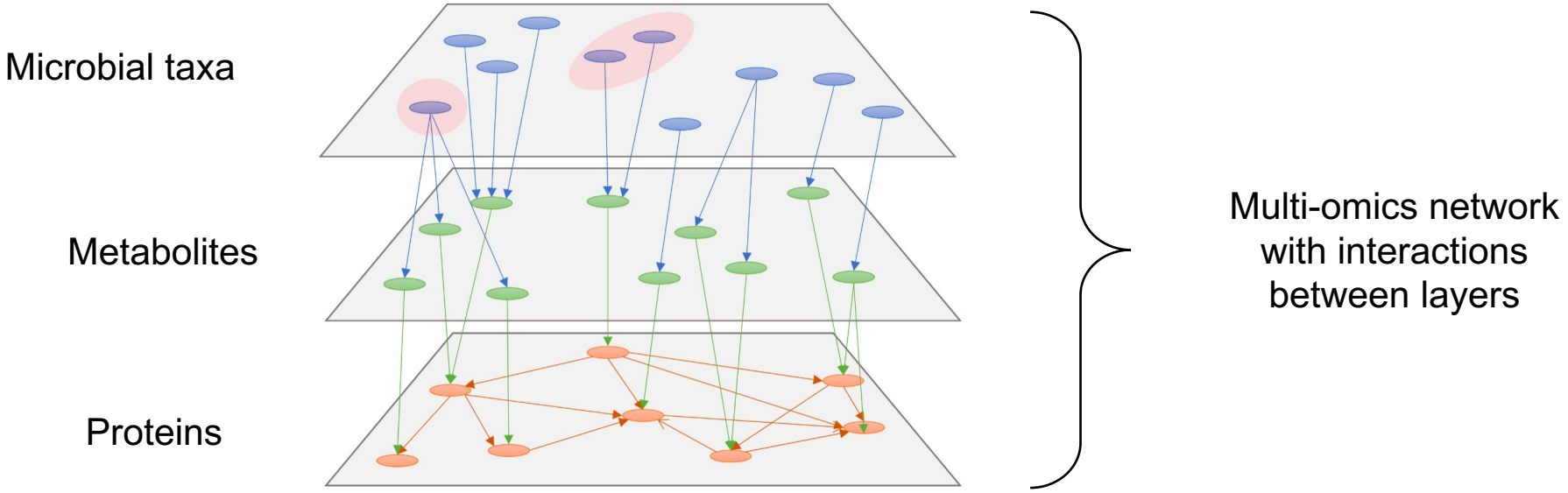
OmicsNet Tutorial: IBD Case Study

Computer Requirement

- Modern browser supporting WebGL
- Chrome 50+, Firefox 47+, Safari 10.1+ and Edge 12+
- Please make sure WebGL is enabled in your browser
  - Please consult this web page to verify: https://get.webgl.org/
- If not enabled, please consult our FAQ page for instructions
- For best performance and visualization, use:
- Latest version of Google Chrome
- A modern computer with at least 4GB of physical RAM
- A 15-inch screen or bigger (larger is better)
- Retina Display is supported

Motivation

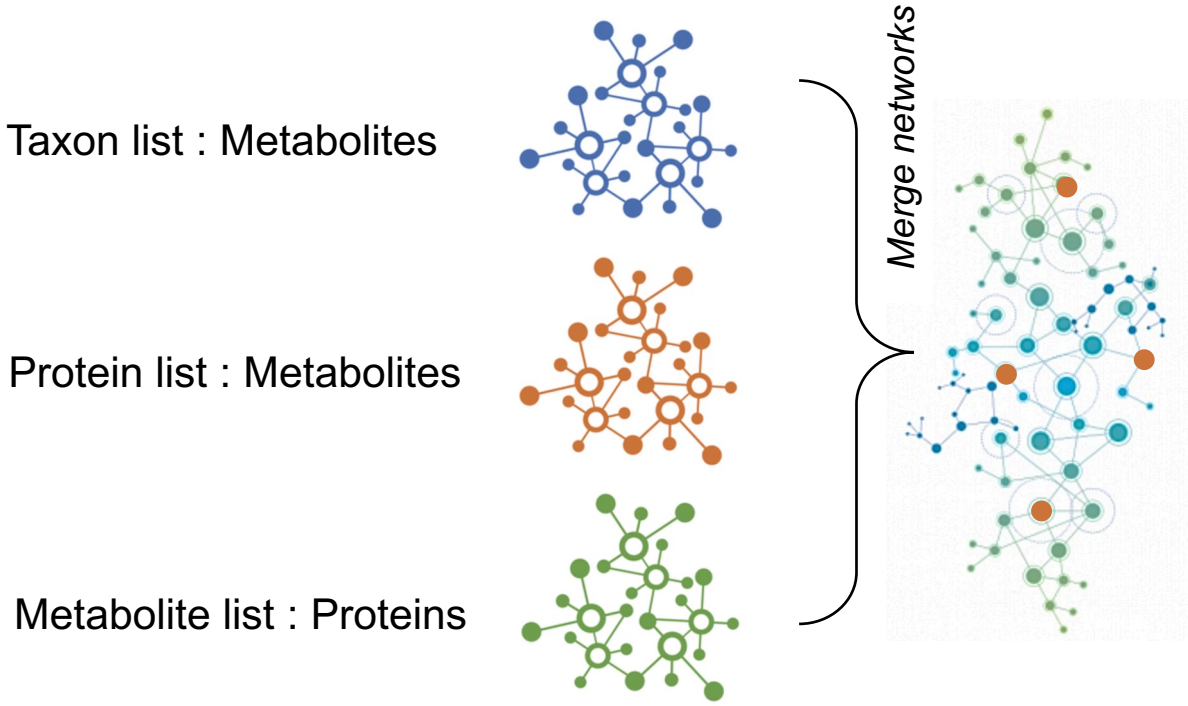
A recent study collected multi-omics data from stool samples from patients with Crohn's disease, a sub-type of inflammatory bowel disease (IBD), to try and understand gut-microbiome drivers of dysbiosis. Here, we analyze lists of molecular features (metabolites, proteins, microbial taxa) that were significantly different between samples collected from dysbiotic and non-dysbiotic patients. The main motivation of this case study is to demonstrate how OmicsNet can be used to integrate and provide background context for multi-omics lists.



