

Multiomics data integration and pathway mapping

MANA 2023 Conference Workshop

10.24.2023

Presenters:

İşin Tuna Sakallioglu

Stephanie Bishop

Pawan Sandhu

Sara J. Gosline

Contributors:

Kim-Ahn Le Cao

Ewy Mathé



Purpose of this workshop

- Introduce multi-omics techniques to the metabolomics community.
- Explain how multi-omics is helpful in the metabolomics research.
- Introduce available tools to the community.

Useful resources for our tutorials:

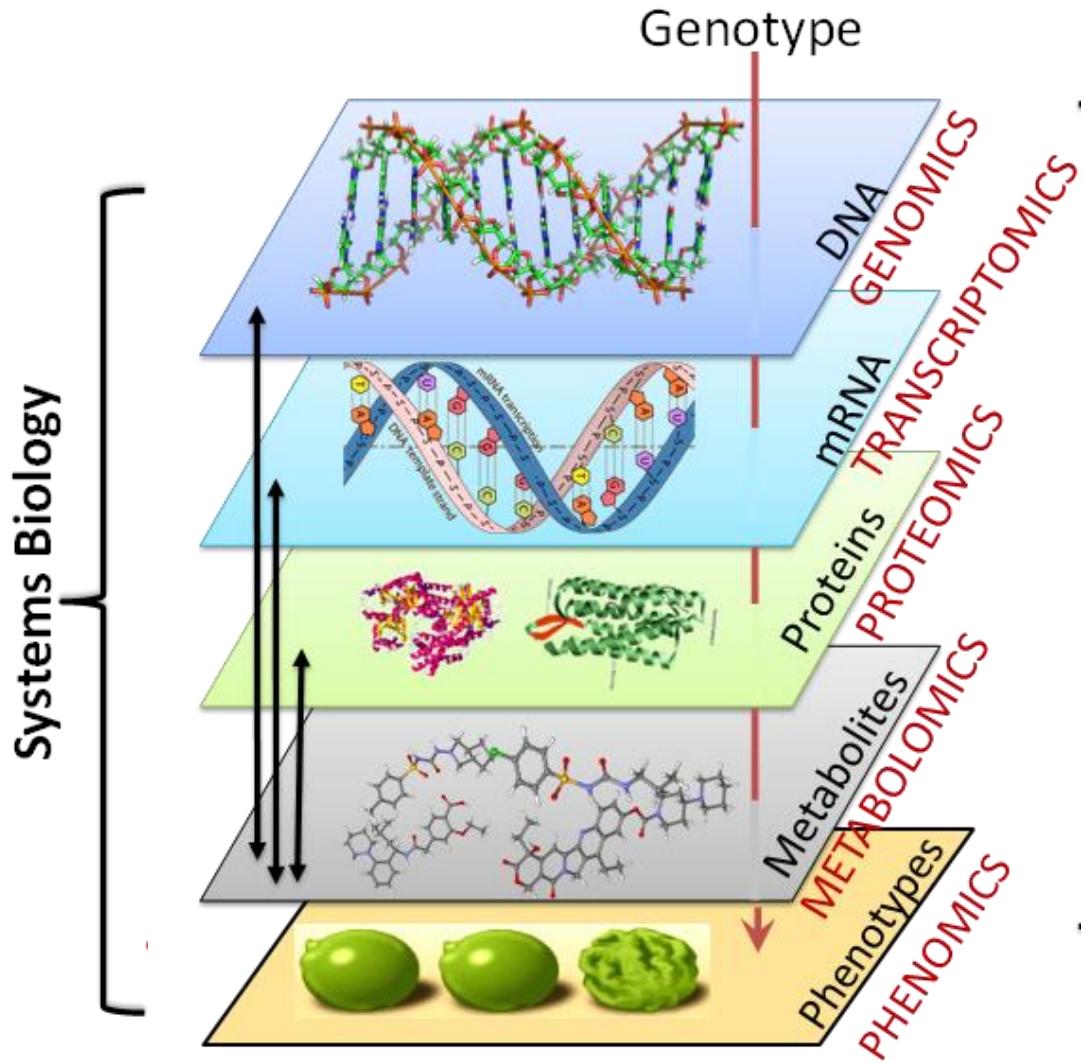
<https://github.com/sgosline/metabolomics-integration/tree/main>



Workshop outline

Time	Topic	Presenters
1:00 – 1:10 (10 mins)	Statistical approaches background	İşin Tuna Sakallioglu
1:10 – 1:40 (30 mins)	Multi-omics tools for non-computational scientists background and tutorial <ul style="list-style-type: none">• PaintOmics• MINNO	Pawan Sandhu Stephanie Bishop
1:40 – 1:50 (10 mins)	Break and download R software for next tutorial	
1:50 – 2:20 (30 mins)	Multi-omics tools for correlation and network analysis	Sara Gosline
2:20 – 2:45 (25 mins)	Group discussion – best approaches for multi-omics workflows	İşin Tuna Sakallioglu

Multi-omics: Background

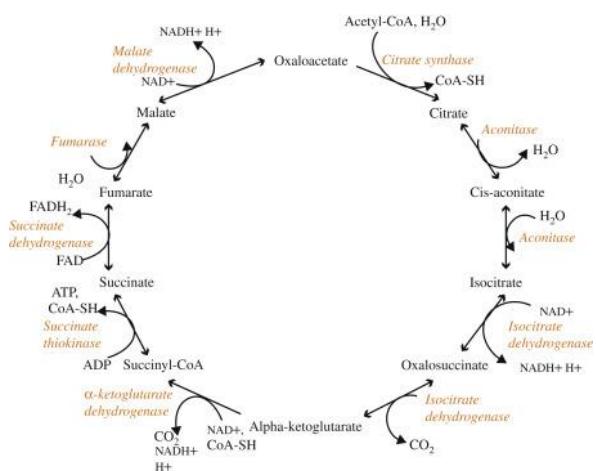
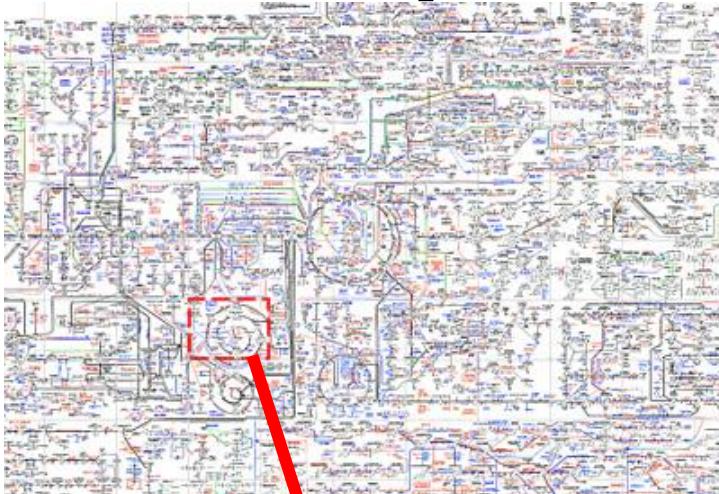


- Keywords: integrated omics, pan-omics, and trans-omics.
- An approach aiming to improve the understanding of
 - ✓ systems regulatory biology,
 - ✓ molecular central dogma and
 - ✓ genotype-phenotype relationship.
- Aims to combine two or more omics data sets to aid in data analysis, visualization and interpretation to determine the mechanism of a biological process.

So far it is being used:

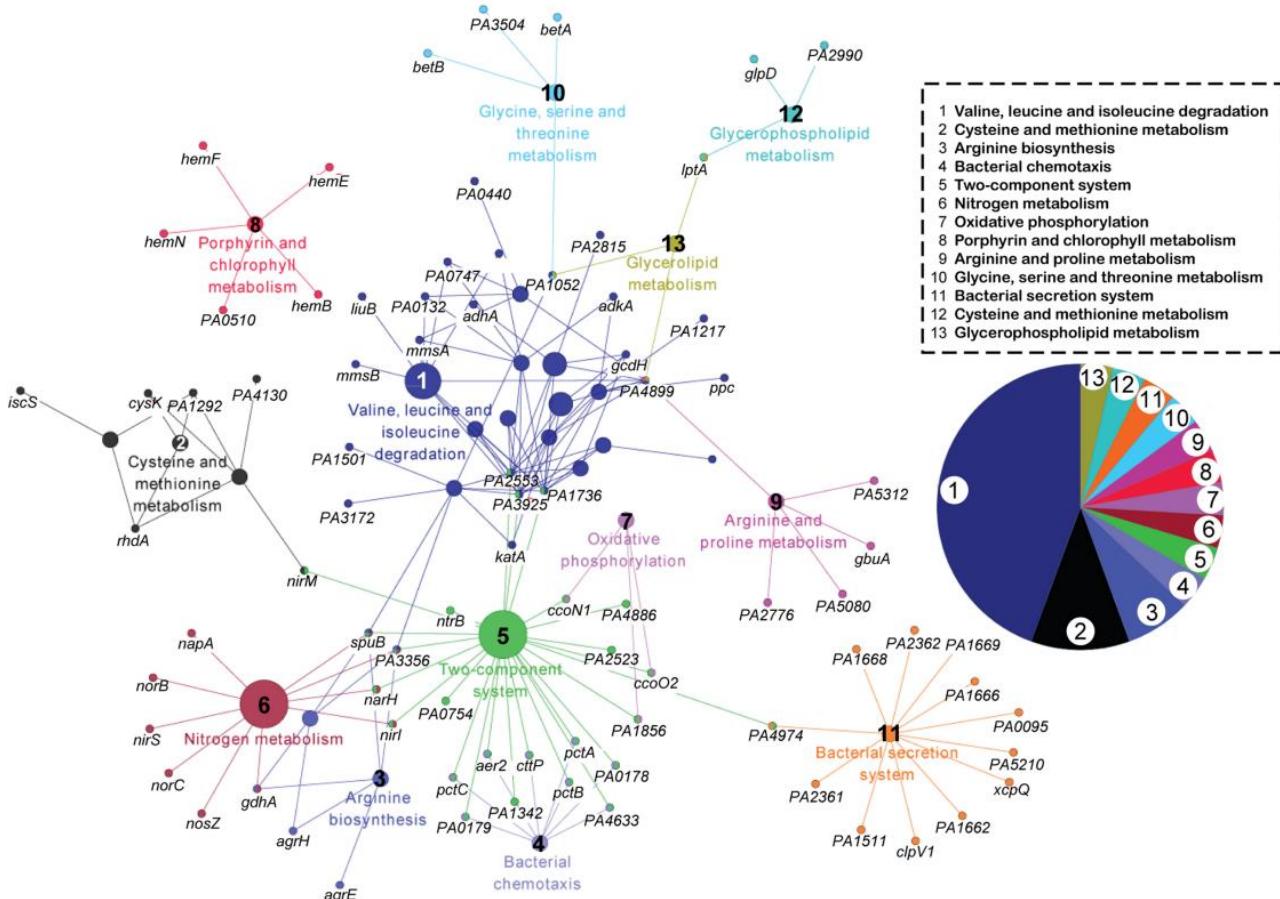
- ✓ To increase the diagnostic yield for analysis of global regulome,
- ✓ To provide insights into drug targets and
- ✓ To determine specific biological markers for a disease.

Metabolic pathways



- The metabolic pathways are biochemical pathways that are composed of enzymes, active proteins, metabolites and co-factors.
- One metabolite or protein can have more than one interconnections with other pathways which makes 3D network maps.

Complexity of network maps of single-omics data



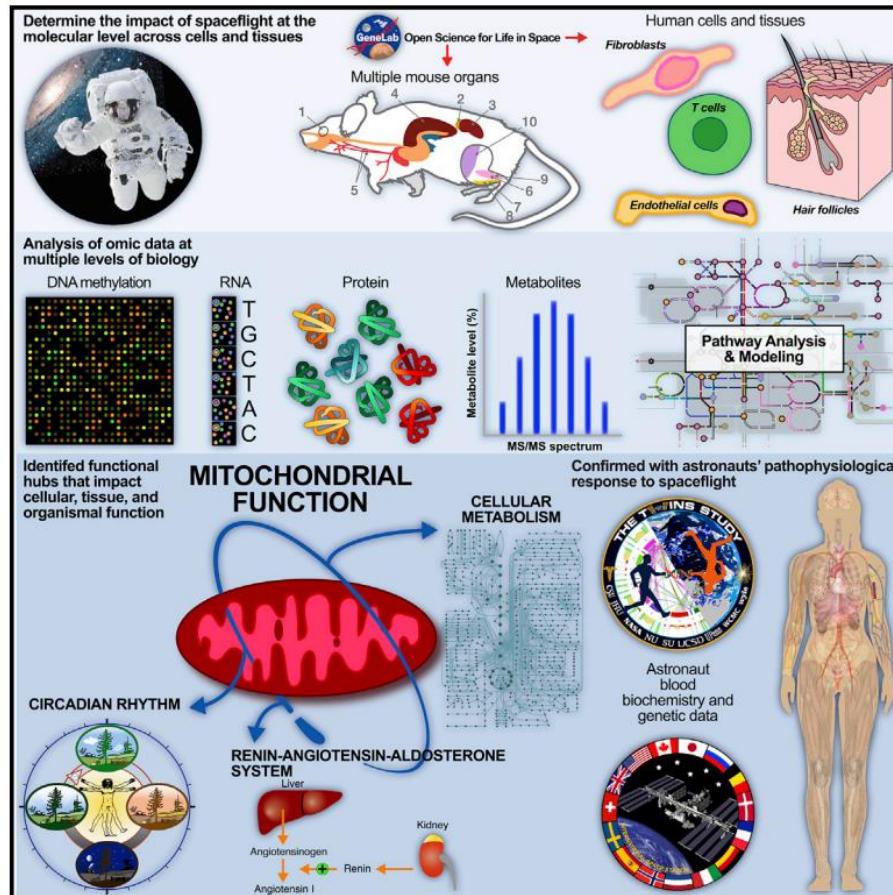
- Kegg pathway network analysis of downregulated genes from the proteomics data.

- The interpretation of these network data is one of the bottlenecks of Omics.
- So many unknowns in these pathways and networks.
- Cannot identify every single association of metabolite -metabolite or metabolites-proteins.
- These needs traditional enzyme activity assay approach.
- Now functional Omics, AI approaches, multi-omics data integrations overcoming these issues.

Literature multi-omics examples

Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact

Graphical Abstract



Authors

Willian A. da Silveira, Hossein Fazelinia, Sara Brin Rosenthal, ..., Christopher E. Mason, Sylvain V. Costes, Afshin Beheshti

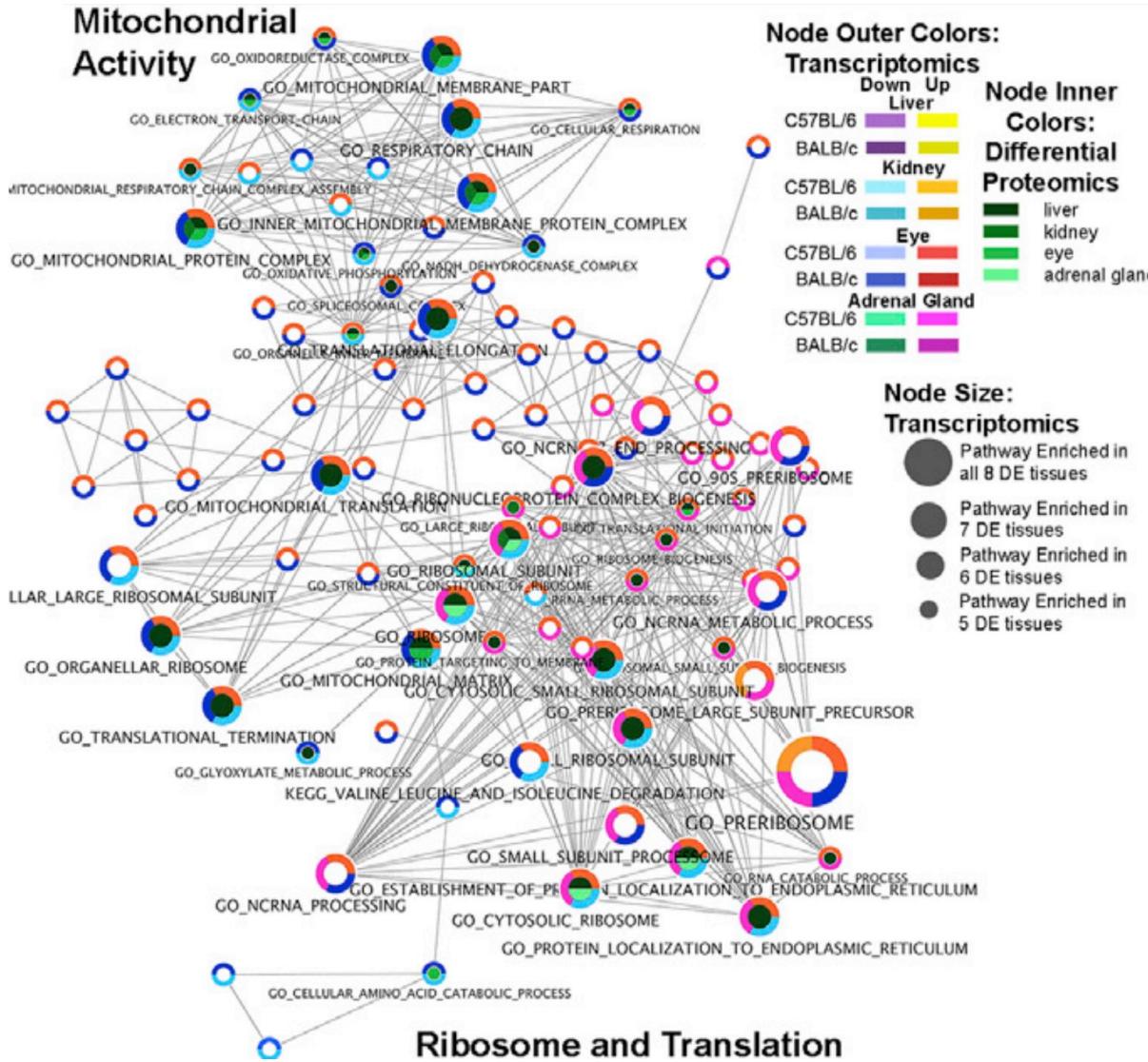
Correspondence

afshin.beheshti@nasa.gov

In Brief

A comprehensive multi-omics analysis from 59 astronauts and hundreds of samples flown in space provides insight into fundamental biological mechanisms affected by spaceflight and highlights mitochondrial dysregulation as a central hub for space biology.

