



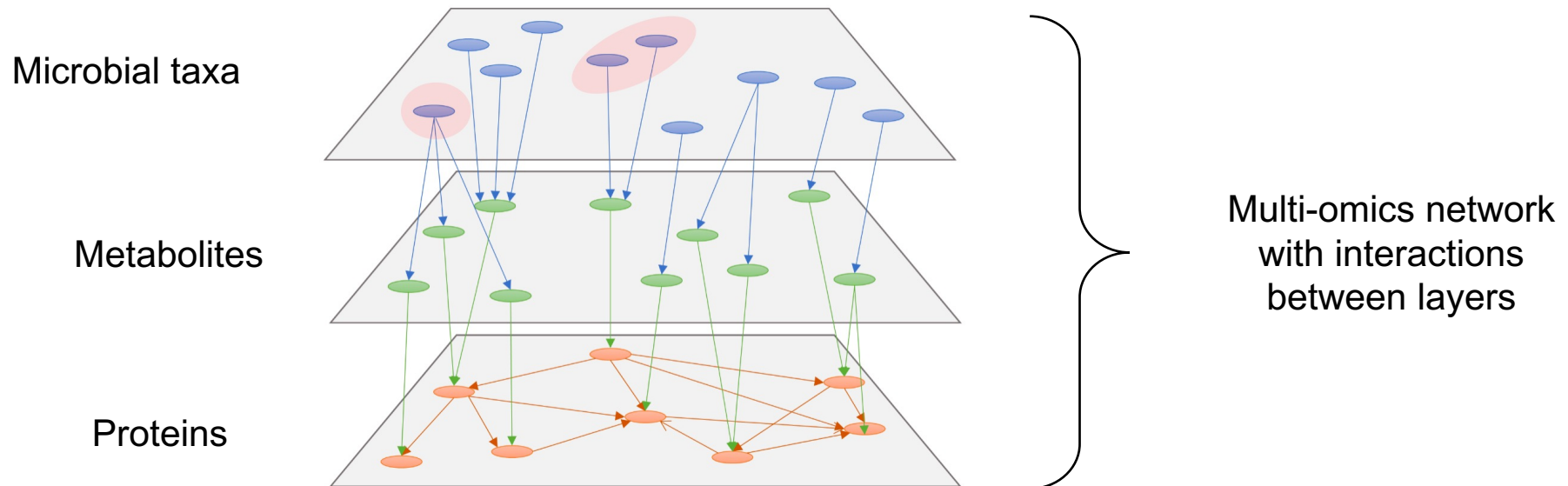
OmicNet Tutorial: IBD Case Study

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

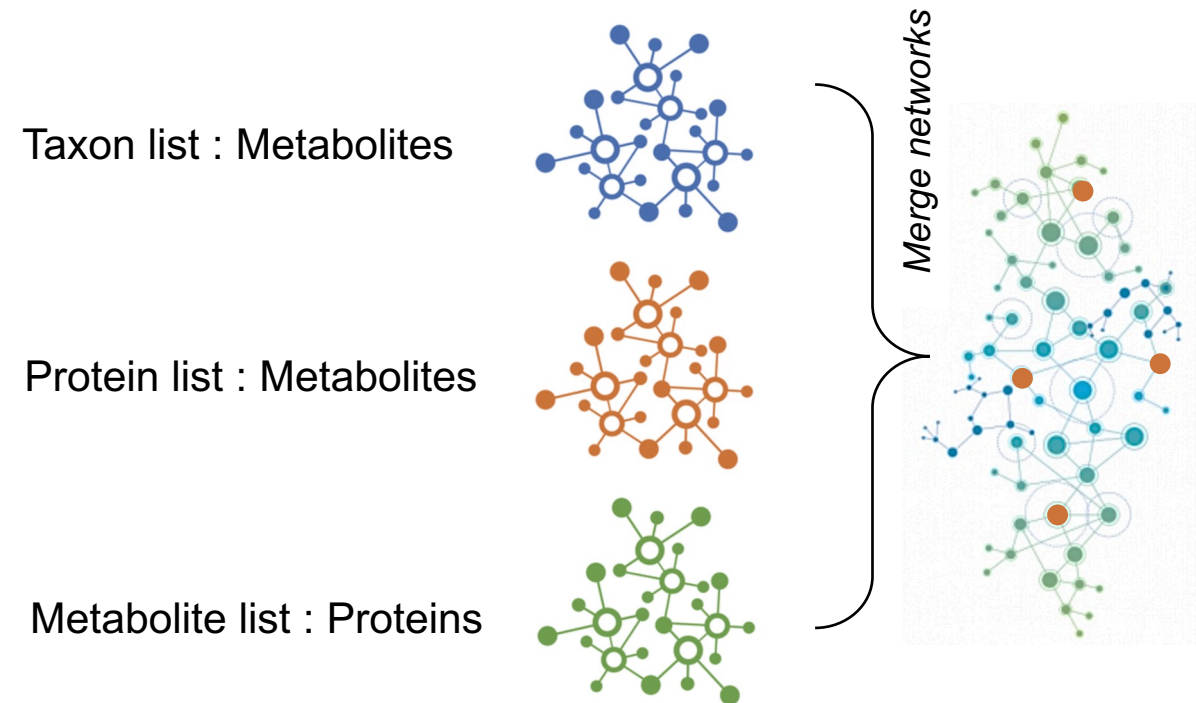
A recent study collected multi-omics data from stool samples from patients with Crohn's disease, a subtype of inflammatory bowel disease (IBD), to try and understand gut-microbiome drivers of dysbiosis. Here, we analyze lists of molecular features (metabolites, proteins, microbial taxa) that were significantly 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.

Objective	Click on a panel below to start				
Explore networks in 2D or 3D space			A Graph File		
Annotate SNPs, taxa, or LC-MS peaks for network analysis		SNPs	Microbial Taxa	LC-MS Peaks	
Network analysis of one or more list(s) of molecules	Genes	Proteins	Transcription Factors	miRNAs	Metabolites

Upload a list of proteins

Enter your data below: ?

Specify organism: Non-specific (microbiome) v

Set ID type: KEGG Ortholog (KO) v

K00262
K10200
K15633
K02355
K04077
K02931
K03809
K00626
K00074
K02996
K00041
K00874
K01006
K04043

Use example Default IBD example

Upload Cancel

Upload a list of taxon names

Enter your data below: ?

Taxon Name Level: Species Name v

Faecalibacterium_prausnitzii