Analysis Overview

The main steps are:

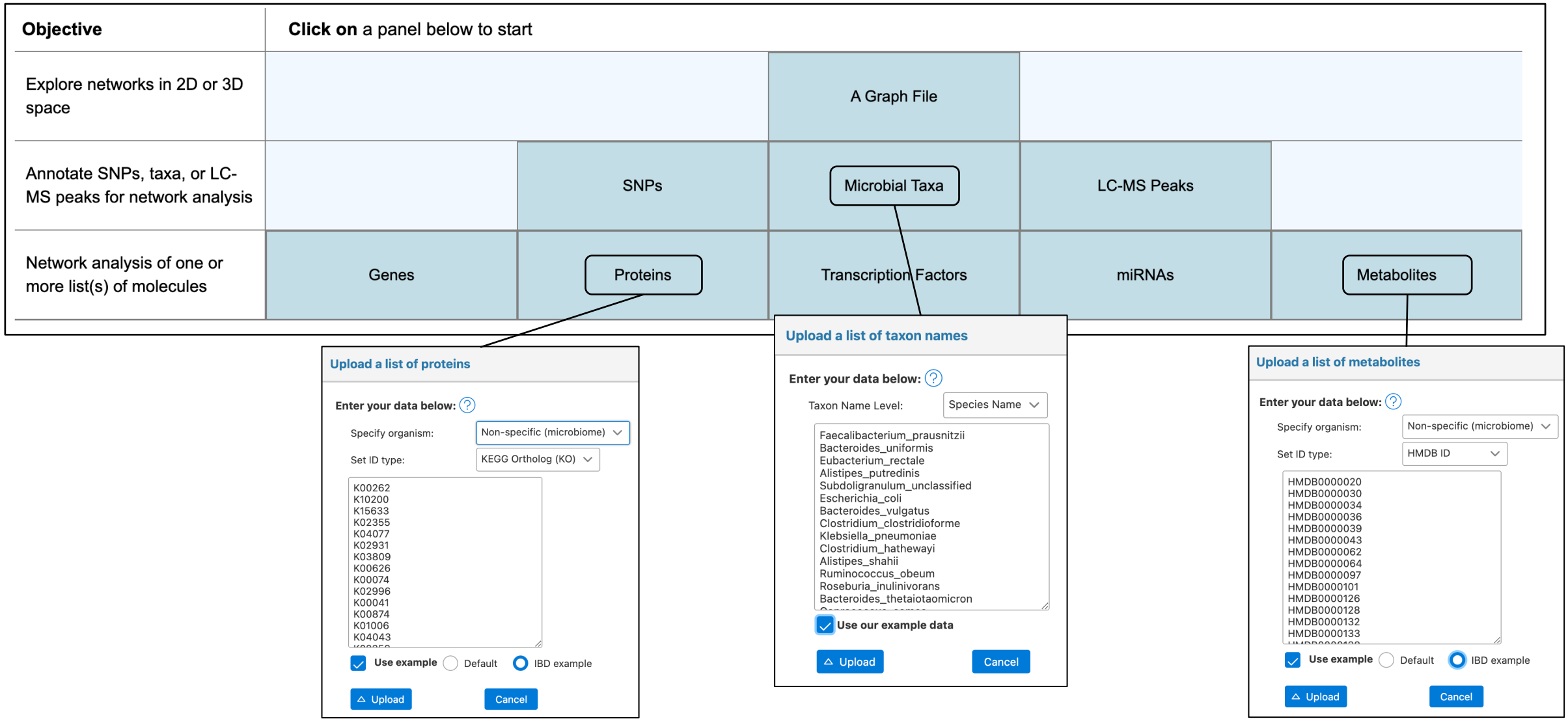
1. ***Database selection*** - for each input list, build an independent network by retrieving all interacting features from an appropriate database.
2. ***Network building*** – upon navigation to this page, all networks from the previous step are merged. There are additional tools for trimming the resulting network if it is too large.
3. ***Network analytics*** – use the 2D and 3D interactive network viewer to visualize and analyze the trimmed network.