Complexity of network maps of multi-omics data



To understand the system wide effect:

- Pathway results were integrated across multiple tissues and RNAseq and proteomics platforms.
- Enriched GO terms and KEGG pathways were integrated using a network framework.
- Complexity increase as you integrate Omics data sets
- Goal of the multiomics researcher is to harmonize this complexity to present it.

Multi-omics Investigation into the Mechanism of Action of an Anti-tubercular Fatty Acid Analogue

Isin T. Sakallioglu, Amith S. Maroli, Aline De Lima Leite, Darrell D. Marshall, Boone W. Evans, Denise K. Zinniel, Patrick H. Dussault, Raúl G. Barletta, and Robert Powers*

 Cite This: *J. Am. Chem. Soc.* 2022, 144, 21157–21173

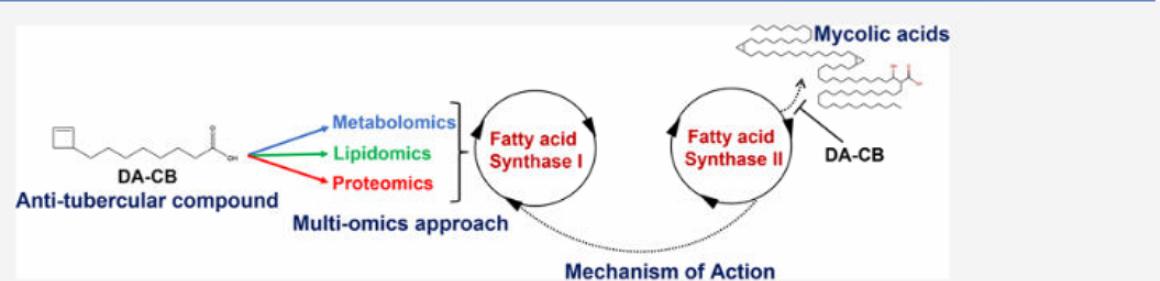
 Read Online

ACCESS |

 Metrics & More

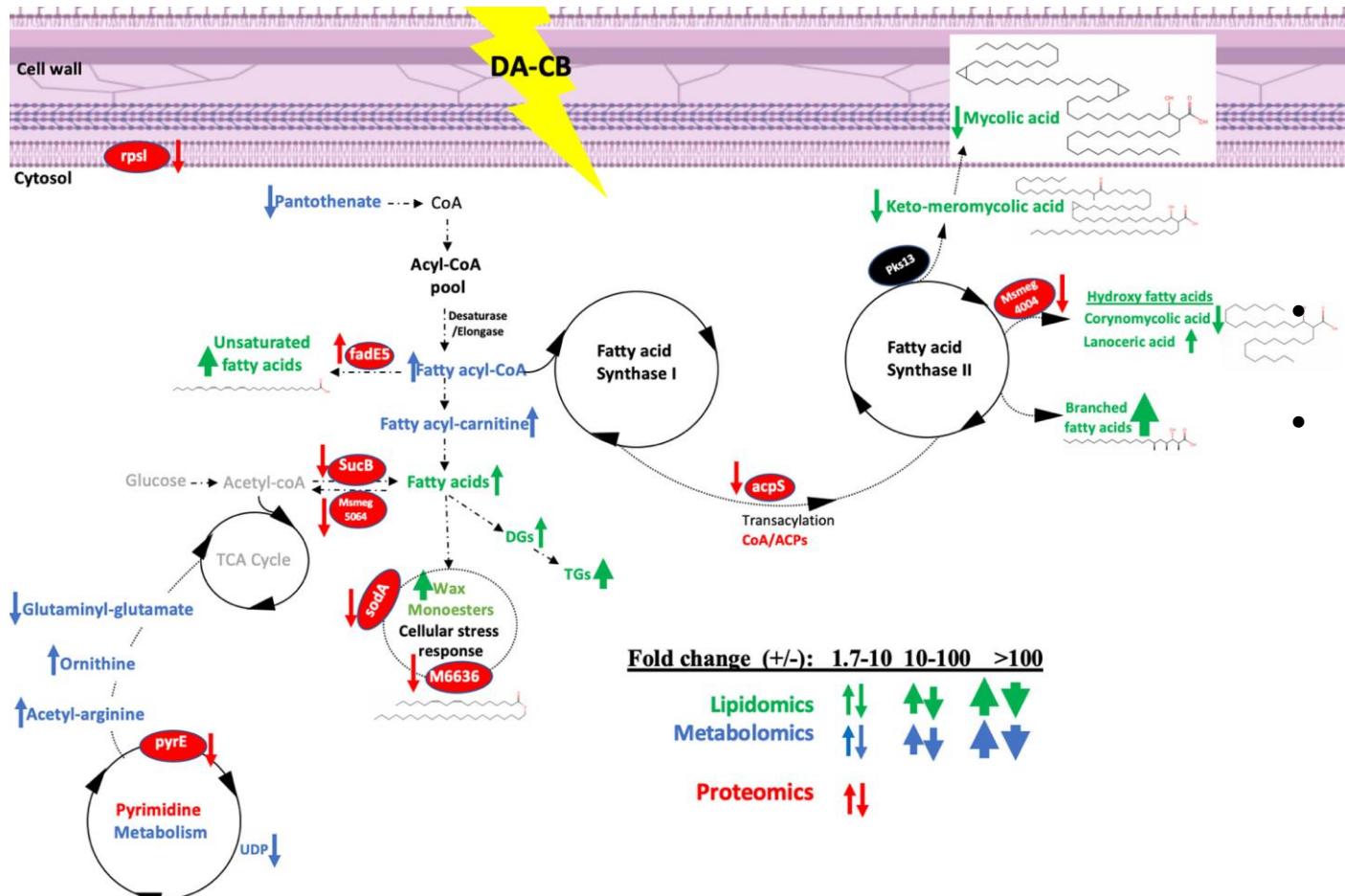
 Article Recommendations

 Supporting Information



- Understand the mechanism of action of a drug-like molecule,
- Same sample, lipid-metabolite-protein content.
- Separately processed without data concatenation.
- One of the major bottleneck of this work was the lack of database for *Msm.* that would accept multiomics data IDs.

Multiomics illuminated the broader regulatory biomolecules that changed during drug perturbation



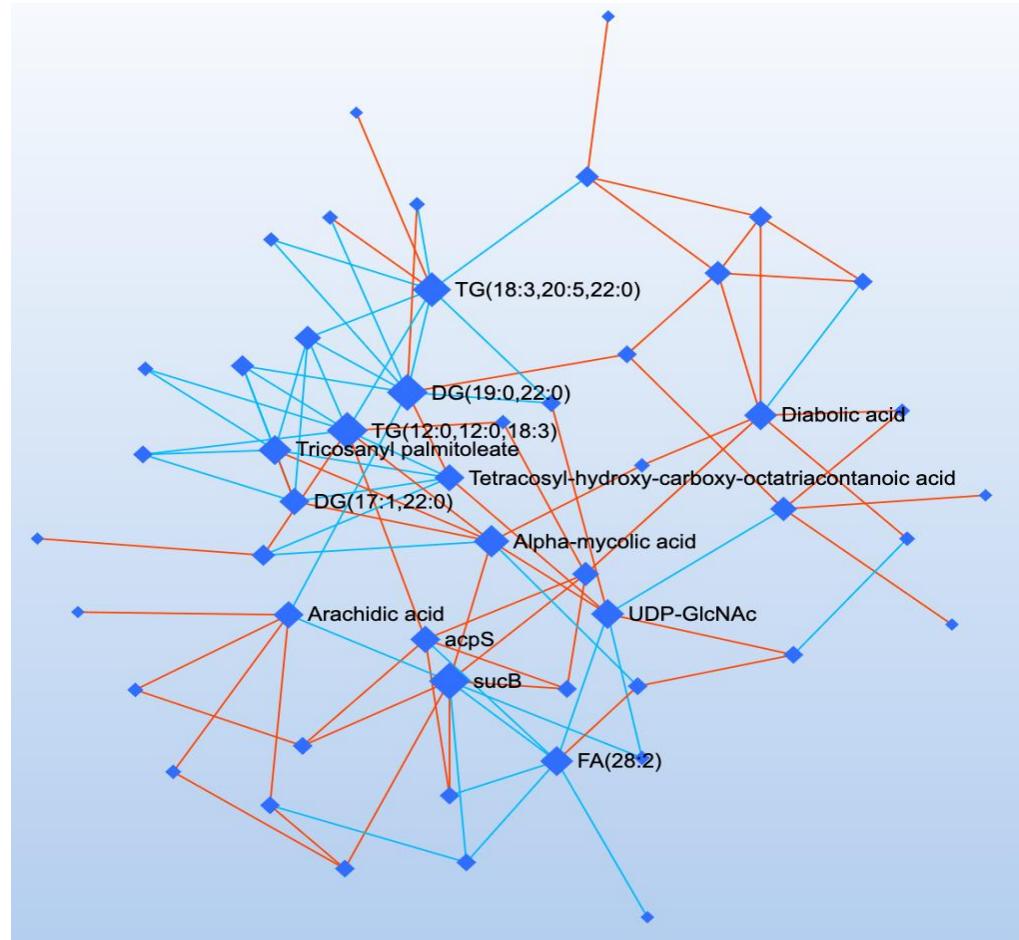
“This wider span of coverage enabled the construction of a unified and consensus network.

- This permitted elimination of off-target or secondary perturbations that commonly confound the analysis of the single-omics data.”

- Literature, KEGG database, manual pathway analysis lead to generation of the network.
- The resulting multi-omics network showed a specific pathway being regulated.

Correlation analysis provides support to the consensus integrated network

- How statistics helps to understand accuracy of the consensus network.
- Provide the evidence to the consensus network.



- A hierarchical clustered heatmap, summarizes the entire set of pairwise correlation coefficient.
 - Showing a strong positive or negative correlation between all the significantly altered metabolites, lipids, and proteins.
- A network map based on the Spearman's rank correlation coefficient for the metabolites, lipids, and proteins on consensus network is shown.
 - Demonstrating a unified cellular response to the treatment and the accuracy of the consensus integrated network.

Multi-block multi-omics approach

Multiomics Approach Captures Hepatic Metabolic Network Altered by Chronic Ethanol Administration

by  **Isin Tuna Sakallioglu** ¹,  **Bridget Tripp** ^{2,3} ,  **Jacy Kubik** ^{4,5},  **Carol A. Casey** ^{4,5} ,
 **Paul Thomes** ^{4,5,*}  and  **Robert Powers** ^{1,3,*}  

¹ Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, USA

² Center for Biotechnology, University of Nebraska-Lincoln, Lincoln, NE 68588-0665, USA

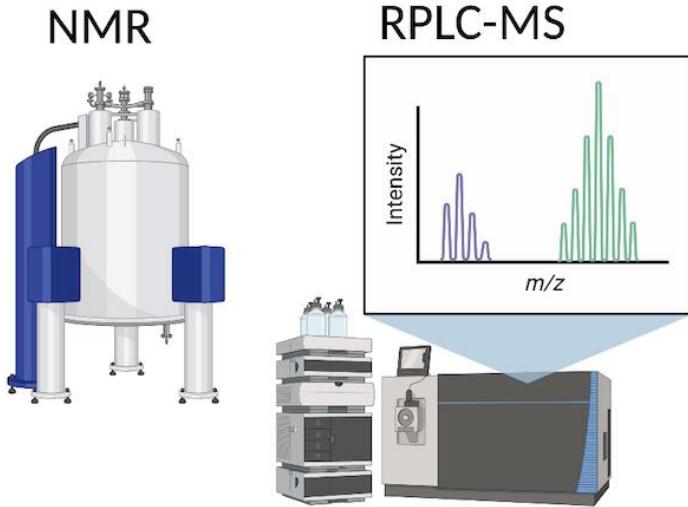
³ Nebraska Center for Integrated Biomolecular Communication, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, USA

⁴ Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE 68198-5870, USA

⁵ Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, NE 68198-5870, USA

* Authors to whom correspondence should be addressed.

Biology **2023**, *12*(1), 28; <https://doi.org/10.3390/biology12010028>



- “This study highlights the beneficial impact of a combined multiplatform, multiomics approach in achieving a system-wide characterization of EtOH-induced hepatocellular changes.”

Tools available for multi-omics integration and visualization

State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing

Michał Krassowski^{1†}, Vivek Das^{2†}, Sangram K. Sahu^{3†} and Biswapriya B. Misra^{4*†}

¹ Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, United Kingdom, ² Novo Nordisk Research Center Seattle, Inc, Seattle, WA, United States, ³ Independent Researcher, Bengaluru, India, ⁴ Independent Researcher, Namburu, India

Journal of Molecular
Endocrinology

B B Misra *et al.*

Approaches and tools in
integrated omics

62:1

R21–R45

REVIEW

Integrated omics: tools, advances and future approaches

Biswapriya B Misra¹, Carl Langefeld^{1,2}, Michael Olivier¹ and Laura A Cox^{1,3}

¹Center for Precision Medicine, Section on Molecular Medicine, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North California, USA

²Department of Biostatistics, Wake Forest School of Medicine, Winston-Salem, North California, USA

³Southwest National Primate Research Center, Texas Biomedical Research Institute, San Antonio, Texas, USA

Correspondence should be addressed to L A Cox: laurocox@wakehealth.edu

List of various tools, software statistical approaches and databases available for integrated Omics approach

Tool	Computational platform	User friendliness	Purpose
IntegrOmics	R	Difficult	Efficiently performs integrative analyses of two types of 'omics' variables that are measured on the same samples
SteinerNet	R	Moderate	Integrating transcriptional, proteomic and interactome data by searching for the solution to the prize-collecting Steiner tree problem
Omics Integrator	Python	Easy	Integrate proteomic data, gene expression data and/or epigenetic data using a protein–protein interaction network
MixOmics	R	Difficult	Provides a wide range of linear multivariate methods for data exploration, integration, dimension reduction and visualization of biological data sets
Mergeomics	R	Difficult	Genetic association (e.g., GWAS or exome sequencing), transcriptome-wide association (e.g., TWAS from microarray or RNA sequencing studies), and epigenetic association (e.g., EWAS from methylome association studies), functional genomics (such as eQTLs and ENCODE annotations), biological pathways, and gene networks

Tool	Computational platform	User friendliness	Purpose
XCMS Online	Web	Easy	Systems biology scale workflow, that allows rapid metabolic pathway mapping from raw metabolomics data to integration of genomic and proteomics data for mechanistic insights
Paintomics	Web	Easy	Integrated visual analysis of transcriptomics and metabolomics data
IMPaLA	Web	Easy	Joint pathway analysis of transcriptomics or proteomics and metabolomics data that also performs over-representation or enrichment analysis
3Omics	Web	Easy	Integrating multiple inter- or intra-transcriptomic, proteomic, and metabolomic human data
Ingenuity Pathway Analysis, Qiagen License	Web, Licence	Easy	Integration and mapping of genomics, transcriptomics, proteomics, and metabolomics datasets

A list of various resources for supporting FAIR* and interactive multi-omics study.

Popular multi-omics tools	Purpose
mixOmics	A tool with a framework that provides wide range of multivariate statistical methods for exploratory data analysis. This involves feature identification, extraction, selection.
Paintomics	A web-based systems biology tool for multi-omics integration and visualization across multi-species.
3Omics	A web-based application for integration and analysis of multi-omics data.
Data sharing	Purpose
OmicsDI	An aggregated database facilitating the discovery of heterogeneous published omics datasets across studies.
Zenodo	A general-purpose open-access data, softwares etc repository that allows user to obtain a citable DOI.
Code sharing	Purpose
Github	A version controlled code sharing and collaborative platform

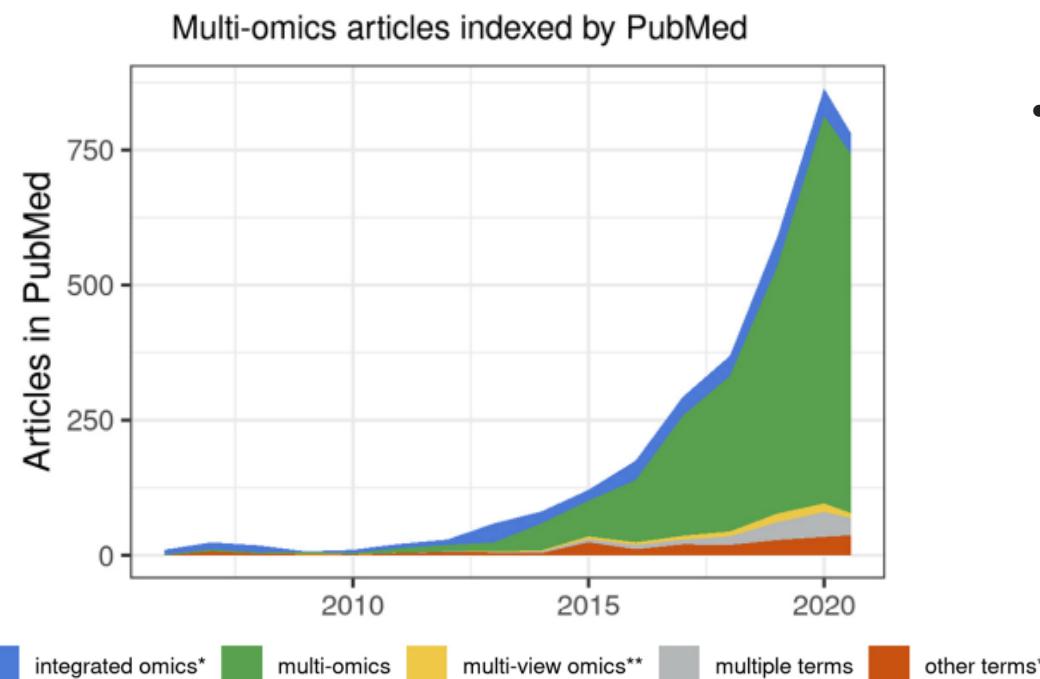
FAIR*: findability, accessibility, interoperability, and reusability.