Bacteroides_uniformis
Eubacterium_rectale
Alistipes_putredinis
Subdoligranulum_unclassified
Escherichia_coli
Bacteroides_vulgatus
Clostridium_clostridioforme
Klebsiella_pneumoniae
Clostridium_hathewayi
Alistipes_shahii
Ruminococcus_obeum
Roseburia_inulinivorans
Bacteroides_thetaiotaomicron

Use our example data

Upload Cancel

Upload a list of metabolites

Enter your data below: ?

Specify organism: Non-specific (microbiome) v

Set ID type: HMDB ID v

HMDB0000020
HMDB0000030
HMDB0000034
HMDB0000036
HMDB0000039
HMDB0000043
HMDB0000062
HMDB0000064
HMDB0000097
HMDB0000101
HMDB0000126
HMDB0000128
HMDB0000132
HMDB0000133

Use example Default IBD example

Upload Cancel

Database Selection: Microbial Taxa

Input list(s) ?

- Microbial taxa (46)
- Protein (191)
- Metabolite (56)

Note - databases organized by input list type – click to select

Metabolite-protein **Taxon-metabolite**

Predicting Metabolic Potential of Microbial Taxa

The prediction is obtained based on logistic regression models trained based on high-quality genome-scale meta could be further enriched by introducing protein-metabolite to find out potential enzymes. In the network viewer, toolbar for overview of the potential scores across all metabolites for your input taxa.