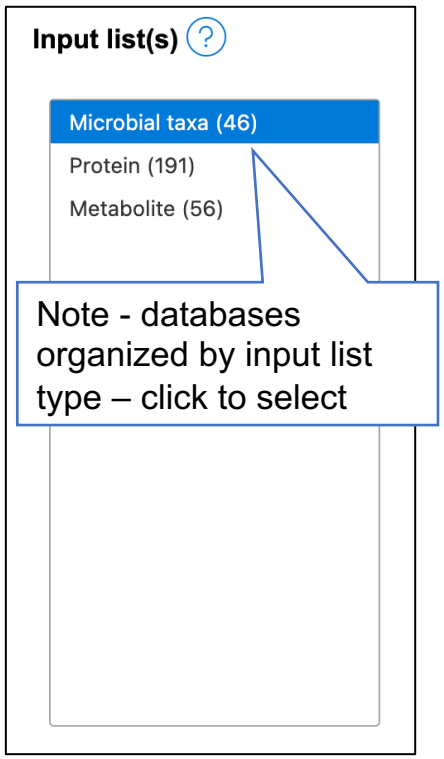
Note: This tutorial makes extensive use of AGORA genome-scale metabolic models (GEMs) to understand interactions between microbial taxa, metabolites, and proteins. GEMs are mathematical representations of metabolism, including reactions between genes, proteins, and metabolites. AGORA derives their GEMs from gut—microbiome data, using logistic regression to predict the potential of different taxa to produce different metabolites.

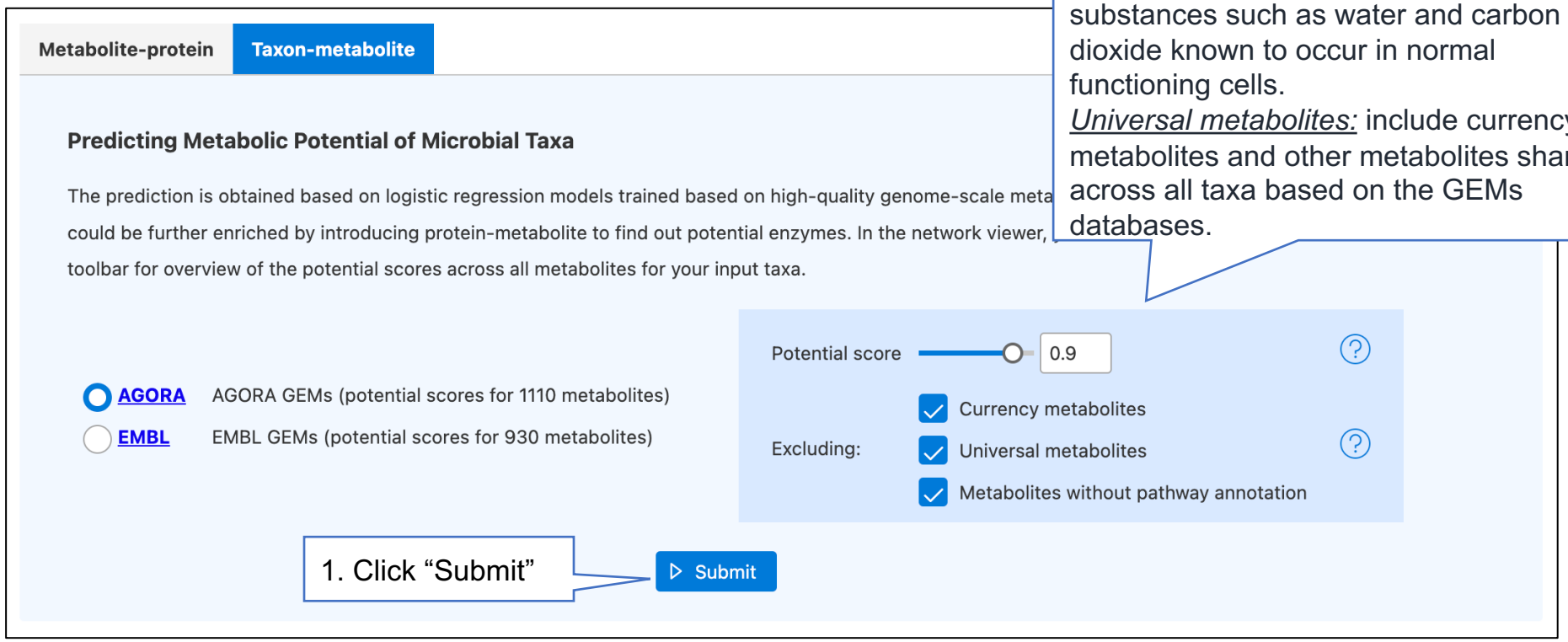
Data Upload

Select the example data as below, clicking “Upload” for each list. Make sure to select the *IBD example* for **Proteins** and **Metabolites**. Then, click the blue “Proceed” button below the list input.