Krassowski et al, 2020, *Frontier in Genetics*

Workflow sharing	Purpose
Common workflow	An open standard for describing analysis workflows which makes them portable and scalable across a variety of software and hardware environments.
Nextflow	An enterprise level workflow language for writing scalable and reproducible scientific pipelines
Environment sharing	Purpose
Conda	A package manager and computation environment management system
Biocontainers	A community driven project that provides docker based containerized bioinformatics software
Data visualization tools	Purpose
Shiny	A framework in R for doing GUI based
bokeh	A python library for interactive data visualization in browser
Cytoscape	A platform for network data integration, analysis and visualization

Future directions in multi-omics

- More data, more complexity, heterogeneity, more analysis!
- But more data, more omics brings accuracy to that approach!
- Data driven decisions and accelerated translational science is possible by developing new computational method/data analysis platforms.
- Disease specific multi-omics databases.
- Knowledge dependent community:
This is possible by data sharing and data repositories.



- Simply combining multiple omics techniques may uncover new biological insights that are not likely to be accomplished with a single approach.

Multi-omics tools for non-computational scientists



PaintOomics 4

<https://www.paintomics.org/>

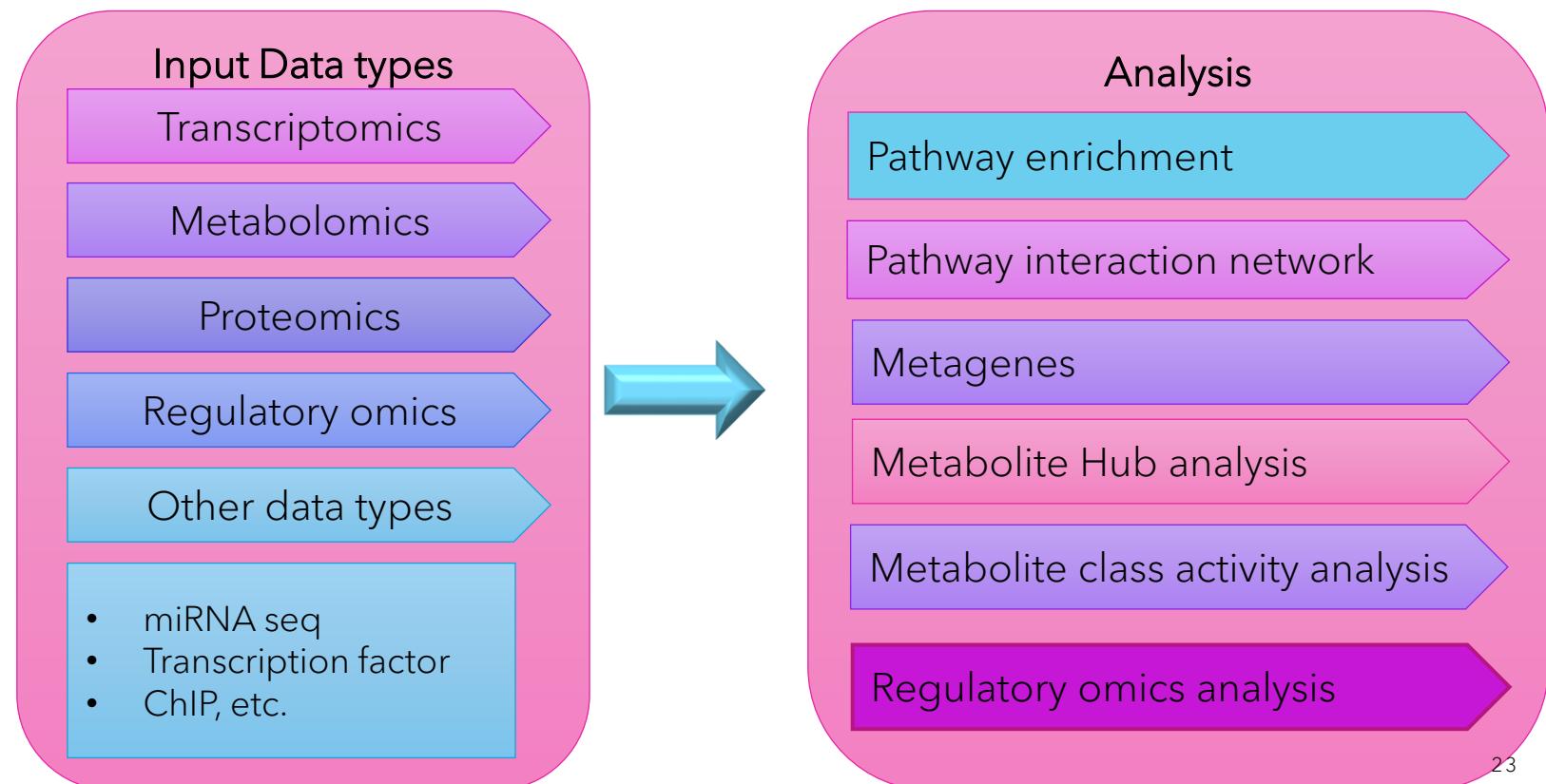
Liu et al., 2022, Paintomics 4, *Nucleic Acids Research*



Dr. Ana Conesa

Background

- Paintomics 4 is a web server for the integrative analysis and visualization of multi-omics datasets using biological pathway maps including KEGG, Reactome and MapMan.
- User-friendly.
- Requires no background in coding.
- Supports >100 organisms including plants, microbes, animals.
- Supports any omics type.



Background

- Create an account to save jobs, access previous jobs.
- Get personal storage.

<https://www.paintomics.org/>

PaintOmics 4 v1.00

Sign In

Email Address:

Password:

→ Sign in

Forgot your password?

No login

Continue using Paintomics without using accounts.

Please remember to save the provided URL to recover your job.

→ Continue without account

Want your own space? [Sign up now.](#)

Job view

Personal storage

Supporting tools

Resources

Publications

Contact

Sign in/Log in

Load example

Run PaintOmics

Reset

Data uploading

1. Organism selection

Organism: Please choose an organism

Not your organism? Request new organisms [clicking here.](#)

Enter a job description:

Databases:

KEGG (required)

MapMan

Reactome

For some species more than one database might be available. Please check [Supported ID and databases](#). Choose which ones you want to include in the analysis.

Download example data

2. Choose the files to upload

Available omics

- + Proteomics
- + Regulatory omic
- + Region based omic
- + Other omics

Selected omics

Gene expression

Data file: Browse... [?](#)

Relevant features file: Browse... [?](#)

Metabolomics

Data file: Browse... [?](#)

Relevant features file: Browse... [?](#)

Help

Drag and drop omics from "Available" to "Selected" area or click the + button.

If you do not need them, delete with [Delete](#).

Once you are done, click on the "Run PaintOmics" button on the upper-right corner.

Make sure to [choose an organism](#) from the select box first!

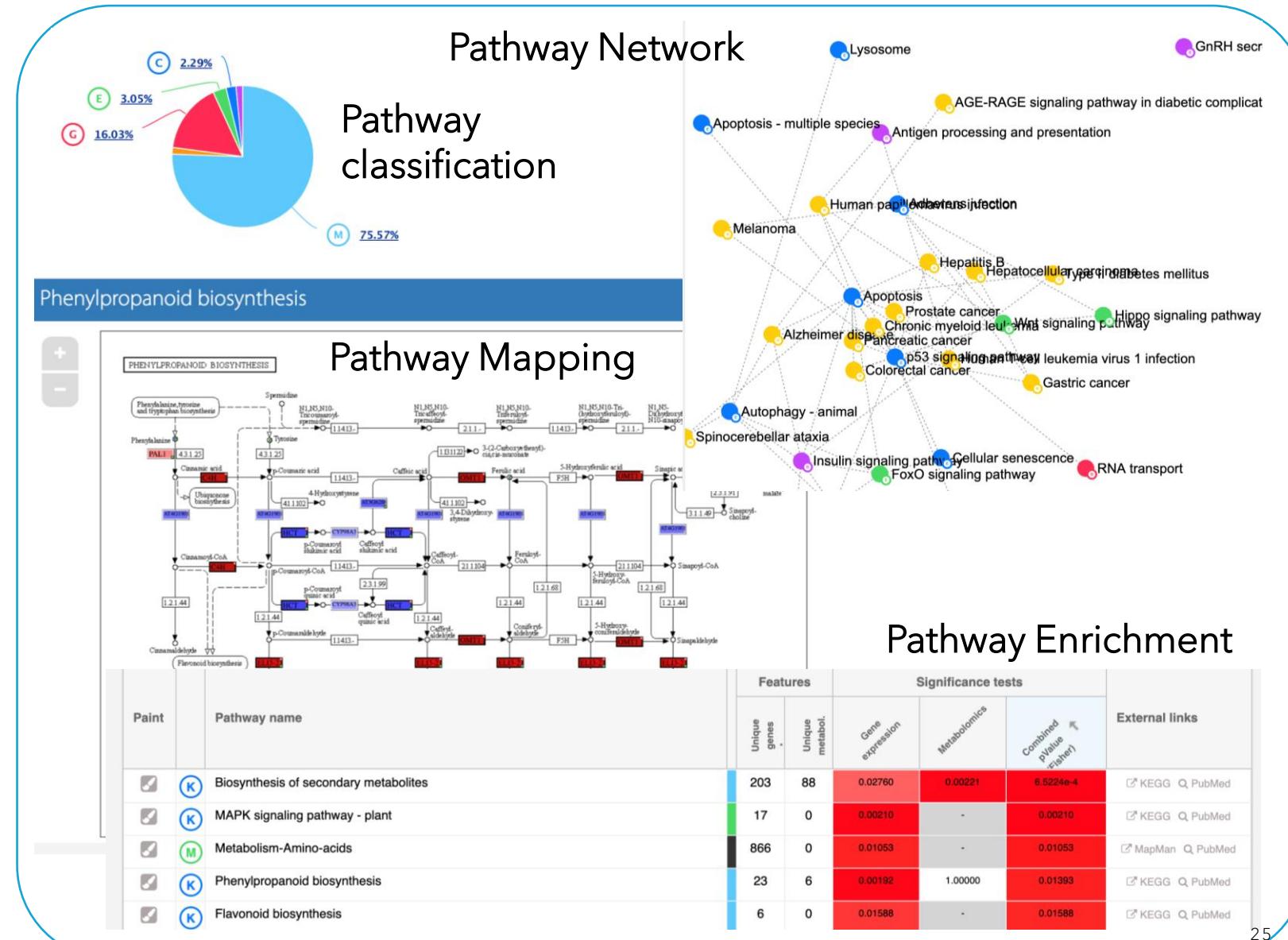
Background

Visualization

Input

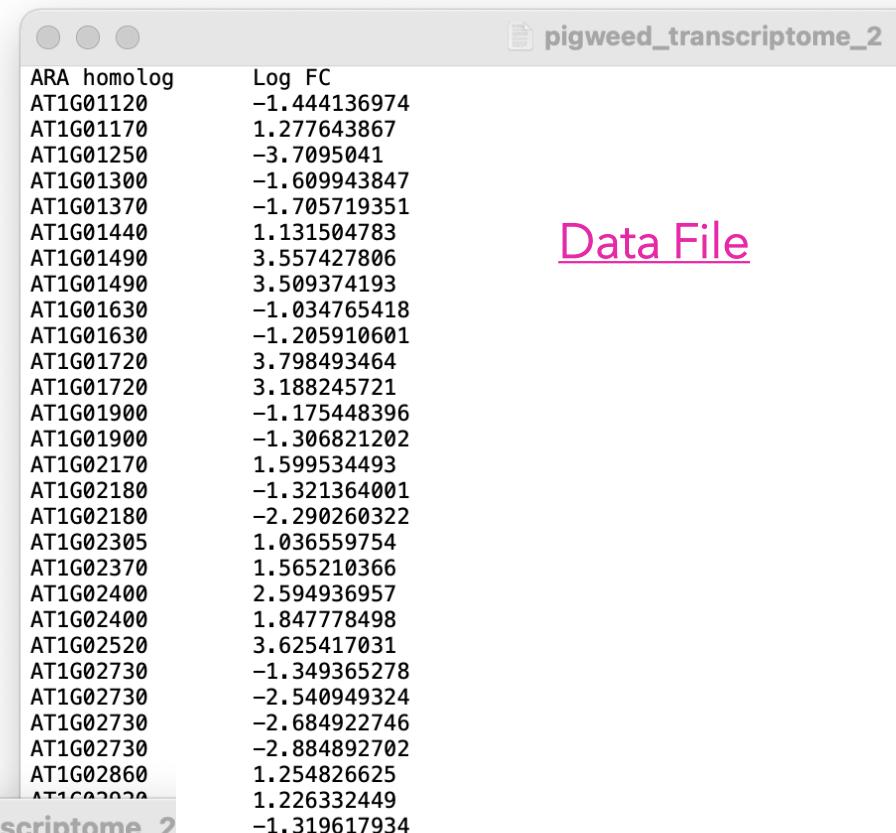
- **Format:** Tab-delimited text.
 - **Data file:** List of annotated features in each omics dataset with unique identifiers and quantification values.
 - **Relevant Features File:** List of significant features.

ANALYSIS



Step-by-step instructions - Step 1

- Format the data in tab-delimited text files.
- Requires Entrez IDs.
- Two files:
 - Data File: Two columns with features and **quantification values**.
 - Relevant Features File: List of features **statistically significant**.



ARA homolog	Log FC
AT1G01120	-1.444136974
AT1G01170	1.277643867
AT1G01250	-3.7095041
AT1G01300	-1.609943847
AT1G01370	-1.705719351
AT1G01440	1.131504783
AT1G01490	3.557427806
AT1G01490	3.509374193
AT1G01630	-1.034765418
AT1G01630	-1.205910601
AT1G01720	3.798493464
AT1G01720	3.188245721
AT1G01900	-1.175448396
AT1G01900	-1.306821202
AT1G02170	1.599534493
AT1G02180	-1.321364001
AT1G02180	-2.290260322
AT1G02305	1.036559754
AT1G02370	1.565210366
AT1G02400	2.594936957
AT1G02400	1.847778498
AT1G02520	3.625417031
AT1G02730	-1.349365278
AT1G02730	-2.540949324
AT1G02730	-2.684922746
AT1G02730	-2.884892702
AT1G02860	1.254826625
AT1G03220	1.226332449
AT1G05060	-1.319617934

[Data File](#)



ARA homolog
AT1G01250
AT1G01490
AT1G01490
AT1G01720
AT1G01720
AT1G02180
AT1G02400
AT1G02520
AT1G02730
AT1G02730
AT1G02730
AT1G03220
AT1G03220
AT1G03670
AT1G03780
AT1G05060

[Relevant Features File](#)

Step-by-step instructions - Step 1

- Upload Data.
- Select Organism (Arabidopsis thaliana in this case).
- Enter Job Description (Optional): *MANA Multi-omics workshop*.
- Select MapMan.
- Click **► Run PaintOmics**

The screenshot shows the PaintOmics 4 interface. On the left, a sidebar lists 'Job view', 'Personal storage', 'Supporting tools', 'Resources', 'Publications', and 'Contact'. A purple box labeled 'Other Omics Options' is overlaid on the sidebar. The main area has a dark header with the PaintOmics logo, version 4.1.0, and 'Log out'. The top right has buttons for 'Load example', 'Run PaintOmics', and 'Reset'.
1. Organism selection: A blue box highlights the 'Organism' dropdown set to 'Arabidopsis thaliana (thale cress)'. Below it is a link: 'Not your organism? Request new organisms [clicking here](#)'.
Optional: A blue box highlights the 'Enter a job description' field containing 'MANA Multi-Omics Workshop'.
Databases: A pink box highlights checkboxes for 'KEGG (required)' (checked), 'Reactome' (unchecked), and 'MapMan' (checked).
2. Choose the files to upload:
- **Available omics** (purple box): 'Proteomics', 'Regulatory omic', 'Region based omic', 'Other omics'.
- **Selected omics** (pink box):

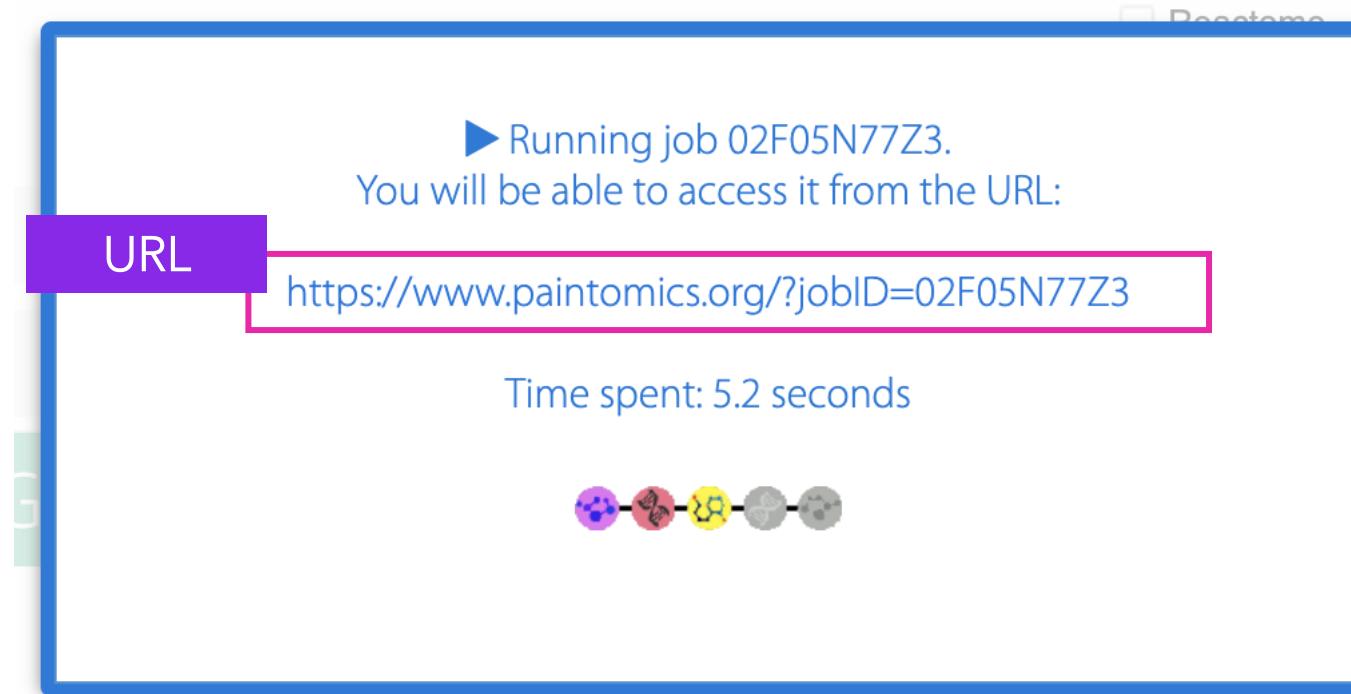
- Gene expression**: Data file: [MyData]/pigweed_transcriptome.txt, Relevant features file: [MyData]/relevant_transcriptome_2.txt. A context menu is open over the 'Upload file from my PC' button with options: 'Upload file from my PC', 'Use a file from My Data', and 'Clear selection'.
- Metabolomics**: Data file: [MyData]/pigweed_metabolome.txt, Relevant features file: [MyData]/relevant_metabolome.txt.