AGORA AGORA GEMs (potential scores for 1110 metabolites)

EMBL EMBL GEMs (potential scores for 930 metabolites)

Potential score 0.9 ?

Excluding:

- Currency metabolites ?
- Universal metabolites ?
- Metabolites without pathway annotation

1. Click "Submit"

Currency metabolites: abundant substances such as water and carbon dioxide known to occur in normal functioning cells.

Universal metabolites: include currency metabolites and other metabolites shared across all taxa based on the GEMs databases.

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

Input list(s) ?

- Microbial taxa (46)
- Protein (191)**
- Metabolite (56)

1. Select the protein list

Protein-protein miRNA-gene **Metabolite-protein** TF-gene

- [KEGG \(Organism-specific\)](#) Metabolite-protein interaction data based on all KEGG reactions (updated on 01/04/2022)
- [Recon3](#) High-quality genome-scale metabolic reconstruction (human) (updated on 01/04/2022)
- [AGORA](#) Agora based microbial metabolic reactions (updated on 01/12/2022)
- [EMBL](#) EMBL GEMs based microbial metabolic reactions (updated on 01/12/2022)
- [KEGG Generic](#) Non-organism specific metabolic reactions from KEGG metabolic network (updated on 01/04/2022)

2. Make sure AGORA is selected and click "Submit"

Database Selection: Metabolites

Input list(s) ?

- Microbial taxa (46)
- Protein (191)
- Metabolite (56)**

1. Select the metabolite list

Protein-protein miRNA-gene **Metabolite-protein** TF-gene

- [KEGG \(Organism-specific\)](#) Metabolite-protein interaction data based on all KEGG reactions (updated on 01/04/2022)
- [Recon3](#) High-quality genome-scale metabolic reconstruction (human) (updated on 01/04/2022)
- [AGORA](#) Agora based microbial metabolic reactions (updated on 01/12/2022)
- [EMBL](#) EMBL GEMs based microbial metabolic reactions (updated on 01/12/2022)
- [KEGG Generic](#) Non-organism specific metabolic reactions from KEGG metabolic

Submit

2. Make sure AGORA is selected and click "Submit"

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"

Proceed >>

Network Building

Home > Network Builder

Network Tools ?

- Degree Filter
- Betweenness Filter
- Minimum Network
- Steiner Forest (PCSF)
- Tissue Filter
- P-value Filter
- Zero-order Network
- Reset to Default

Multi-omics Network Building

If more than one network was generated in the previous page, they will be merged together to form multi-omics network through shared nodes. The network is then decomposed into connected subnetworks available for visual analysis in the next page. If the resulting subnetwork1 is too large, you can trim the network to be suitable for visual analytics (< 2000 nodes) using the **Network Tools** on the left.

Subnetworks	Sizes (node# - edge# - seed#)	Topology	Download
subnetwork1	170 - 520 - 64	View	Download
subnetwork2	24 - 24 - 5	View	Download
subnetwork3	11 - 10 - 2	View	Download
subnetwork4	10 - 10 - 3		Download
subnetwork5	9 - 9 - 2		Download
subnetwork6	7 - 6 - 3		Download
subnetwork7	7 - 6 - 1		Download
subnetwork8	7 - 6 - 1	View	Download
subnetwork9	6 - 5 - 1	View	Download
subnetwork10	6 - 5 - 1	View	Download

Note - this is a reasonable size, so we do not need to perform network trimming

1. Click "Proceed"

2. Leave as "2D visualization" and click "Proceed"

```
R Command History Save
```

- dataSet<-Init.Data()
- dataSet<-PrepareInputList(dataSet,"Your input list", "microbiome", "mic", "species");
- dataSet<-PrepareInputList(dataSet,"Your input list", "microbiome", "protein", "ko");
- dataSet<-PrepareInputList(dataSet,"Your input list", "microbiome", "met", "hmb");
- dataSet<-QueryNetMulti(dataSet,"mic", "default", "mic");
- CreateGraph()
- dataSet<-QueryNet(dataSet, "met", "agora", "gene");
- CreateGraph()
- dataSet<-QueryNetMulti(dataSet,"met", "agora", "met");
- CreateGraph()
- CreateGraph()

Network Visualization

Please select the type of network visualization to proceed.