Database Selection: Microbial Taxa

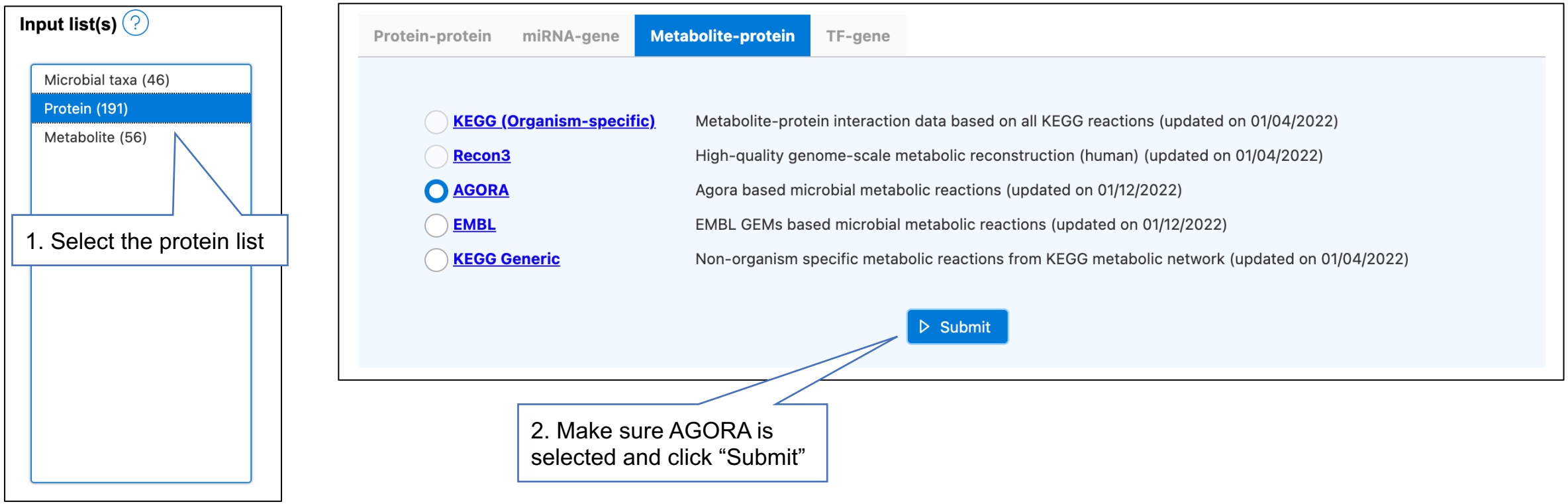




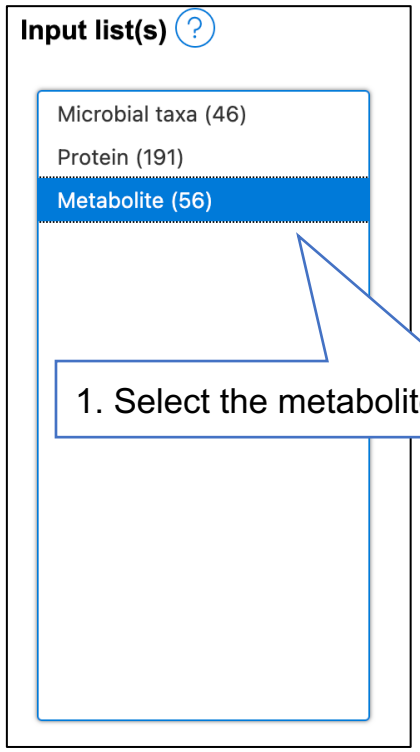
For AGORA GEMs, there are several parameters that can be adjusted. For this analysis we leave them as default, but here is more information in case you want to adjust later:

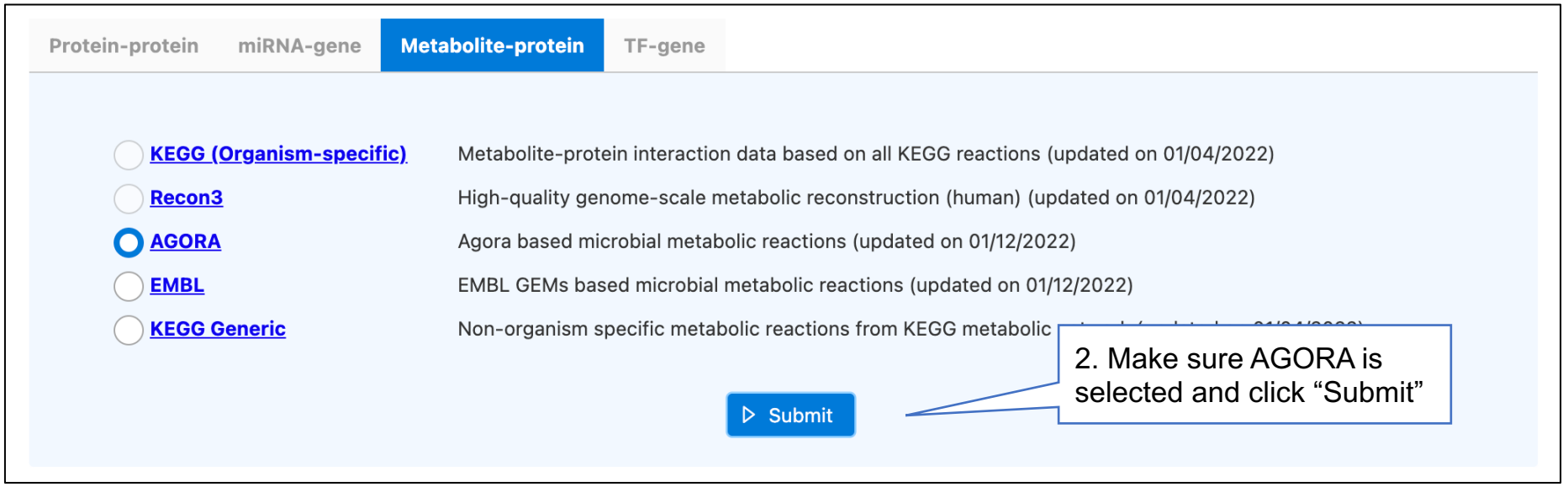
- **Threshold for potential score**: score over 0.5 indicates the taxon is more likely to produce the given metabolite and the increasing score value means the greater production possibility;
- **Exclude metabolites**: exclude currency metabolites, universal metabolites, and metabolites without functional information (pathway annotation) to prune network.

Database Selection: Proteins



Database Selection: Metabolites



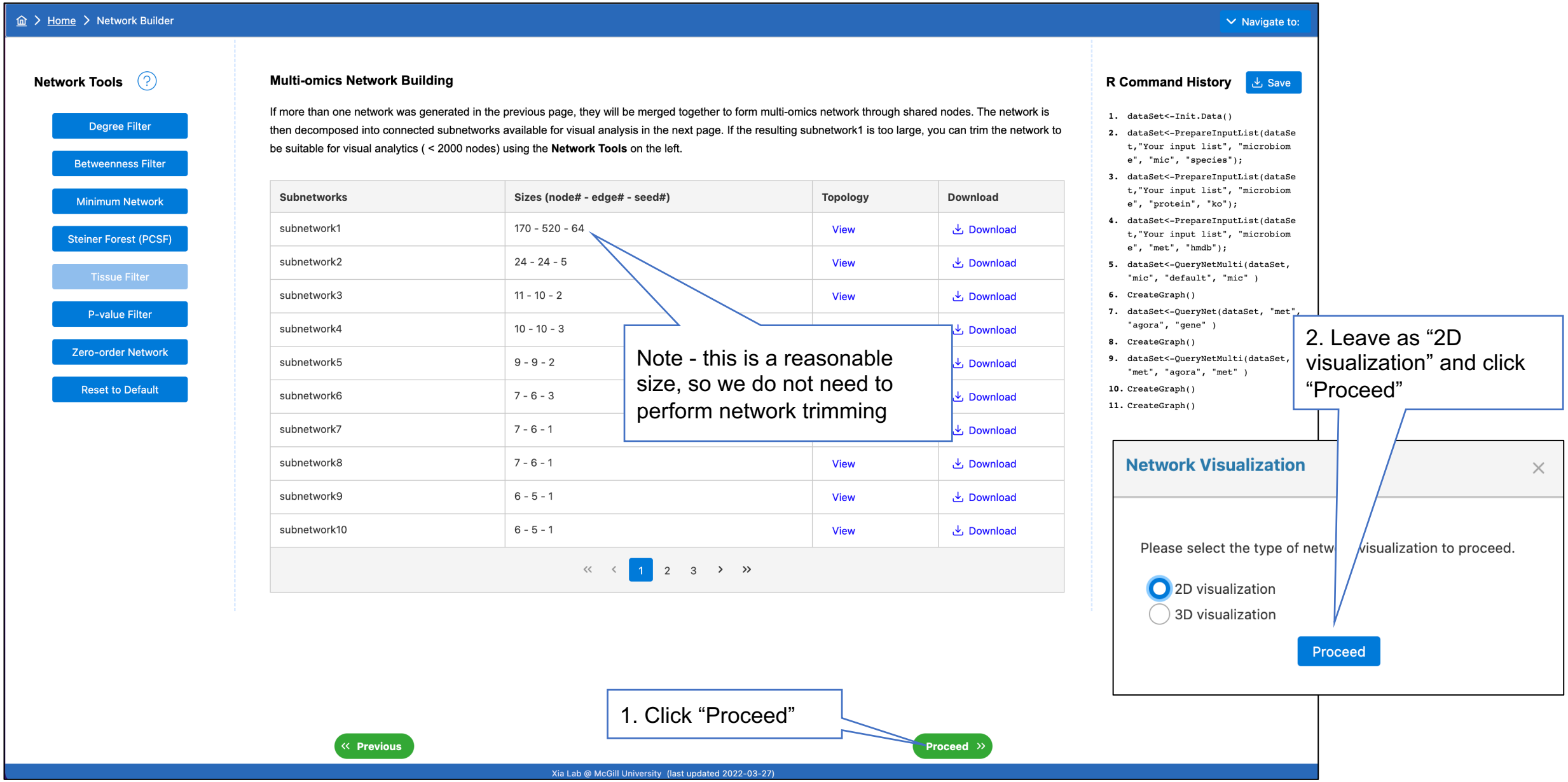


Note that the table at the bottom has been updated with each network:

Input Type	Network Type	Sizes (node# - edge# - seed#)	Browse	Download	Delete
Microbial taxa	Taxon-metabolite	55 - 379 - 28			
Gene	Metabolite-protein	156 - 146 - 54			
Metabolite	Metabolite-protein	109 - 114 - 24			

3. Click “Proceed”

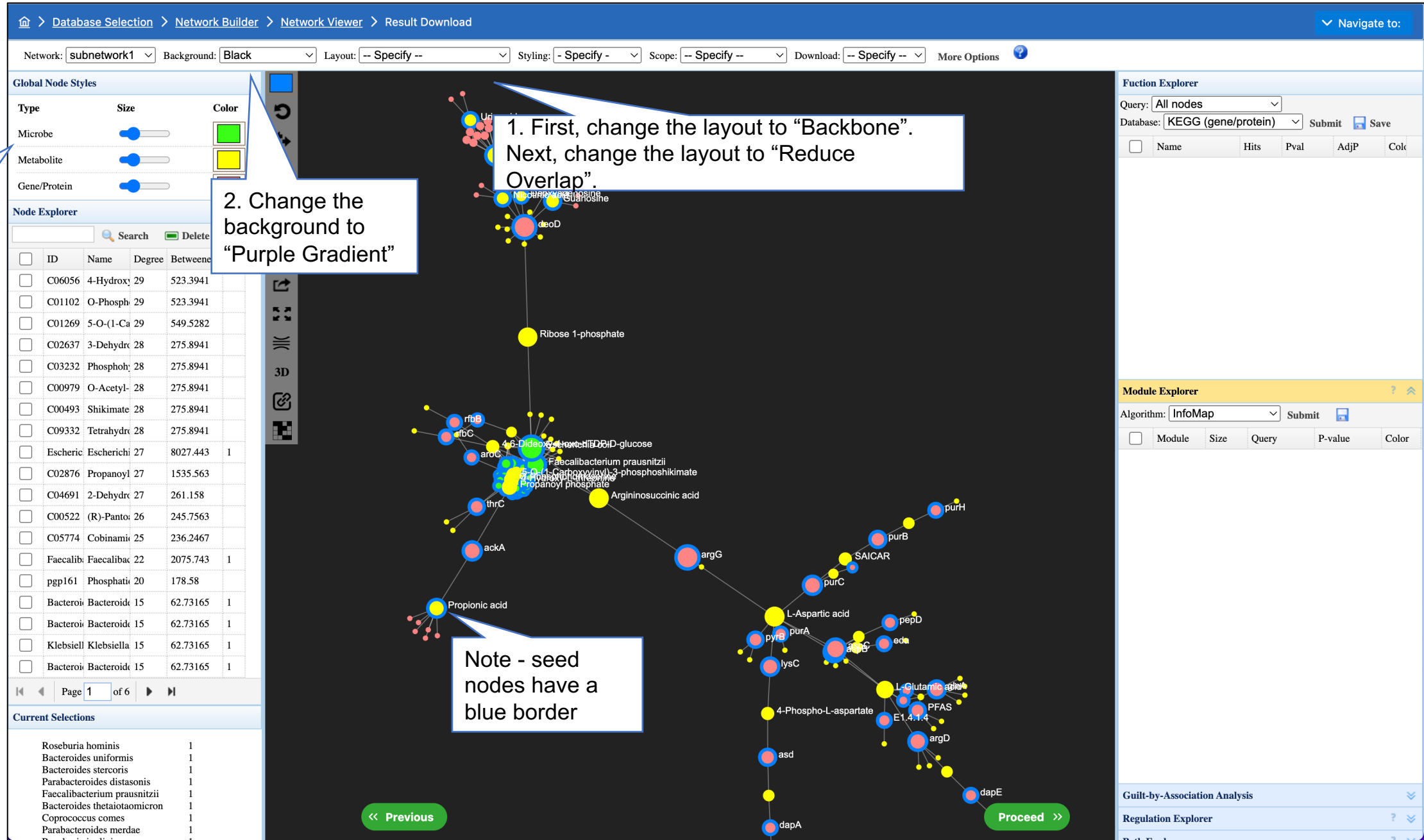
Network Building



Overview of 2D Network:

Here we perform some basic adjustments to make the structure more visible.