Upload Data by clicking on Browse

Help:
Drag and drop omics from "Available" to "Selected" area or click the **+** button.
If you do not need them, delete with **trash**.
Once you are done, click on the "Run PaintOmics" button on the upper-right corner.
Make sure to choose an organism from the select box first!

Step-by-step instructions - Step 1

- Running Job with time and URL to access the job later..



Step-by-step instructions - Step 2

- Data Summary Window on Feature ID mapping.
- Use of multiple datasets, and data distribution summary for each Omics

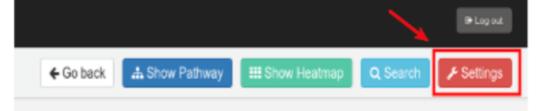
Feature ID/name translation summary

Below you will find an overview of the results after matching the input files against the PaintOmics databases. As a general rule, the bigger the percentage of mapped features, the better the results obtained in later stages. If the mapping percentage was low, manually check your results and input data.

[Download ID/Name mapping results.](#)

Data distribution summary

The following diagrams show the data distribution summary of each data set provided. By default the percentiles 10 and 90 will be taken as a reference for the colour scale when generating the heatmaps, but **in the next step you will be able to change the setting** using the "visual settings" button located in the toolbar, the one showed here.



Multiple databases used

The selected species had more than one database available. Your final analysis contains information about the following databases:

KEGG

[Kyoto Encyclopedia of Genes and Genomes](#) is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.

Omic	Matched
Gene expression	2304 (100%)
Metabolomics	undefined (NaN%)

MapMan

Oriented towards plant species, in combination with [GoMapMan](#), it provides additional pathways as well as an improved and more consolidated annotation for the model species Arabidopsis, and several crop species (potato, tomato, rice).

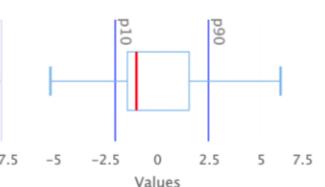
Omic	Matched

Gene expression

Mapped/Unmapped features



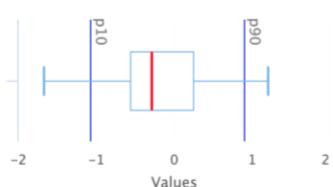
Data distribution



Metabolomics

See [Compounds disambiguation](#)

Data distribution



29

Step-by-step instructions - Step 2

- Compound Disambiguation.,
- Select number of Clusters for genes, and set metabolite activity thresholds.
- Click on the  to run the job.

Configure the number of clusters

In the next step Paintomics will calculate the clusters present in the data provided for each omic. It will use the method k-means using a automatically calculated number of cluster or the ones you define here. In the next step you will also be able to modify them by selecting individual omics in the network.

Gene expression: 

Configure the metabolite class activity threshold

To test the hypothesis of a metabolite class being regulated, Paintomics implements a metabolite class activity analysis tool, where a binomial test is used to assess the hypothesis of the proportion of significant compounds in a given measured metabolite class being higher than a user-defined threshold. In case the user does not define an activity threshold, Paintomics will use the average percentage of significant metabolites as threshold for the "Generate automatically".

Metabolite class activity threshold: 

PaintOmics 4 v1.00 

Job view   

Personal storage

Supporting tools

Resources

Publications

Contact

Compounds disambiguation

Some compounds names need to be disambiguated.

Please check the list below and choose the compounds in which you are interested.

2-Naphthylamine

1 compounds founds

2-Naphthylamine

2 alternative compounds founds 

3-dehydroshikimate

1 compounds founds

3-Dehydroshikimate

1 alternative compounds founds 

3-Methoxytyramine

1 compounds founds

3-Methoxytyramine

1 alternative compounds founds 

Aconitic acid

2 compounds founds

cis-Aconitic acid trans-Aconitic acid

Step-by-step instructions - Step 2

PaintOomics 4 v1.0.0

Log out

Go back Next step Reset

6 alternative compounds founds Show

beta-Tyrosine D-Tyrosine

39 alternative compounds founds Show

Urea

1 compounds founds

Urea

40 alternative compounds founds Show

▶ Running job 02F05N77Z3

D-Valine

Xylitol

1 compounds founds

Xylitol

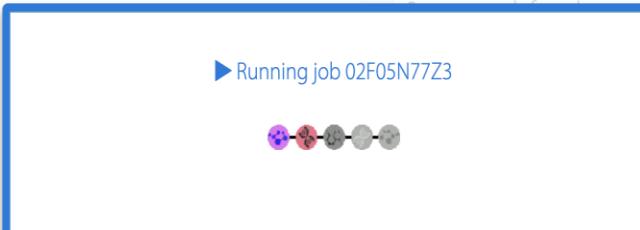
1 alternative compounds founds Show

Xyloonic acid

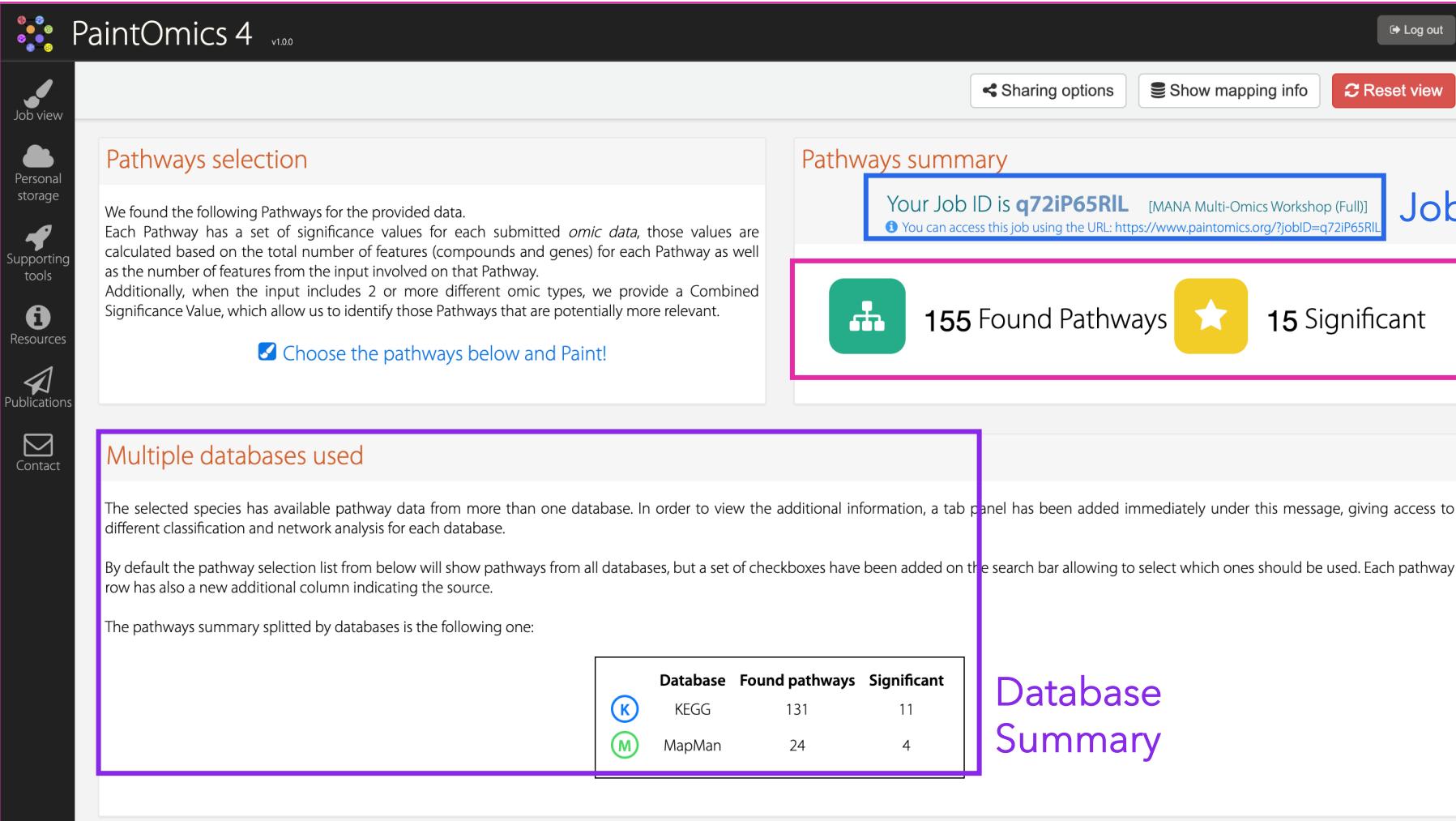
1 compounds founds

L-Xyloonic acid

0 alternative compounds founds Show



Step-by-step instructions - Output Summary



Job ID and URL

Pathways Summary

Database Summary

PaintOmics 4 v1.00

Pathways selection

We found the following Pathways for the provided data. Each Pathway has a set of significance values for each submitted *omic data*, those values are calculated based on the total number of features (compounds and genes) for each Pathway as well as the number of features from the input involved on that Pathway. Additionally, when the input includes 2 or more different *omic* types, we provide a Combined Significance Value, which allow us to identify those Pathways that are potentially more relevant.

Choose the pathways below and Paint!

Pathways summary

Your Job ID is **q72iP65RIL** [MANA Multi-Omics Workshop (Full)]
You can access this job using the URL: <https://www.paintomics.org/?jobID=q72iP65RIL>

155 Found Pathways **15 Significant**

Multiple databases used

The selected species has available pathway data from more than one database. In order to view the additional information, a tab panel has been added immediately under this message, giving access to different classification and network analysis for each database.

By default the pathway selection list from below will show pathways from all databases, but a set of checkboxes have been added on the search bar allowing to select which ones should be used. Each pathway row has also a new additional column indicating the source.

The pathways summary splitted by databases is the following one:

Database	Found pathways	Significant	
K	KEGG	131	11
M	MapMan	24	4

Step-by-step instructions - Output

KEGG MapMan

Pathways classification (KEGG database)

Category Distribution

Click on each slice to view the distribution of the subcategories.

Category	Percentage
M	75.57%
G	16.03%
E	3.05%
C	2.29%

Filter by category

Use this tool to [Show or Hide Pathways](#) based on their classification

- ▶ Cellular Processes
- ▶ Environmental Information Processing
- ▶ Genetic Information Processing
- ▶ Metabolism
- ▶ Organismal Systems [Show](#) [Hide](#) [Custom](#)

✓ [Apply](#)

Database Information

Pathways Classification

Step-by-step instructions - Metabolite analysis

Click on Paint Icon to paint expression of metabolites and its neighboring genes

Metabolites Hub Analysis

Neighbouring genes for each metabolite at 1 to 4 network steps are identified.

The percentile and binomial tests are used to identify metabolites with a high density of DEGs in their proximal network.

Paint	Search	Metabolite	ID	Step.	DE neighbor	not DE neighbor	Percentage	Percentile	P-values..	FDR BH
		Succinic ...	C00...	4	9	62	0.1268	0.48	0.99844	1.00000
		L-Malic a...	C00...	4	9	47	0.1607	0.58	0.97635	1.00000
		D-Glucose	C00...	2	8	12	0.4	0.84	0.13066	1.00000
		alpha-Ke...	C00...	3	8	67	0.1067	0.50	1.00000	1.00000
		L-Threon...	C00...	4	8	47	0.1455	0.53	0.98819	1.00000
		L-Serine	C00...	3	7	48	0.1273	0.53	1.00000	1.00000

Select a step: --Select--

Expression Value

Use this tool to show expression details of metabolites regulated features

Metabolite Expression Value

★ Succinic aci...
Succinic acid

Metabolite class activity analysis

Paint	Name	Unique Features	P Value	FDR BH	FDR BY
	Amines	4	1.00000	1.00000	1.00000
	Amino acids	34	0.03874	0.21310	0.64350
Paint this classification					
	Carboxylic acids	9	5.9117e-4	0.00650	0.01960
	Cofactors	1	1.00000	1.00000	1.00000
	Fatty acids	3	1.00000	1.00000	1.00000
	Monosaccharides	14	0.99473	1.00000	1.00000
	Neurotransmitters	4	1.00000	1.00000	1.00000

Expression Value

Use this tool to show expression details of metabolites based on their classification

Metabolomics Only relevant

★ D-Lysine
Lysine: D-Lysine...

★ L-Lysine
lysine: L-Lysine...

★ L-Isoleucine
isoleucine: L-...

★ D-Isoleucine
isoleucine: D-...

★ L-Valine
Valine: L-Vali...

★ D-Valine
Valine: D-Vali...

Expression values for metabolites

Step-by-step instructions - Pathway Enrichment

Pathway enrichment

Search Regular expression Case sensitive Search by gene/compound **Databases to view:** KEGG MapMan **Show FDR:** No

Paint	Pathway name	Features		Significance tests			External links
		Unique genes	Unique metabol.	Gene expression	Metabolomics	Combined pvalue (Fisher)	
	Biosynthesis of secondary metabolites	203	88	0.02760	3.6100e-4	1.2472e-4	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	Citrate cycle (TCA cycle)	4	14	1.00000	1.18336e-4	0.00119	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	MAPK signaling pathway - plant	17	0	0.00210	-	0.00210	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	Metabolism-Amino-acids	866	0	0.01053	-	0.01053	<input type="checkbox"/> MapMan <input type="checkbox"/> PubMed
	Biosynthesis of amino acids	34	46	0.83674	0.00211	0.01297	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	Phenylpropanoid biosynthesis	23	6	0.00192	1.00000	0.01393	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	Flavonoid biosynthesis	6	0	0.01588	-	0.01588	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	Pentose and glucuronate interconversions	12	18	0.00337	0.67287	0.01607	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	Glyoxylate and dicarboxylate metabolism	14	28	0.54517	0.00427	0.01643	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed
	Metabolism2nd-Synthesis-JA	13	0	0.01861	-	0.01861	<input type="checkbox"/> MapMan <input type="checkbox"/> PubMed
	Stilbenoid, diarylheptanoid and gingerol biosynthesis	4	0	0.02600	-	0.02600	<input type="checkbox"/> KEGG <input type="checkbox"/> PubMed

Paint this pathway

Step-by-step instructions - Pathway Mapping

PathwayES v1.0

Change visual settings

History Show Pathway Show Heatmap Search Settings

Draw heatmap

Citrate cycle (TCA cycle)

Download

Pathway information

Search in this pathway

Citrate cycle (TCA cycle)

Classification:

- Metabolism
- Carbohydrate metabolism

Matched features p-value

Gene expression	7 (7)	1.5457e-5	⊕
Metabolomics	7 (7)	1.1836e-4	⊕

Gene expression
No data for this pathway.

CITRATE CYCLE (TCA CYCLE)

The diagram illustrates the Citrate cycle (TCA cycle) with various enzymes and their corresponding EC numbers. Key enzymes include PCK1 (4.1.1.32), MDH (1.1.5.4), ACO3 (2.3.3.1, 2.3.3.3), and SDH2 (1.3). The cycle consists of Citrate, Isocitrate, and α-Ketoglutarate. Various metabolites are shown, including Acetyl-CoA, Oxaloacetate, (S)-Malate, Fumarate, Succinate, and Succinyl-CoA. The diagram also shows connections to other metabolic pathways such as Fatty acid biosynthesis, Fatty acid elongation in mitochondria, Val, Leu & Ile degradation, Fatty acid metabolism, Alanine, aspartate and glutamate metabolism, Glyoxylate and dicarboxylate metabolism, Pyruvate metabolism, and Amino acid metabolism. Enzymes are color-coded: blue for mitochondrial enzymes and red for cytosolic enzymes.

Step-by-step instructions – Pathway Mapping

Search in the pathway

Pathway information ×

Search in this pathway

Pyruvic acid 🔍 Search

← Back to Pathway details

Found 1 features.

Pyruvic acid

Heatmap Line chart

★ Metabolo...
pyruvic ac...

Relevant for this omic

External links

Search at KEGG Database
Search at PubChem Compound
Search at ChEBI Database

Find in Pathway Show details

Heatmap Settings

Global heatmap



This panel contains the heatmap for all the features involved on this pathway. Choose the visible omics features will be visible using the Settings button.

Choose the omics to draw

ⓘ Drag and drop to change the order in which heatmaps will be drawn.

- 1 Gene expression
 All features (Genes or compounds)
 Only relevant features

- 2 Metabolomics
 All features (Genes or compounds)
 Only relevant features

Advanced options

ⓘ Depending on the selected settings, heatmap generation can take up to 10 seconds.