2D visualization

3D visualization

Proceed

<< Previous

Proceed >>

Xia Lab @ McGill University (last updated 2022-03-27)

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.

Database Selection > Network Builder > Network Viewer > Result Download

Network: subnetwork1 Background: Black Layout: -- Specify -- Styling: - Specify - Scope: -- Specify -- Download: -- Specify -- More Options

Global Node Styles

Type	Size	Color
Microbe	<input type="range"/>	<input type="checkbox"/>
Metabolite	<input type="range"/>	<input type="checkbox"/>
Gene/Protein	<input type="range"/>	<input type="checkbox"/>

Node Explorer

ID	Name	Degree	Betweenness
<input type="checkbox"/>	C06056 4-Hydrox	29	523.3941
<input type="checkbox"/>	C01102 O-Phosph	29	523.3941
<input type="checkbox"/>	C01269 5-O-(1-Ca	29	549.5282
<input type="checkbox"/>	C02637 3-Dehydr	28	275.8941
<input type="checkbox"/>	C03232 Phosphoh	28	275.8941
<input type="checkbox"/>	C00979 O-Acetyl-	28	275.8941
<input type="checkbox"/>	C00493 Shikimate	28	275.8941
<input type="checkbox"/>	C09332 Tetrahydr	28	275.8941
<input type="checkbox"/>	Escheric Escherichi	27	8027.443 1
<input type="checkbox"/>	C02876 Propanoyl	27	1535.563
<input type="checkbox"/>	C04691 2-Dehydr	27	261.158
<input type="checkbox"/>	C00522 (R)-Panto	26	245.7563
<input type="checkbox"/>	C05774 Cobinami	25	236.2467
<input type="checkbox"/>	Faecalib: Faecalibac	22	2075.743 1
<input type="checkbox"/>	pgp161 Phosphati	20	178.58
<input type="checkbox"/>	Bacteroid Bacteroid	15	62.73165 1
<input type="checkbox"/>	Bacteroid Bacteroid	15	62.73165 1
<input type="checkbox"/>	Klebsiell Klebsiella	15	62.73165 1
<input type="checkbox"/>	Bacteroid Bacteroid	15	62.73165 1

Page 1 of 6

Current Selections

Roseburia hominis	1
Bacteroides uniformis	1
Bacteroides stercoris	1
Parabacteroides distasonis	1
Faecalibacterium prausnitzii	1
Bacteroides thetaiotaomicron	1
Coprococcus comes	1
Parabacteroides merdae	1