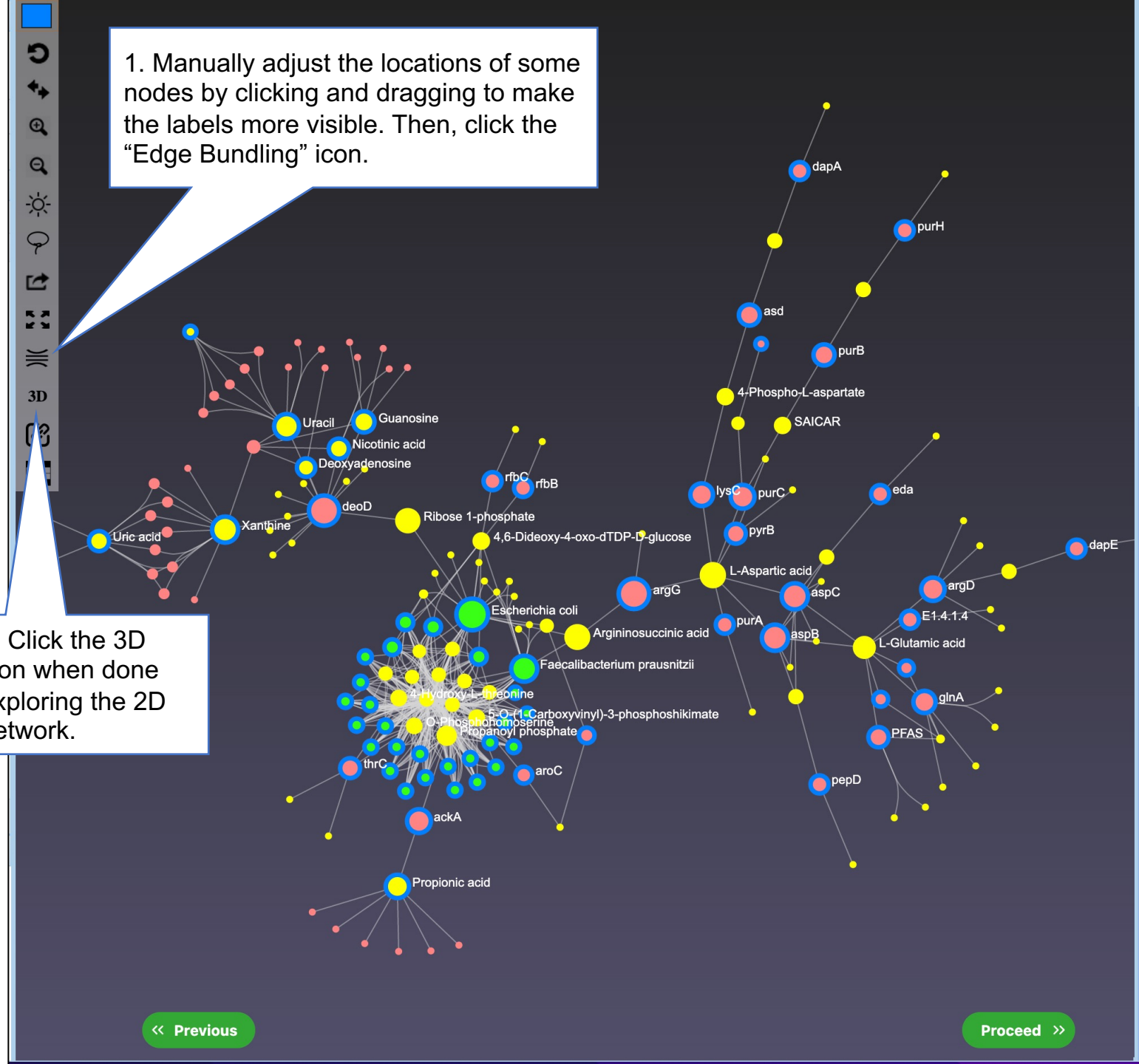
Note - each feature type is given a different color. You can change the color by clicking the boxes.