- Force order for features.
 Clusterize data
 Hierarchical clustering
 K-means clustering

✓ Apply

Change visual settings

Visual settings ×

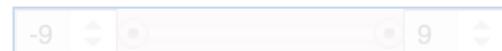
Metabolomics

Coloring options

Reference values

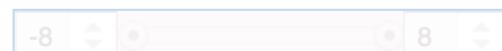
Gene expression

- Percentiles 10 and 90
 Global Min/Max (including outliers).
 Global Min/Max (without outliers).
 Custom values



Metabolomics

- Percentiles 10 and 90
 Global Min/Max (including outliers).
 Global Min/Max (without outliers).
 Custom values



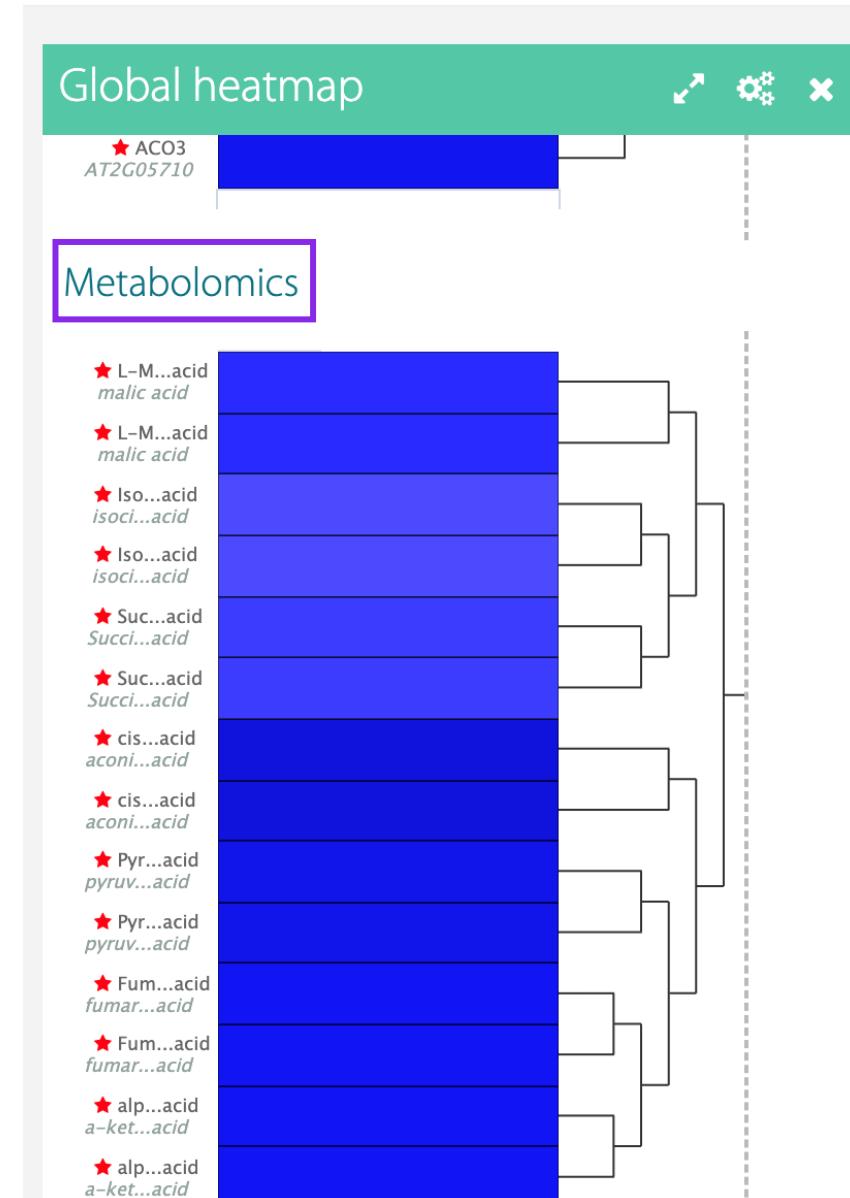
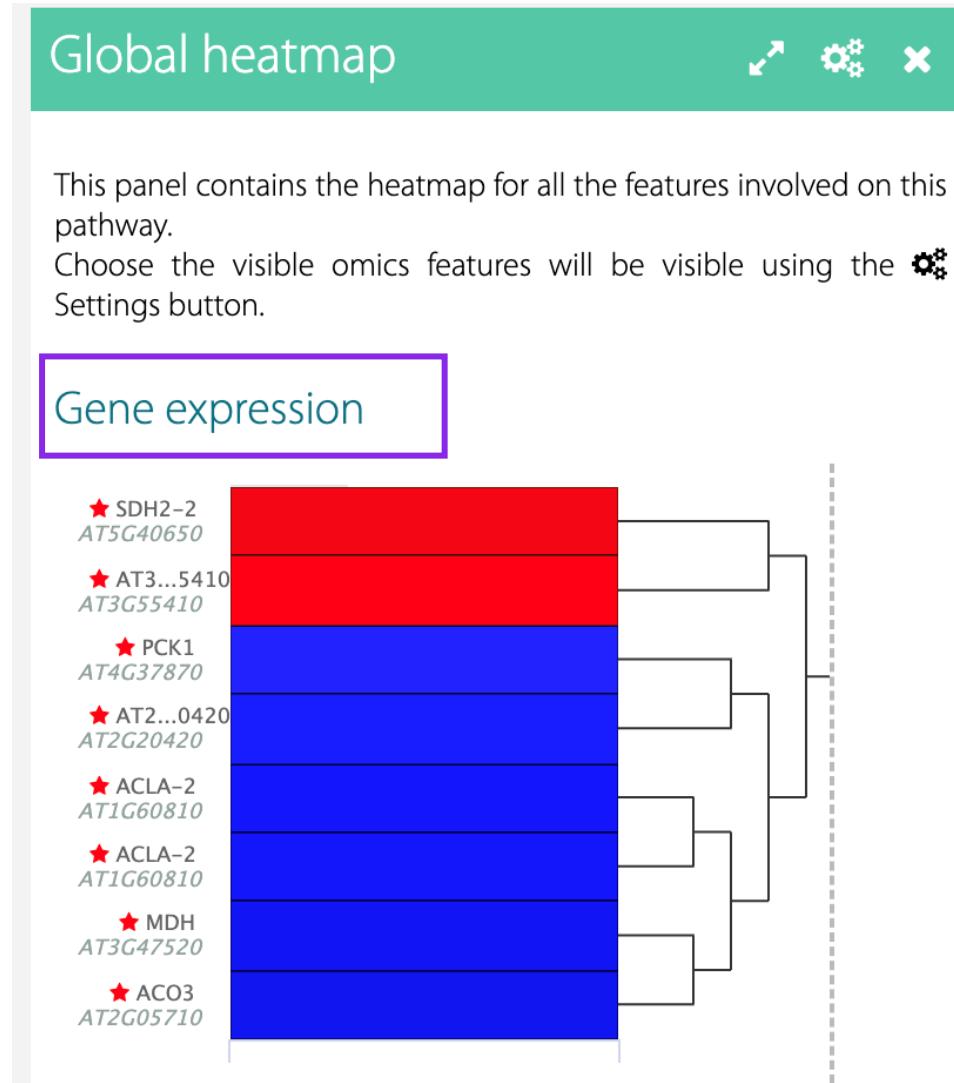
Color scale

- Blue-White-Red
 Green-Black-Red



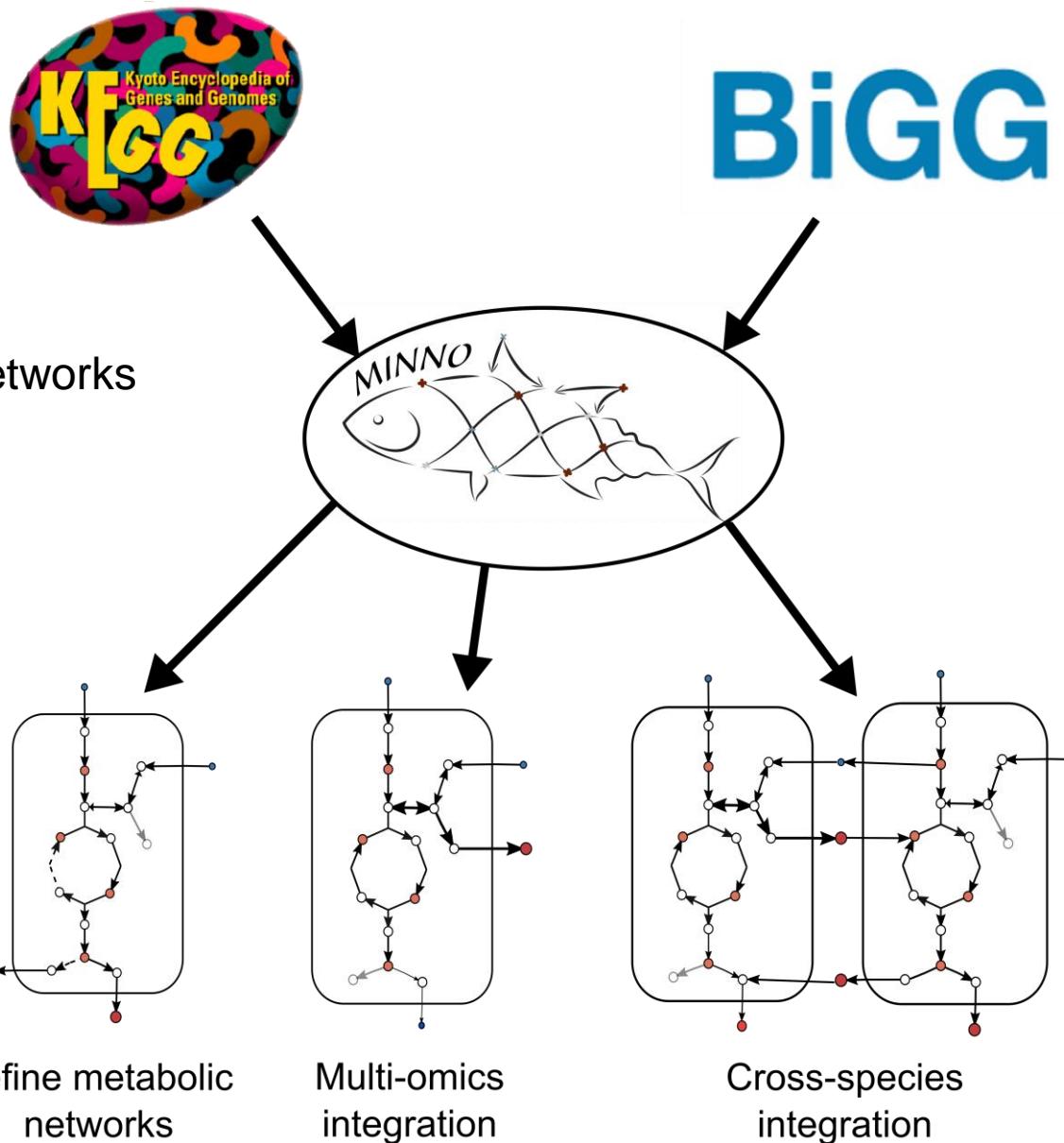
✓ Apply

Step-by-step instructions - Global Heatmap



Metabolic Interactive Nodular Network for Omics

- Thousands of organism networks
- Modular design
- Flexible user-interface
- No coding required



Ayush Mandwal



Access at
lewisresearchgroup.github.io/MINNO/

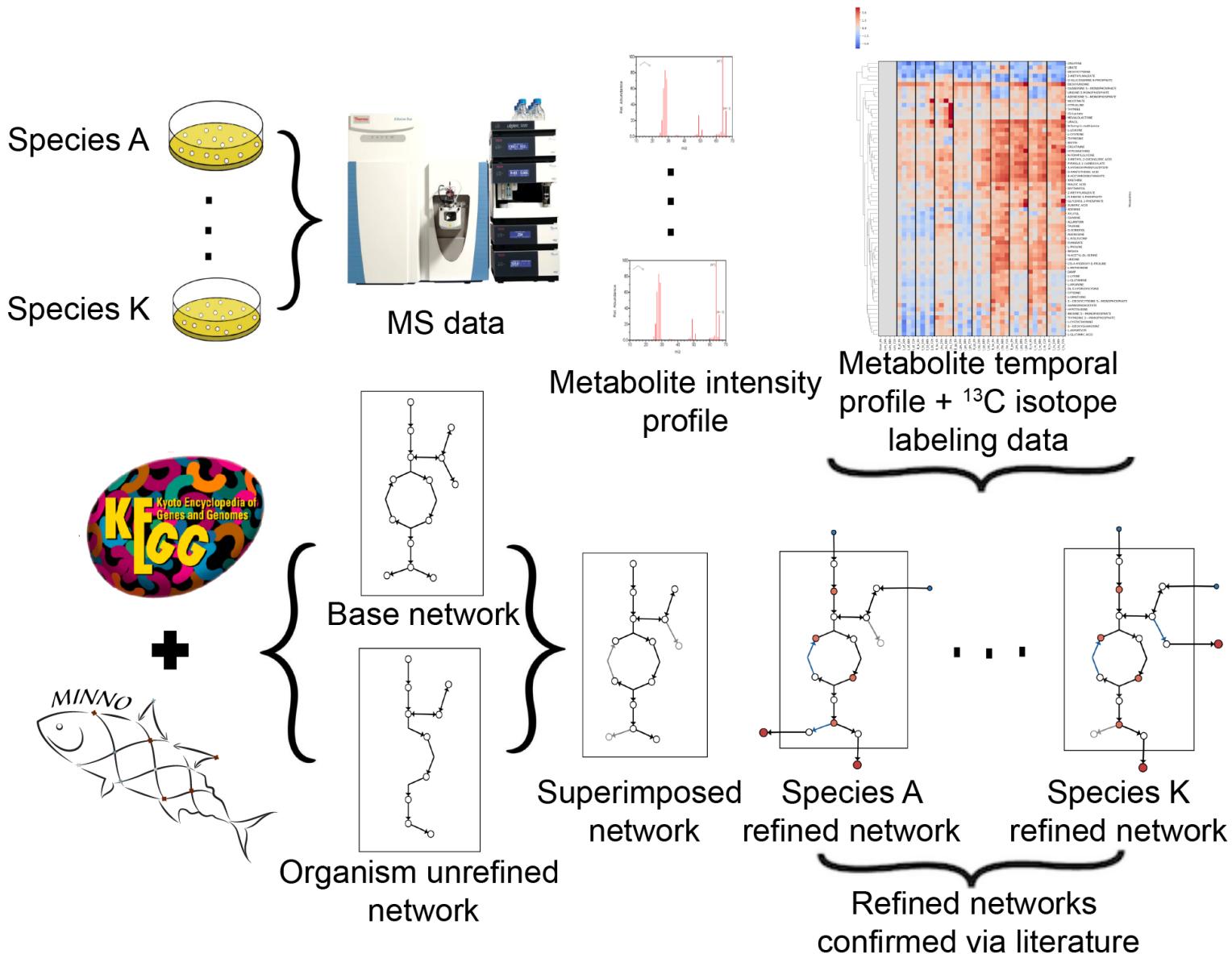
App layout and features

The screenshot shows a metabolic pathway simulation interface with the following features highlighted:

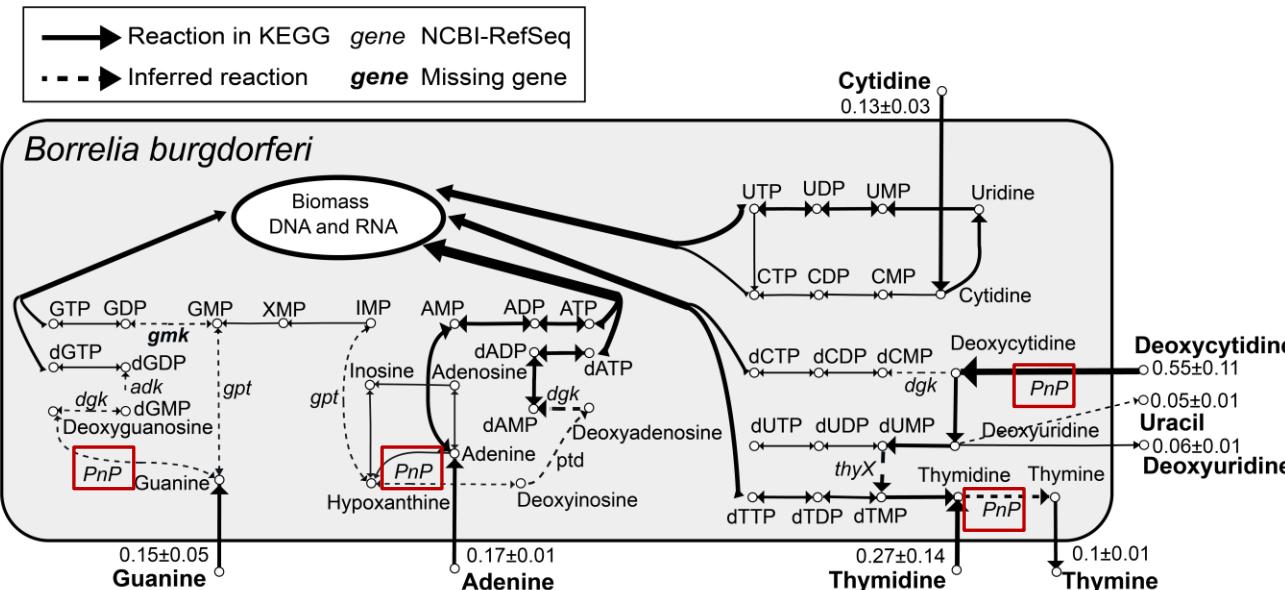
- Organism & reaction selection**: A red dashed circle highlights the left sidebar where the user can select an organism (KEGG Organisms list) and reactions (BiGG Models, alpha-Linolenic acid).
- Simulation OFF**: A red dashed circle highlights the top left button.
- Freeze OFF**: A red dashed circle highlights the top right button.
- Data selection**: A red dashed circle highlights the first row of data groups in the sidebar.
- Data & network formatting**: A red dashed circle highlights the bottom row of data groups and the zoom controls (1, 12, 100) in the sidebar.
- Added reaction**: A red dashed circle highlights a reaction in the pathway network.
- Annotated reaction**: A red dashed circle highlights a reaction in the pathway network.
- Non-annotated reaction**: A red dashed circle highlights a reaction in the pathway network.
- Tooltip menu**: A red dashed circle highlights the tooltip menu in the bottom right corner.
- Context menu**: A blue sidebar menu on the right includes: Import, Export, Delete, Insert, Select, Geometric operations, Undo, Redo, and Help.

