcalá-Corona, Jesús Espinal-Enríquez, Enrique Hernández-Lemus
Applied Network Science (2019-05-20) https://doi.org/ggt7sj
DOI: 10.1007/s41109-019-0129-0

32. **Loss of Connectivity in Cancer Co-Expression Networks**
Roberto Anglani, Teresa M. Creanza, Vania C. Liuzzi, Ada Piepoli, Anna Panza, Angelo Andriulli, Nicola Ancona
PLoS ONE (2014-01-28) https://doi.org/f5vhd4
DOI: 10.1371/journal.pone.0087075 · PMID: 24489837 · PMCID: PMC3904972

33. **A Mitochondrial Pyruvate Carrier Required for Pyruvate Uptake in Yeast, Drosophila, and Humans**
Daniel K. Bricker, Eric B. Taylor, John C. Schell, Thomas Orsak, Audrey Boutron, Yu-Chan Chen, James E. Cox, Caleb M. Cardon, Jonathan G. Van Vranken, Noah Dephoure, … Jared Rutter
Science (2012-05-24) https://doi.org/f33x88
DOI: 10.1126/science.1218099 · PMID: 22628558 · PMCID: PMC3690818

34. **Control of intestinal stem cell function and proliferation by mitochondrial pyruvate metabolism**
John C. Schell, Dona R. Wisidagama, Claire Bensard, Helong Zhao, Peng Wei, Jason Tanner, Aimee Flores, Jeffrey Mohlman, Lise K. Sorensen, Christian S. Earl, … Jared Rutter
Nature Cell Biology (2017-08-14) https://doi.org/gbrf73
DOI: 10.1038/ncb3593 · PMID: 28812582 · PMCID: PMC6137334

35. **Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier**
Claire L. Bensard, Dona R. Wisidagama, Kristofor A. Olson, Jordan A. Berg, Nathan M. Krah, John C.

Schell, Sara M. Nowinski, Sarah Fogarty, Alex J. Bott, Peng Wei, … Jared Rutter
Cell Metabolism (2020-02) https://doi.org/ggt7w8
DOI: 10.1016/j.cmet.2019.11.002 · PMID: 31813825 · PMCID: PMC7004878

36. **Redundancy: A Critical Obstacle to Improving Cancer Therapy**
O. Lavi
Cancer Research (2015-01-09) https://doi.org/ggt7xd
DOI: 10.1158/0008-5472.can-14-3256 · PMID: 25576083 · PMCID: PMC6250436

37. **The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease**
J. Lamb
Science (2006-09-29) https://doi.org/c92ptt
DOI: 10.1126/science.1132939 · PMID: 17008526

38. **Pathway crosstalk perturbation network modeling for identification of connectivity changes induced by diabetic neuropathy and pioglitazone**
Guillermo de Anda-Jáuregui, Kai Guo, Brett A. McGregor, Eva L. Feldman, Junguk Hur
BMC Systems Biology (2019-01-07) https://doi.org/ggt7zx
DOI: 10.1186/s12918-018-0674-7 · PMID: 30616626 · PMCID: PMC6322225

39. **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

40. **ACP Acylation Is an Acetyl-CoA-Dependent Modification Required for Electron Transport Chain Assembly**
Jonathan G. Van Vranken, Sara M. Nowinski, Katie J. Clowers, Mi-Young Jeong, Yeyun Ouyang, Jordan A. Berg, Jeremy P. Gygi, Steven P. Gygi, Dennis R. Winge, Jared Rutter
Molecular Cell (2018-08) https://doi.org/gd5dhm
DOI: 10.1016/j.molcel.2018.06.039 · PMID: 30118679 · PMCID: PMC6104058

41. **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

42. **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

43. **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

44. **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

45. **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

46. **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

47. **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

48. **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

49. **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

50. **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

51. (2014-01-28) http://conference.scipy.org/proceedings/SciPy2008/paper_2/full_text.pdf

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

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

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

55. **adambisek/string-pixel-width**
    GitHub
    https://github.com/adambisek/string-pixel-width

56. **edeno/d3-save-svg**
    GitHub
    https://github.com/edeno/d3-save-svg

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

58. **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

59. **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

60. (2007-01-16) https://web.mit.edu/dvp/Public/numpybook.pdf

61. **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

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

63. **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

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

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

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

67. **Metaboverse/Metaboverse**
GitHub
https://github.com/Metaboverse/Metaboverse

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