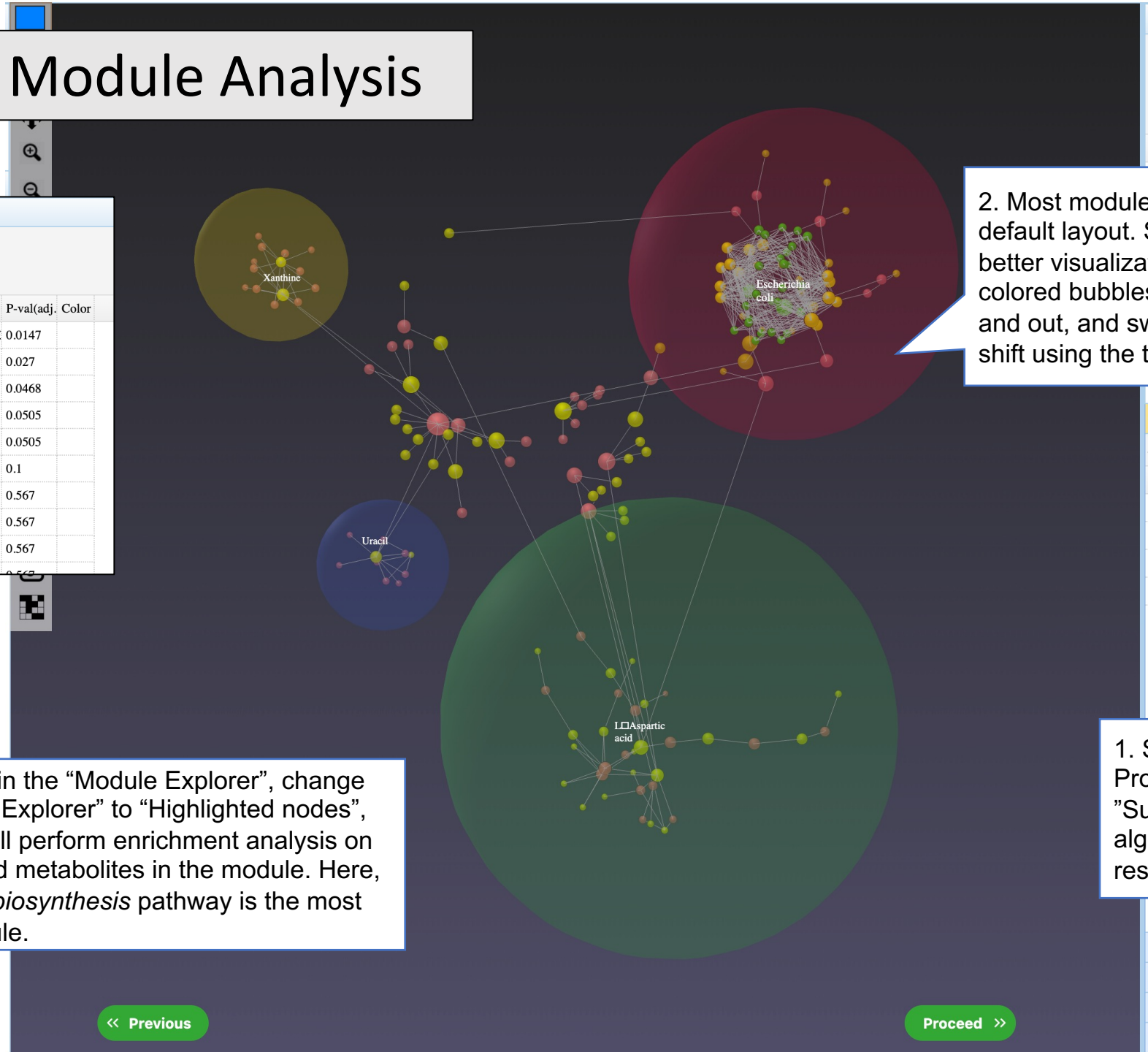
Key nodes in 2D Network

We see that there are two taxa with many connections: *Escherichia coli* and *Faecalibacterium prausnitzii*. In particular, they are closely connected to the large deoD and argG seed protein nodes via the predicted interacting Ribose 1-phosphate and Argininosuccinic acid metabolites respectively.

The deoD protein node is of particular interest - we see here that it is directly connected to five seed metabolites and indirectly connected to two more. Here, seed nodes were differentially abundant between dysbiotic and non-dysbiotic samples, and so it is interesting to see some predicted interactions between multiple perturbed 'omics layers.



3D Network: Module Analysis



Function Explorer

Query: Highlighted nodes
Database: KEGG (gene/protein) Submit

Name	Hits	P-val	P-val(adj.	Color
Streptomycin biosynthesis	2	0.000032	0.0147	
Biosynthesis of amino acids	3	0.000119	0.027	
Polyketide sugar unit biosynthesis	2	0.00031	0.0468	
Phenylalanine, tyrosine and tryptophan bios	2	0.000478	0.0505	
Biosynthesis of antibiotics	4	0.000557	0.0505	
Metabolic pathways	6	0.00133	0.1	
Vitamin B6 metabolism	1	0.0109	0.567	
Taurine and hypotaurine metabolism	1	0.0114	0.567	
Acarbose and validamycin biosynthesis	1	0.0128	0.567	

3. Select a single module in the “Module Explorer”, change the query in the “Function Explorer” to “Highlighted nodes”, and click “Submit”. This will perform enrichment analysis on the joint list of proteins and metabolites in the module. Here, we see the *Streptomycin biosynthesis* pathway is the most significant in the red module.

2. Most modules will overlap in the default layout. Spread them out for better visualization by clicking the colored bubbles and dragging. Zoom in and out, and switch between rotate and shift using the toolbar on the left.

Function Explorer

Query: All nodes
Database: KEGG (gene/protein) Submit

Name	Hits	P-val	P-val(ad

Module Explorer

Algorithm: Label Propagation Submit

Module	Size	P-value	Color
0	67	1.19e-25	
1	34	3.81e-12	
2	14	9.74e-06	
3	10	0.000393	

1. Select “Label Propagation” and click ”Submit”. Note the algorithm is stochastic so results may vary.

Guilt-by-Association Analysis
Regulation Explorer
Path Explorer
Batch Selection

The End