1. First, change the layout to "Backbone".
Next, change the layout to "Reduce Overlap".

2. Change the background to "Purple Gradient"

Note - seed nodes have a blue border

Navigation: << Previous Proceed >>

Fuction Explorer

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

Name	Hits	Pval	AdjP	Colc
------	------	------	------	------

Module Explorer

Algorithm: InfoMap Submit

Module	Size	Query	P-value	Color
--------	------	-------	---------	-------

Guilt-by-Association Analysis

Regulation Explorer

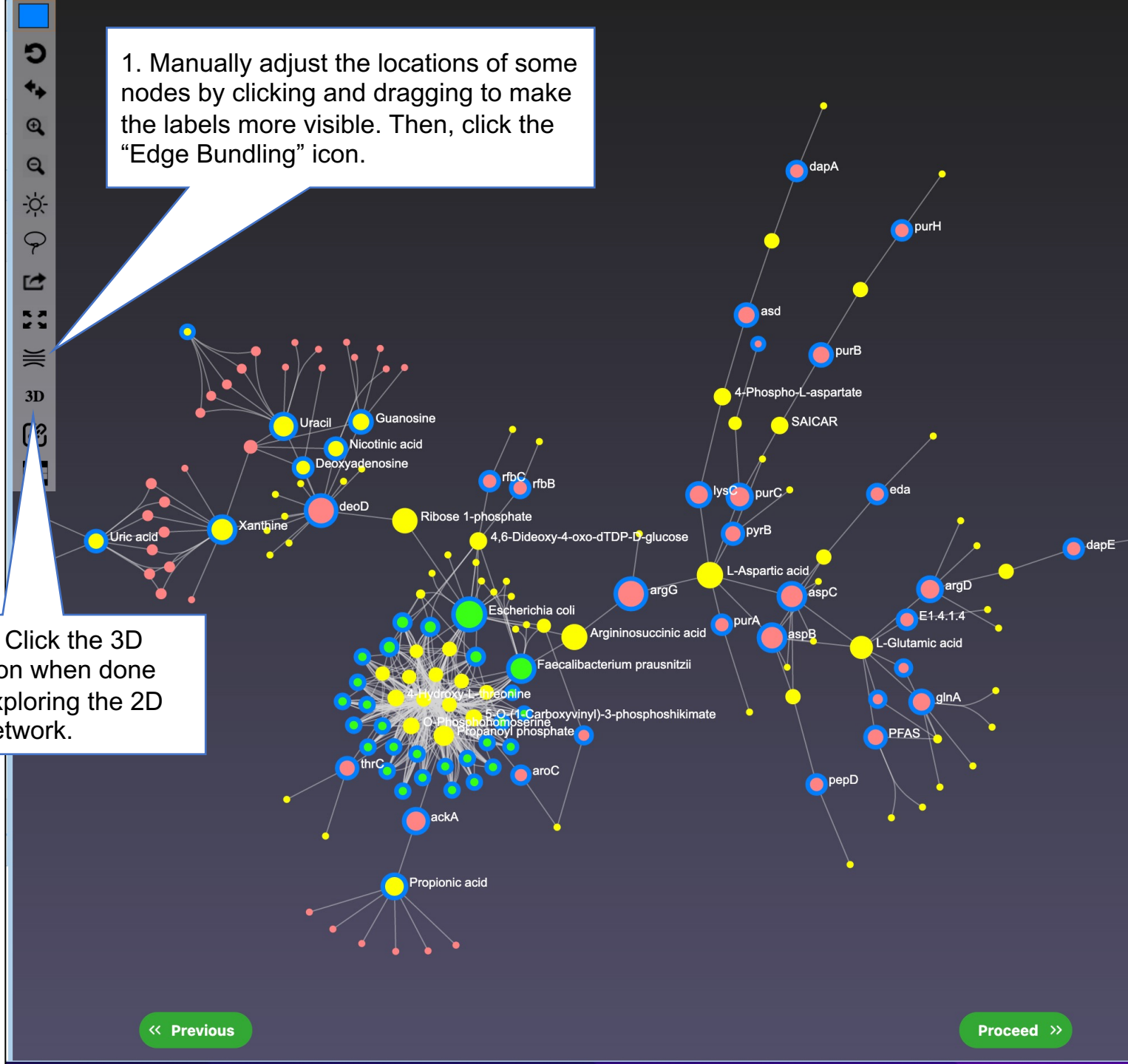
Key nodes in 2D Network

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.

1. Manually adjust the locations of some nodes by clicking and dragging to make the labels more visible. Then, click the "Edge Bundling" icon.