The central area displays a metabolic pathway network with nodes representing metabolites and edges representing reactions. Key nodes include L-Arginine, Fumarate, N-(L-Arginino)succinate, L-Ornithine, L-Citrulline, Urea, Urea-1-carboxylate, CO2, L-Aspartate, N-Acetyl-L-citrulline, N-Acetyl-L-glutamate 5-semialdehyde, LysW-L-ornithine, LysW-L-glutamate, Carbamoyl phosphate, and N-Acetyl-L-glutamate 5-semialdehyde.

Metabolic network refinement strategy

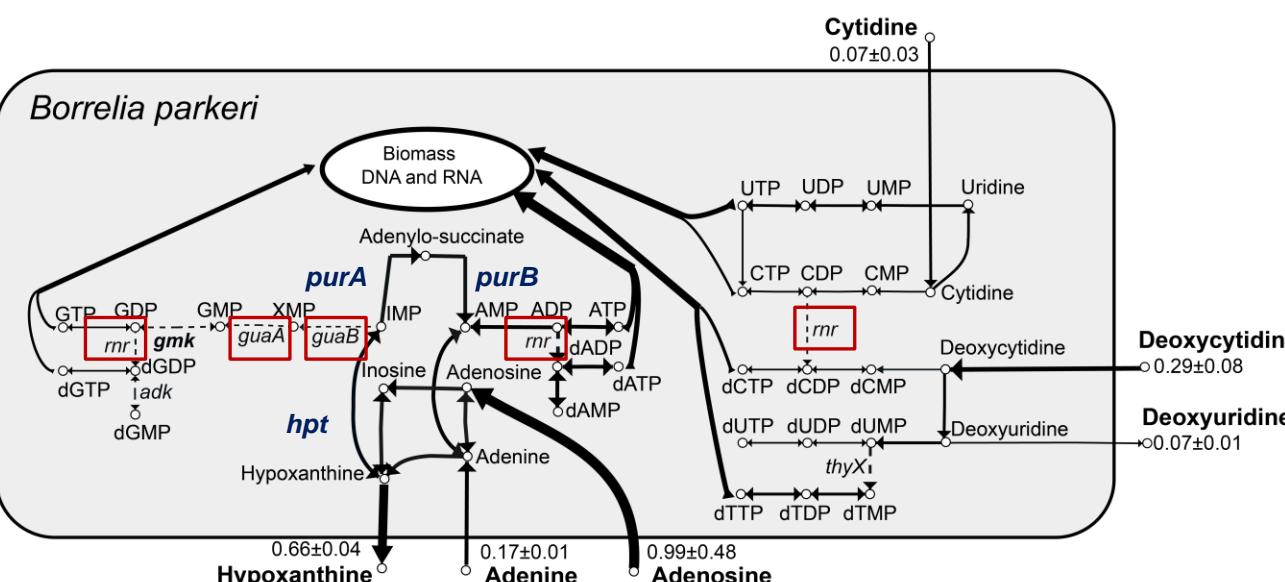


Example: divergent nucleotide metabolism in *Borrelia* species



Borrelia burgdorferi (Lyme disease) uses alternate purine salvage pathways

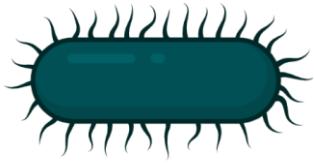
- Synthesizes all purines from adenine and guanine
- Uses pyrimidine-nucleoside phosphorylase (*PnP*) to salvage deoxyribose sugars



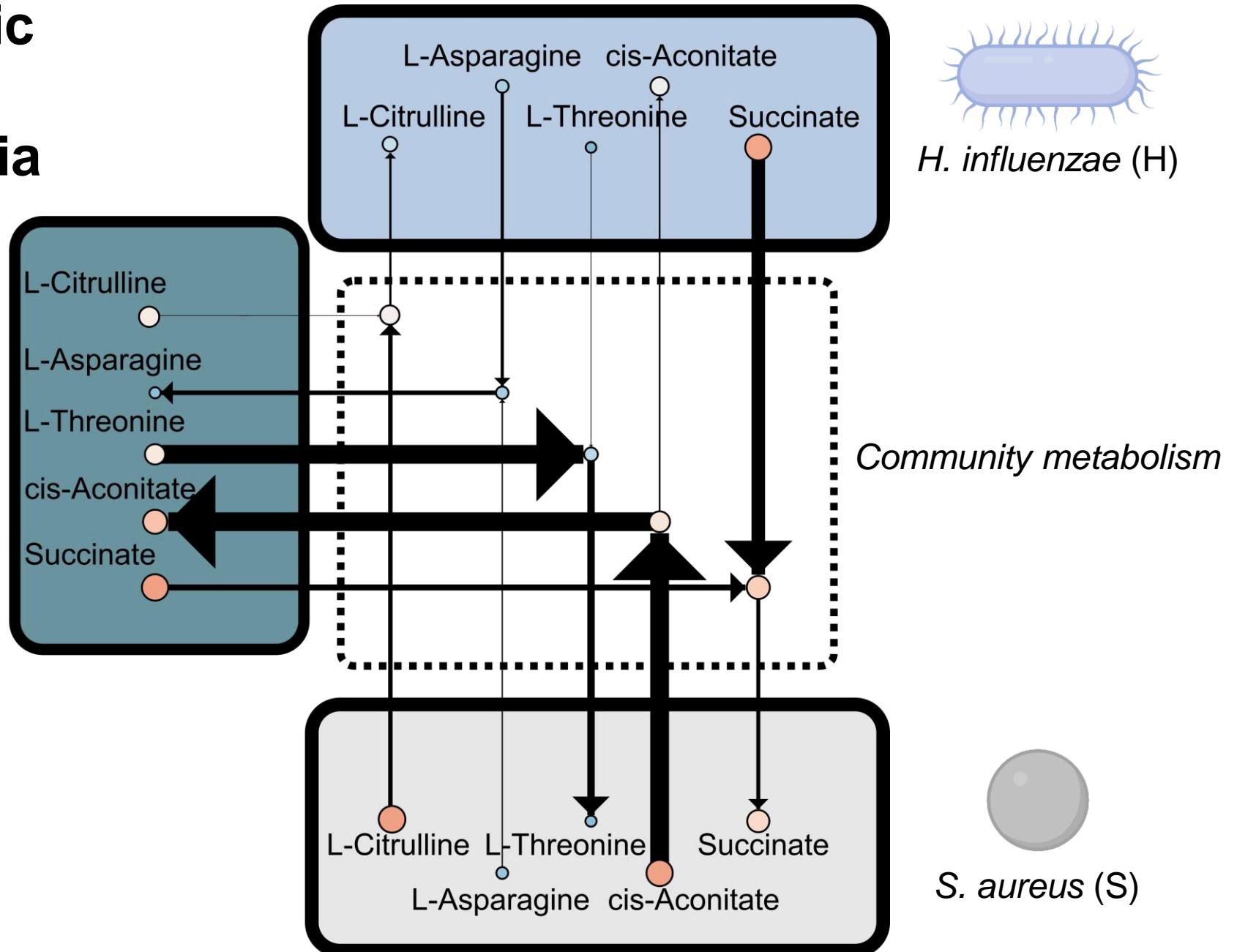
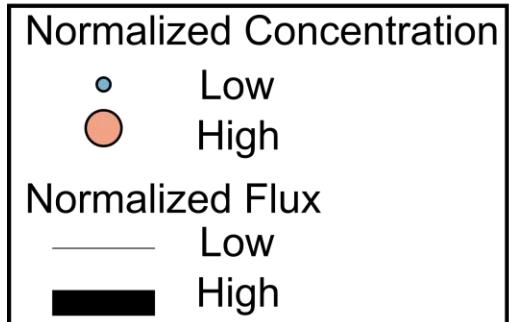
Borrelia parkeri (relapsing fever) has classic purine salvage genes *purA*, *purB*, *hpt*

- Synthesizes all purines from adenine
- Uses GMP synthase (*guaA*), IMP dehydrogenase (*guaB*), and ribonucleotide reductase (*rnr*) for classic purine salvage

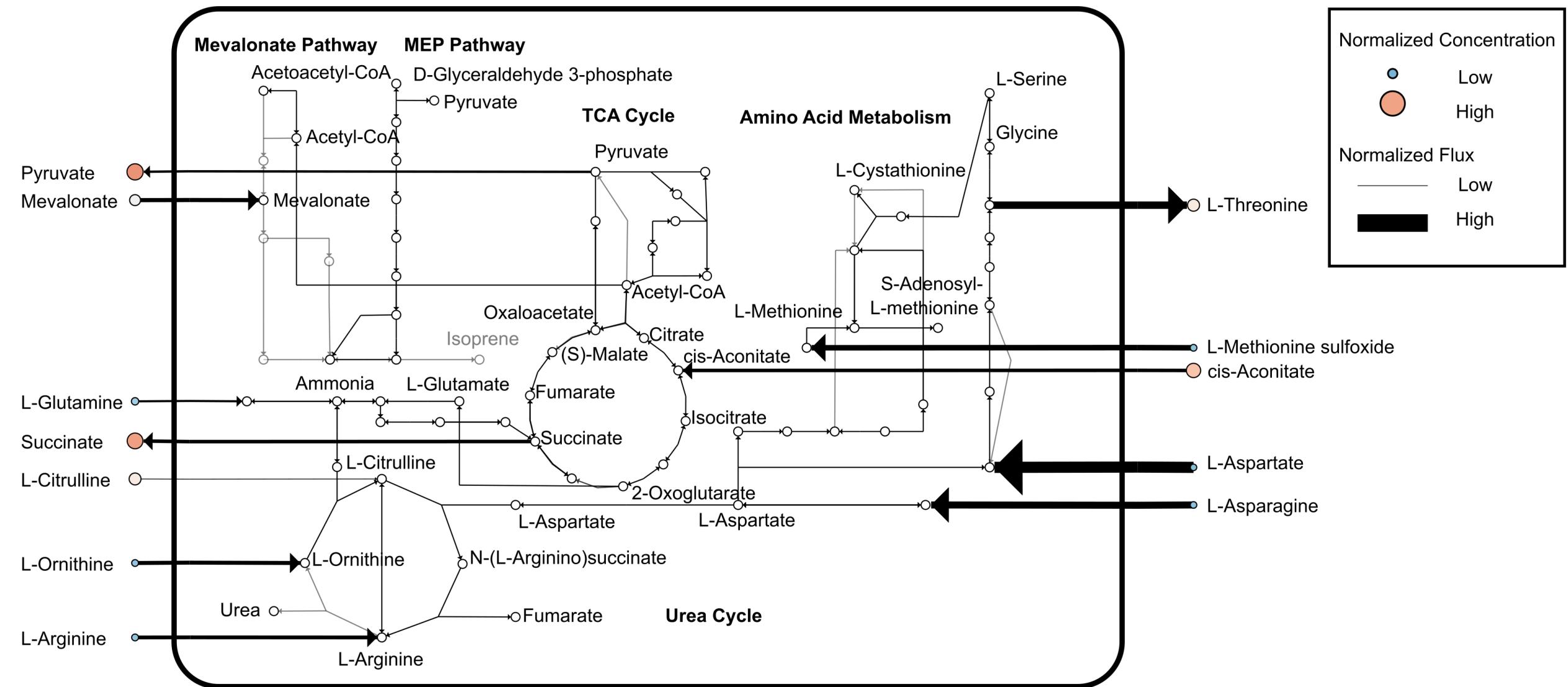
Example: Metabolic crosstalk between pathogenic bacteria



P. aeruginosa (P)



Example: *Pseudomonas aeruginosa* metabolic flux network



Tutorial: Plotting intracellular and proteomics TCA cycle data for Methicillin-resistant *S. aureus* (MRSA) vs. Methicillin-susceptible *S. aureus* (MSSA)

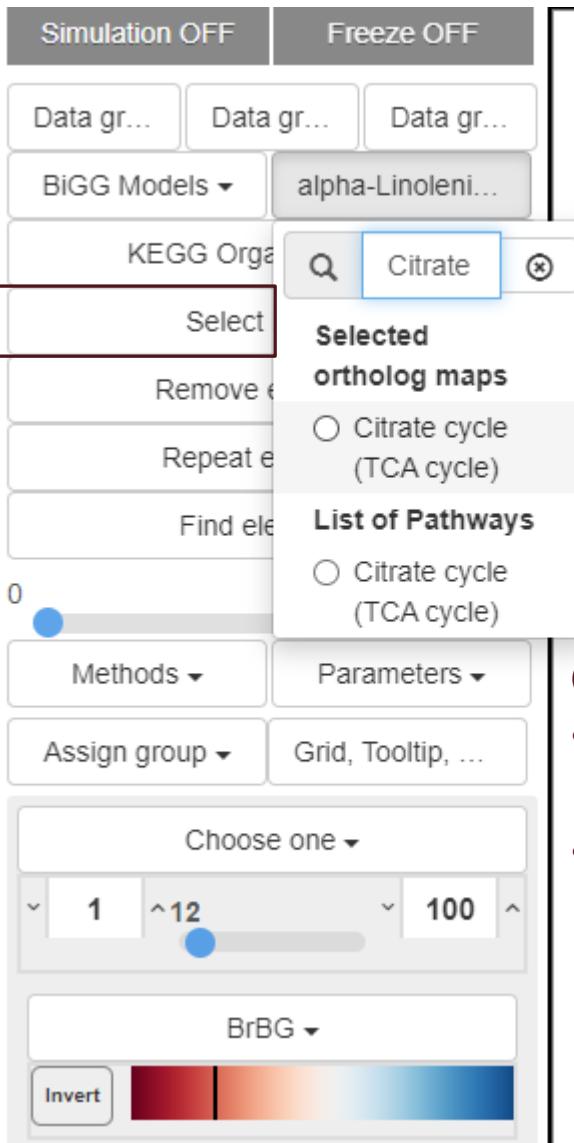
Network from KEGG Organism code saa:
Staphylococcus aureus subsp. *aureus* USA 300 – FPR3757 (CA-MRSA)

Dataset:

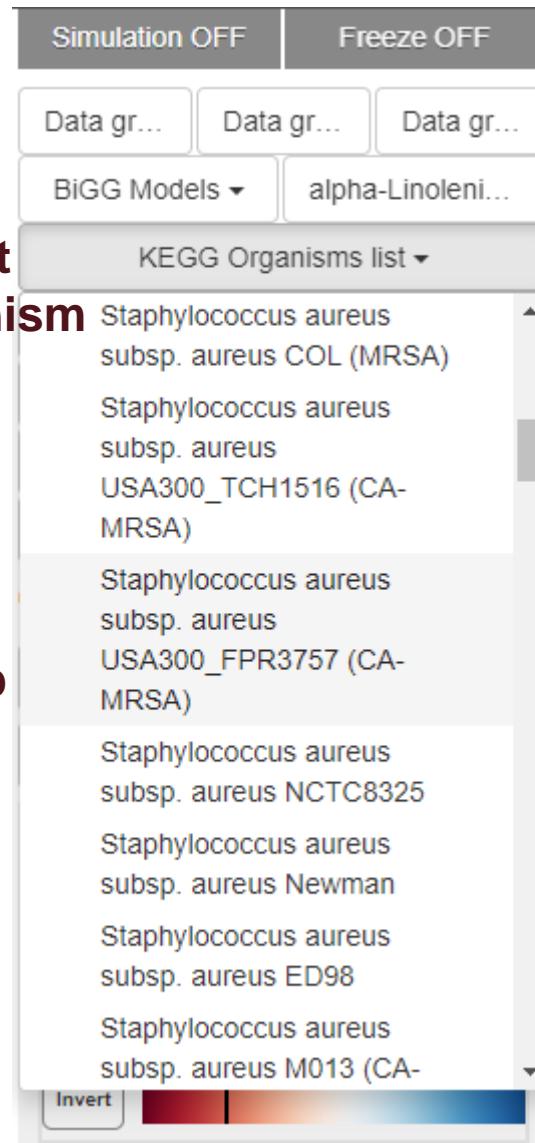
Compounds and protein name (note – no underscores in names)	Omics type	Sample type	Treatment	Control (use “1” for pre-normalized data)
	Label by data type			
Compound	Group	Metabolomics_MSSA_6hr	Metabolomics_MRSA_6hr	Metabolomics_C_6hr
Pyruvate	intracellular	868.0733333	1888.26	12.86009707
pyruvate:ferredoxin 2-oxidoreductase (CoA-acetyl link)	link	7.593511216	4.183030436	1
Citrate	intracellular	761.4166667	90.21333333	1762.048885
citrate hydro-lyase (cis-aconitate-forming)	link	5.739079163	3.657271408	1
cis-Aconitate	intracellular	803.8398679	732.4434115	8.005712896
isocitrate hydro-lyase (cis-aconitate-forming)	link	4.687377938	4.983074167	1
Isocitrate	intracellular	1206.22	6.586666667	2063.38
Isocitrate:NADP+ oxidoreductase	link	5.785368325	4.057682495	1
Succinate:CoA ligase (ADP-forming)	link	4.687377938	4.983074167	1
Succinate	intracellular	973.5512732	952.2254747	1448.89448
succinate:quinone oxidoreductase	link	0.992880306	5.481820183	1
Fumarate	intracellular	28.71666667	52.52333333	12.86009707
(S)-malate hydro-lyase (fumarate-forming)	link	1.750836018	1.407041251	1
(S)-Malate	intracellular	588.4085563	629.1442106	1448.89448

1. Download ortholog and organism networks

This has already been done for you in your sample network file

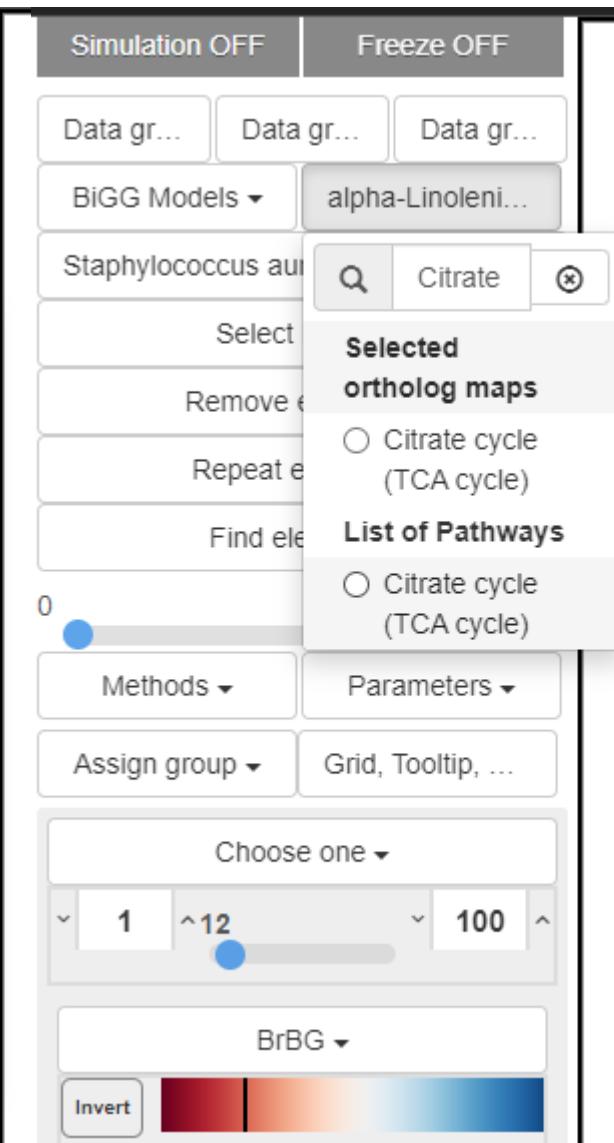


Select organism



Ortholog map

- Select ALL reactions
- Save as JSON file

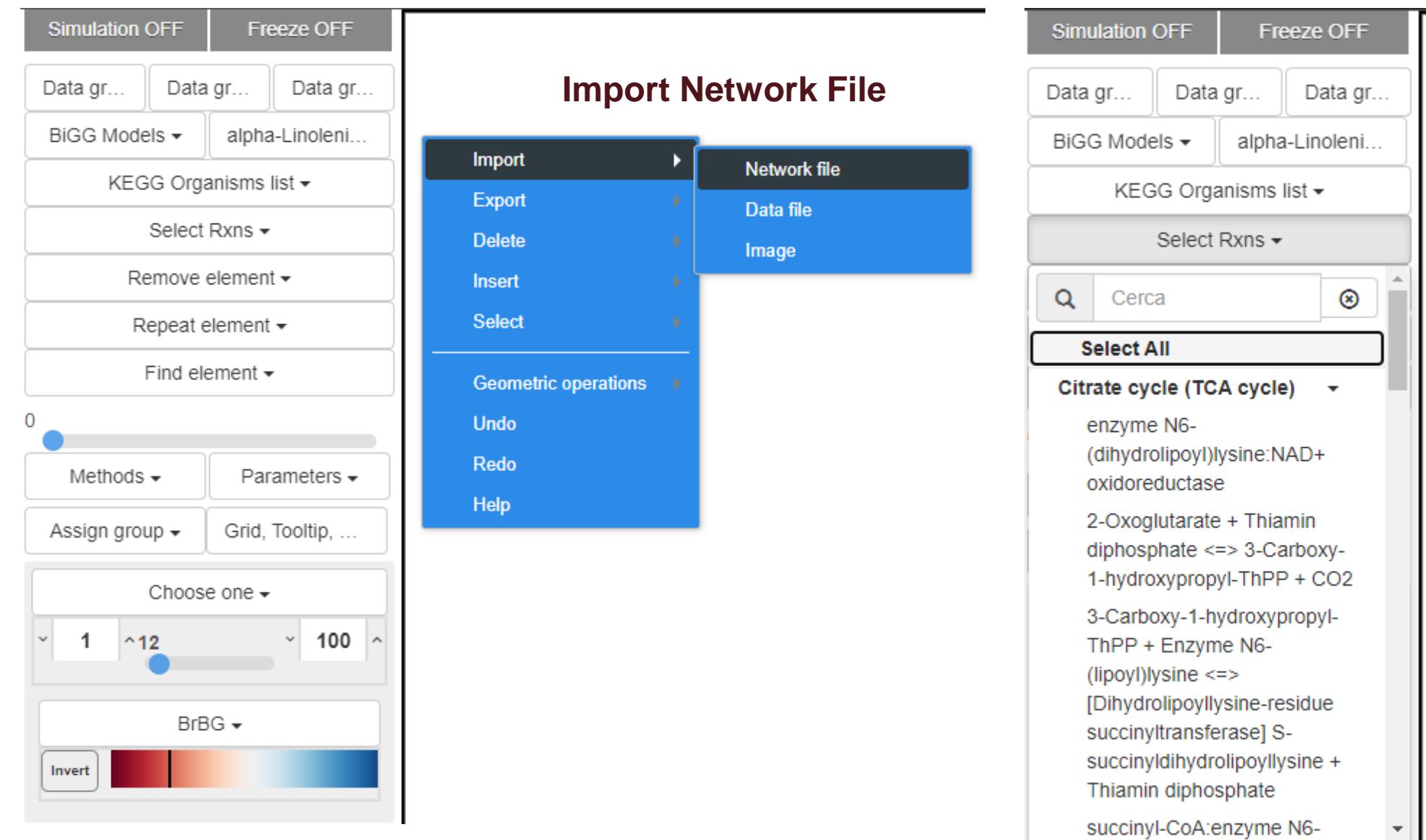


Select pathway

- Save KGML file that opens
- Re-upload KGML file and save as JSON file

2. Upload both ortholog and organism JSON file

This has already been done for you in your sample network file



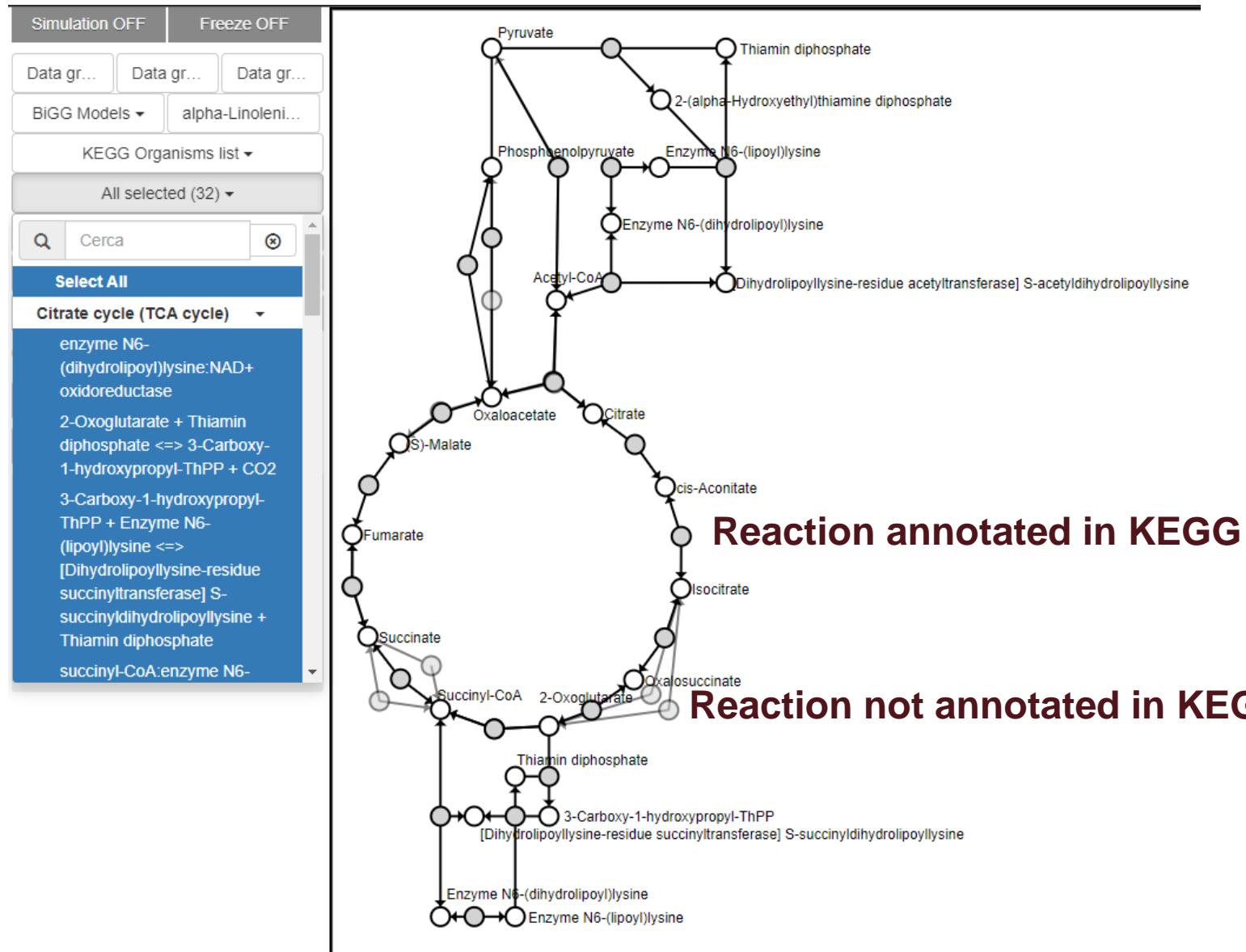
The screenshot shows the Cytoscape software interface. On the left, there is a toolbar with various buttons and dropdown menus. The main area is titled "Import Network File" and shows a dropdown menu with the following options: Import, Export, Delete, Insert, Select, Geometric operations, Undo, Redo, and Help. The "Import" option is highlighted. A sub-menu for "Network file" is open, showing "Data file" and "Image" as options. On the right side of the interface, a list of reactions is displayed under the heading "Select All". The reactions listed are:

- Citrate cycle (TCA cycle)
 - enzyme N6-(dihydrolipoyl)lysine:NAD+ oxidoreductase
 - 2-Oxoglutarate + Thiamin diphosphate <=> 3-Carboxy-1-hydroxypropyl-ThPP + CO2
 - 3-Carboxy-1-hydroxypropyl-ThPP + Enzyme N6-(lipoyl)lysine <=> [Dihydrolipoyllysine-residue succinyltransferase] S-succinylhydrolipoyllysine + Thiamin diphosphate
 - succinyl-CoA:enzyme N6-

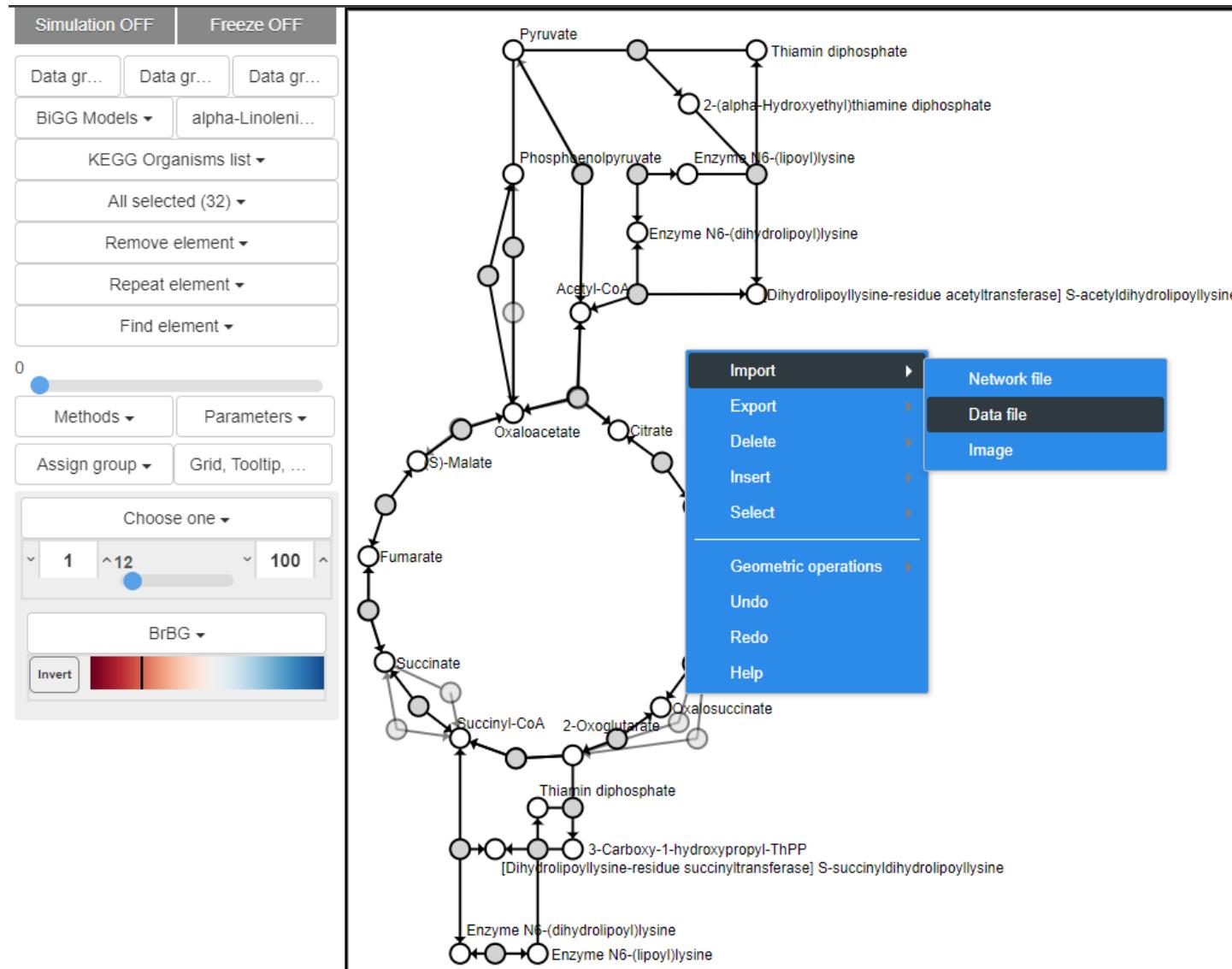
Select All Rxns
(or specific reactions to visualize)

3. Save your combined network as a JSON file

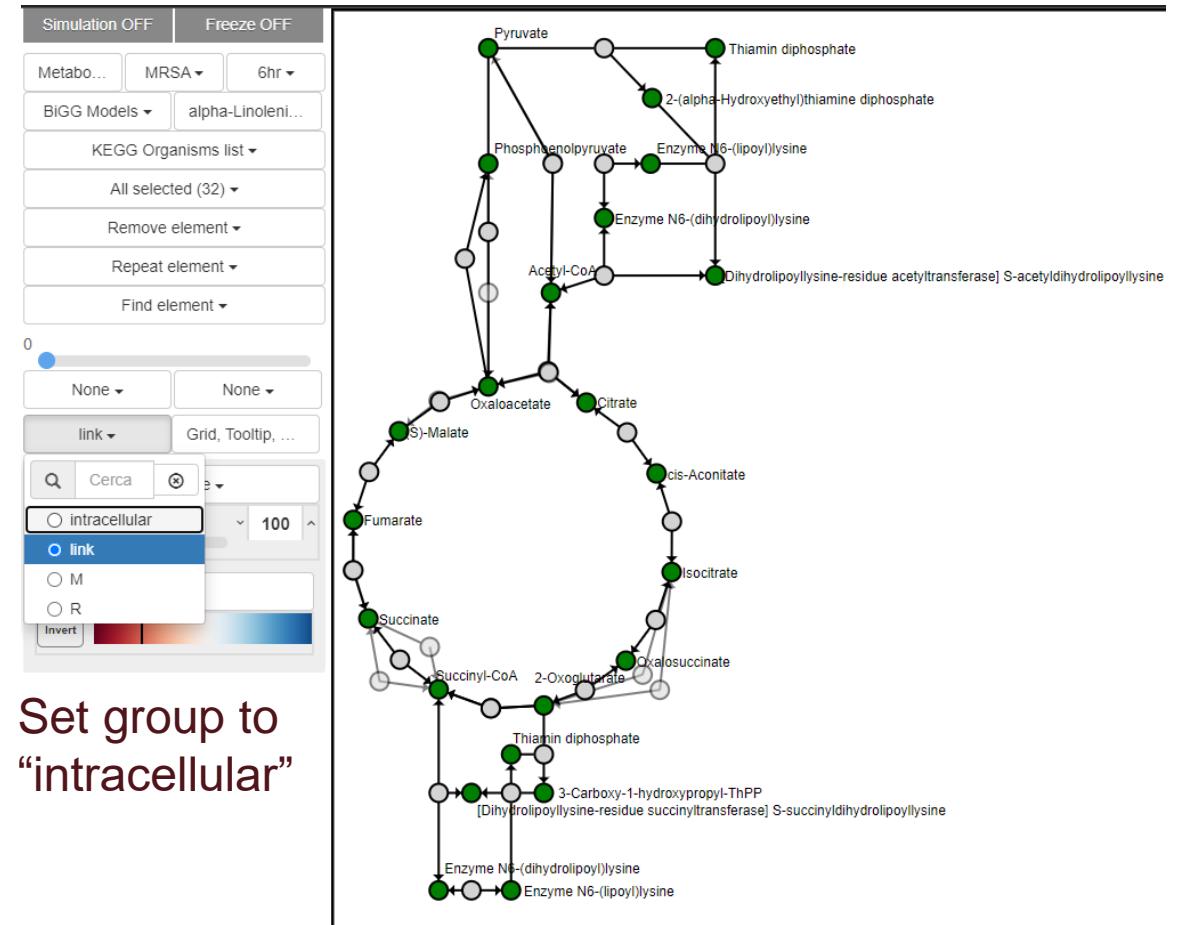
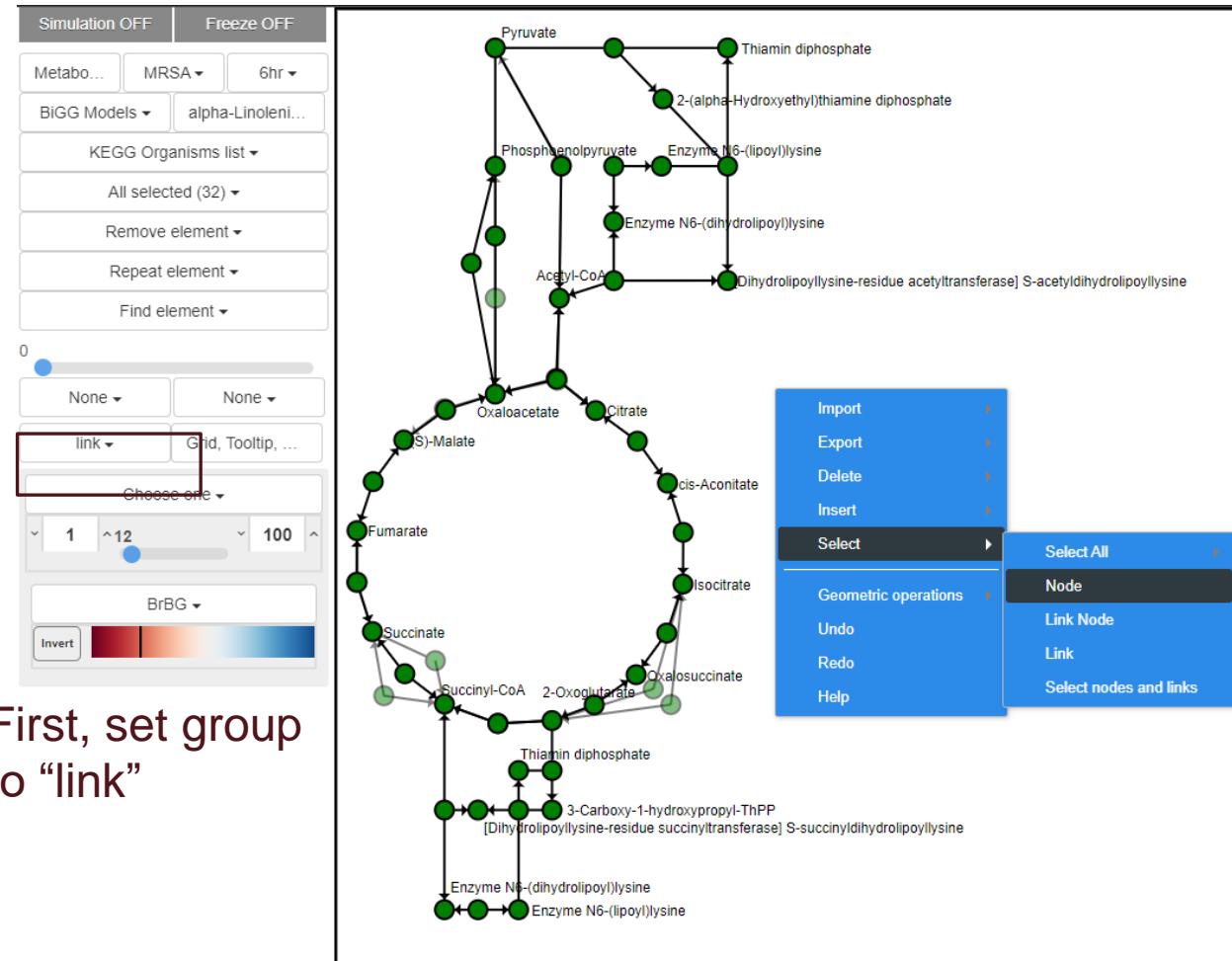
This has already been done for you in your sample network file