2. Click the 3D icon when done exploring the 2D network.

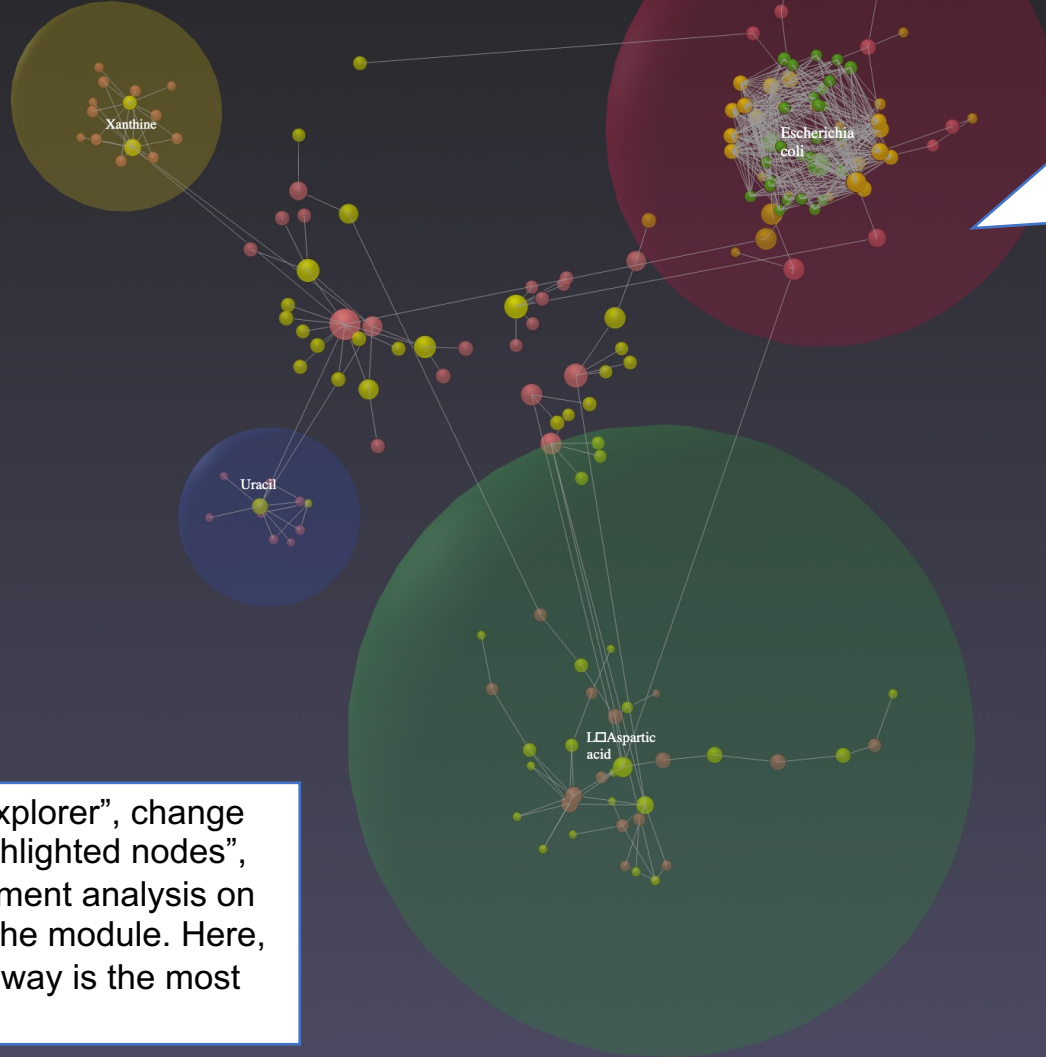


3D Network: Module Analysis

Function Explorer

Query: Database: 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 bio	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	



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.

Module Explorer

Algorithm: Submit

Module	Size	P-value	Color
<input type="checkbox"/> 0	67	1.19e-25	Red
<input type="checkbox"/> 1	34	3.81e-12	Green
<input type="checkbox"/> 2	14	9.74e-06	Yellow
<input type="checkbox"/> 3	10	0.000393	Blue

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.

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

The End