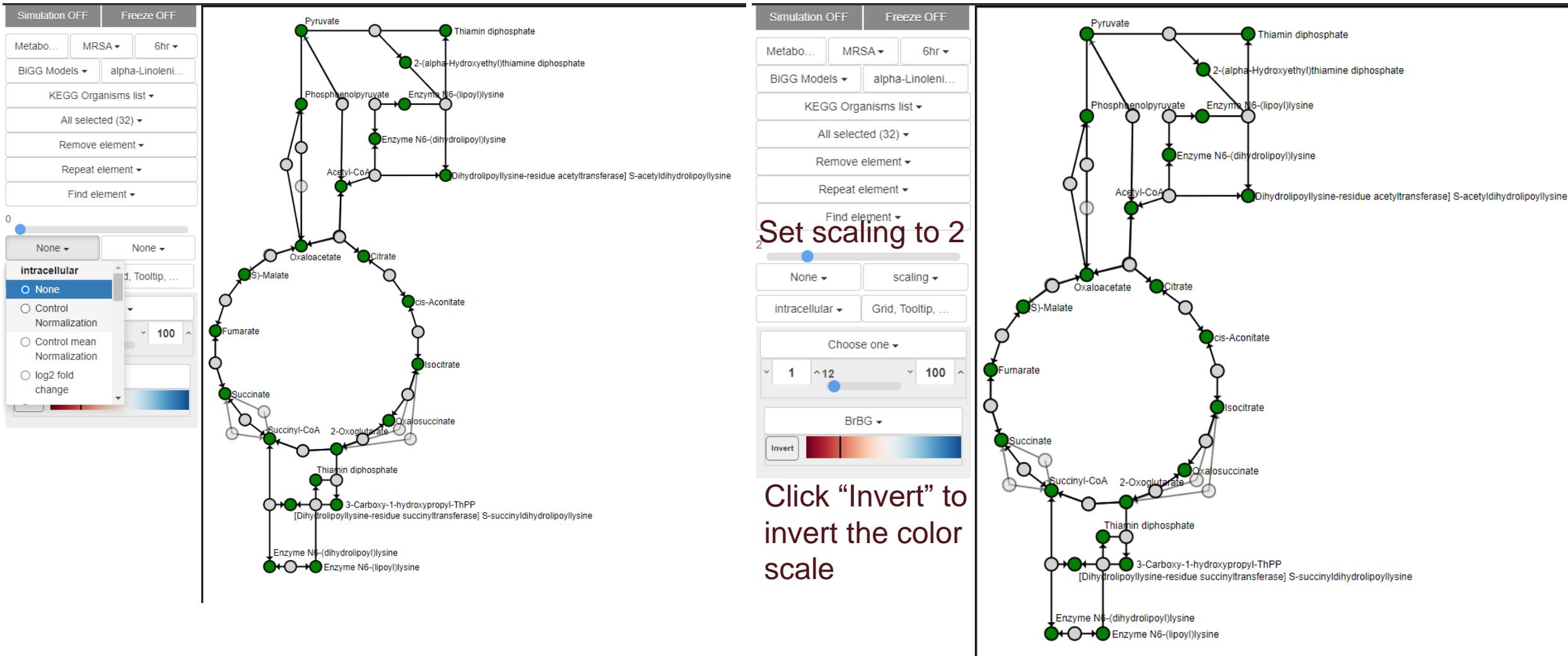
4. Open your combined network (sample network file) and select all reactions. Upload your sample data file.



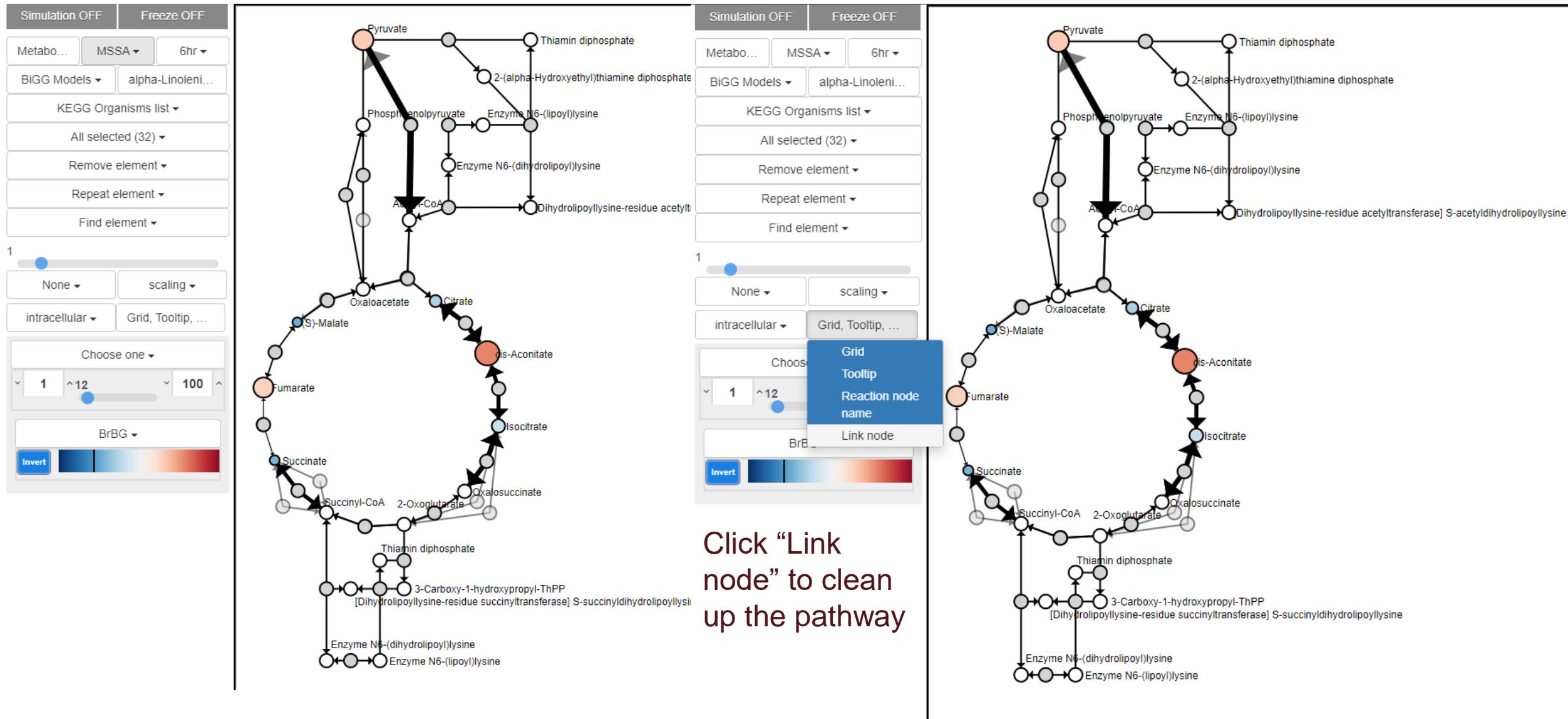
5. Use the SHIFT key to highlight all nodes and select “Node”. Set group to “intracellular”.



6. Use the “control normalization” feature and set the scaling to 2. For the “links” (proteome data), select “none” for normalization and set the scaling to 1. You can also invert the color scale.

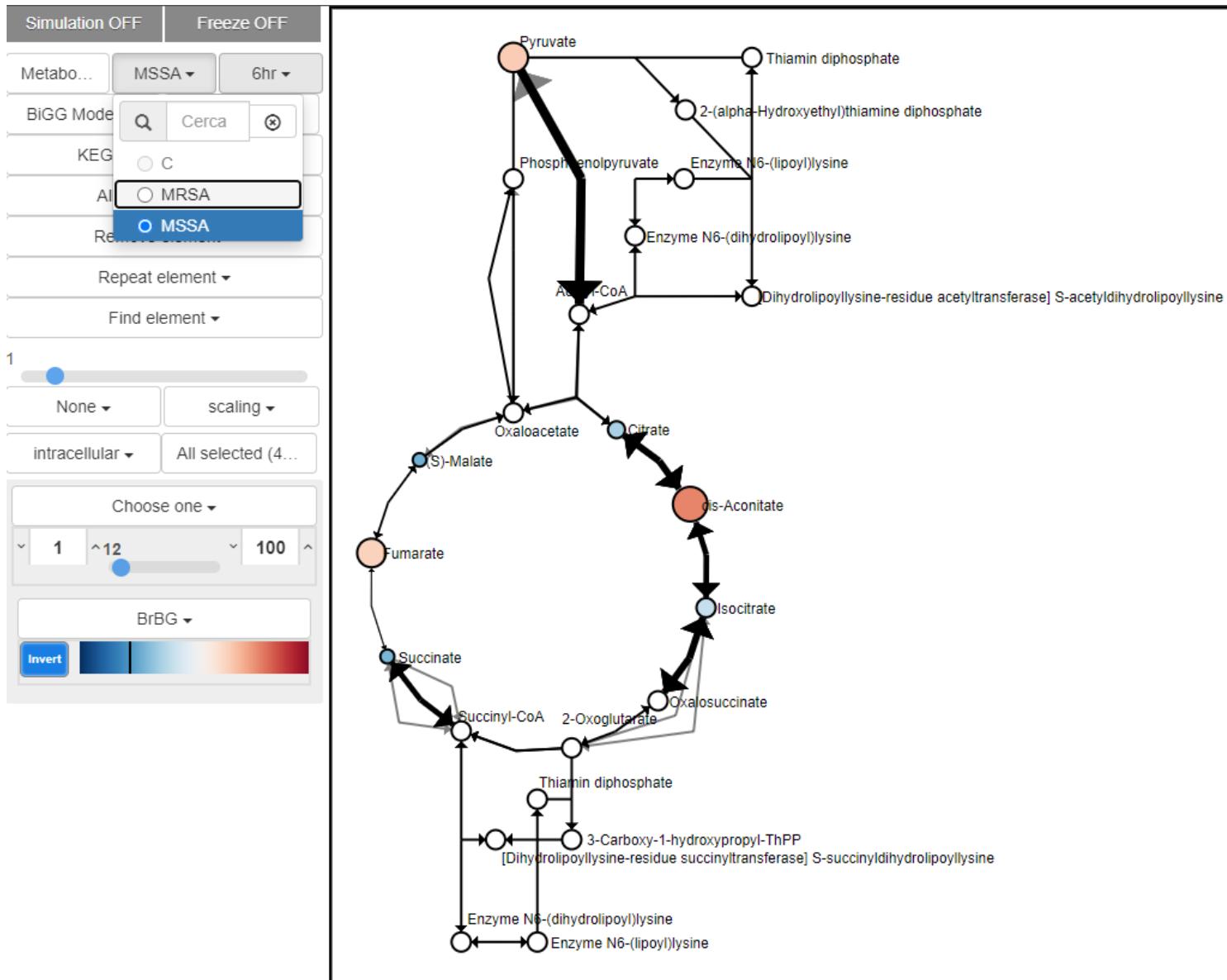


7. Select a treatment group (MSSA) to visualize the data.

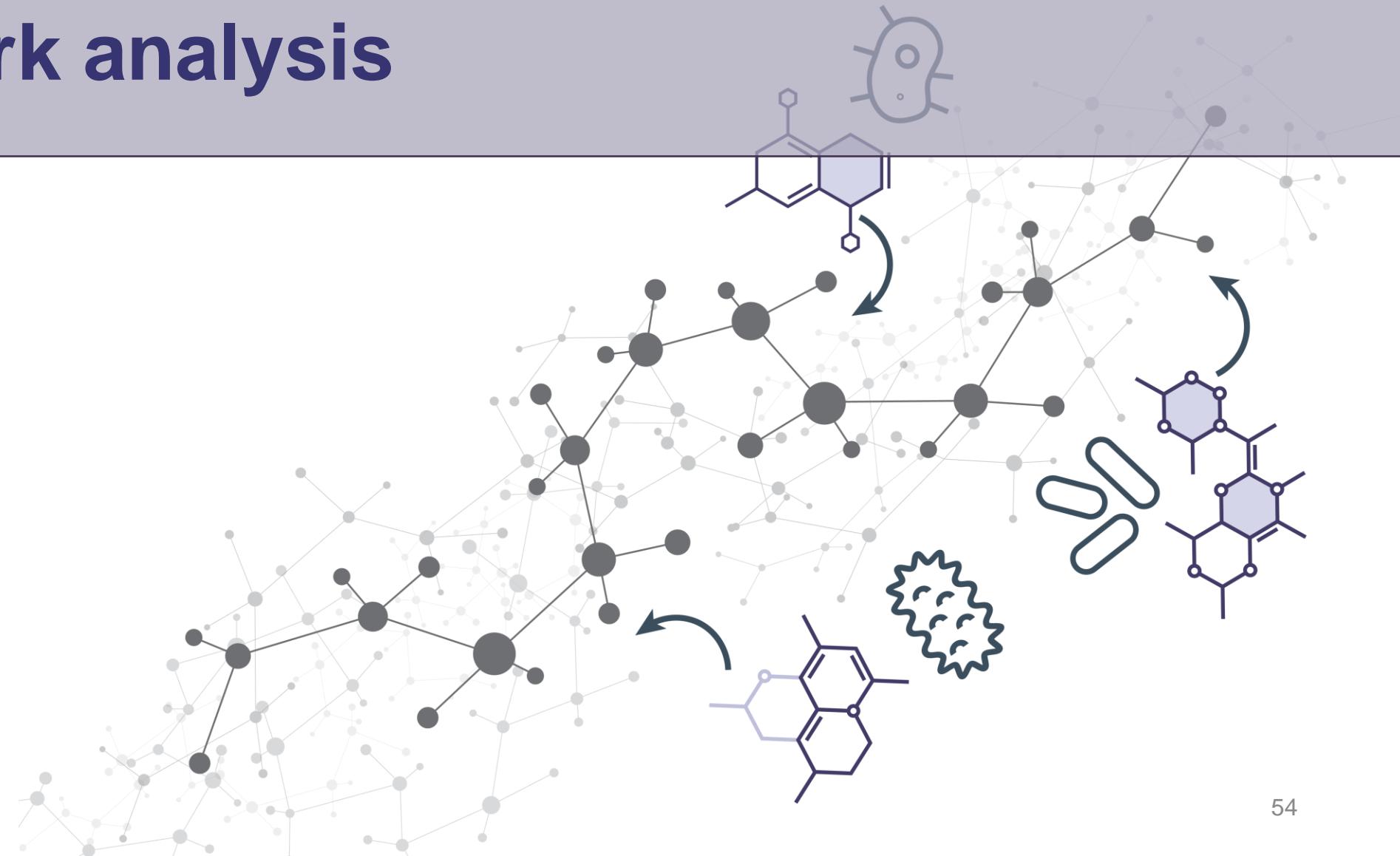


8. Use the data menus to visualize different sample groups and treatments (e.g., timepoints).

Data visualization options



Multi-omics tools for correlation and network analysis



Download R and access files using our GitHub repository

<https://github.com/sgosline/metabolomics-integration/tree/main>

