

National Institute of Allergy and Infectious Diseases

ACE Mali

# Metabolomics & Microbiome

22 April 2024

NIH  
NIAID



National Institute of  
Allergy and  
Infectious Diseases

**Lauren Krausfeldt & Poorani Subramanian**  
Bioinformatics & Computational Biosciences Branch/OCICB

# Ask Questions Here

- Please feel free to interrupt us to ask questions
  - by voice in the Zoom
  - [https://etherpad.p2pu.org/p/Metabolomics\\_microbiome](https://etherpad.p2pu.org/p/Metabolomics_microbiome)
- Also please interrupt us if we are speaking too quickly or you don't understand something!

# About Us

## Lauren Krausfeldt

- Ph.D. Microbiology
- Joined NIAID 2022
- BCBB Metagenomics group
- SME and pipeline dev for Nephele

## Poorani Subramanian

- Ph.D. Applied Math & Scientific Computing
- Joined NIAID in 2017
- BCBB's Metagenomics group
- Lead of microbiome app Nephele  
([nephele.niaid.nih.gov](http://nephele.niaid.nih.gov))

# About BCBB

Science Support: Metagenomics/Microbiome

Bioinformatics and Computational Biosciences Branch (BCBB)

National Institute of Allergies and Infectious Diseases (NIAID)

National Institute of Health (NIH)

Bethesda, MD, USA

**Find out more about what is offered at the BCBB:**

<https://www.niaid.nih.gov/research/metagenomics-bioinformatics-and-computational-biosciences-branch>

Contact: [bioinformatics@niaid.nih.gov](mailto:bioinformatics@niaid.nih.gov)

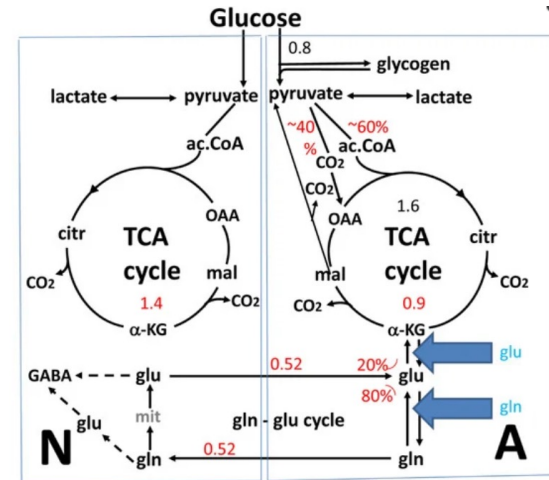


# How do we study the microbiome?

- 16S rRNA sequencing
  - Studying marker gene 16S rRNA – identify who is there
- Metagenomics
  - Studying “all of the DNA” - Identify who is there and functional potential
- Metatranscriptomics
  - Studying “all of the RNA” – Identify who is active and the transcript produced by microbiome
- Proteomics
  - Studying “all of the proteins” – identify what enzymes are being produced and the actual function of the microbiome

# Metabolomics and the microbiome

- Allows us to study the actual **function** of microbiome
  - Characterize the small molecules (<2000 Da) produced by the microbiome
  - Indicative of cellular processes that are occurring, how the microbiome is responding to certain stimuli or conditions
  - Can also be used to study exogenous molecules, like xenobiotics
- Targeted or untargeted
  - Targeted can be more quantitative with abundances
    - Example: Lipids
  - Untargeted is relative abundances
    - Example: All polar or non-polar



# Why is this valuable information?

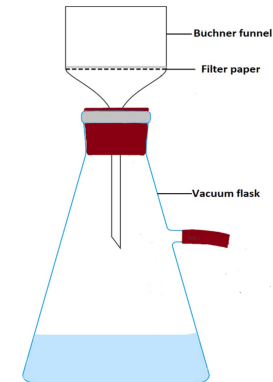
- Metabolites produced by bacteria are known to influence host metabolism and physiology
  - Describes phenotypes that influence the host
- Mechanistic understanding of microbial phenotypes
- Metabolome is a snapshot of what the cell is doing in that moment
- Useful for detecting novel metabolites for drug development
- Can help address variability across studies in taxonomic variability
  - Geography, diet, genetics affect microbiome composition
  - Functional redundancy: different taxa, same function

# How do we study the metabolome of the a microbiome?

- Sample collection
- Data integration
  - Comparing trends in microbiome to metabolome
  - Pathway inference
  - Correlation networks
  - Stable isotope probing
  - 3D cartography

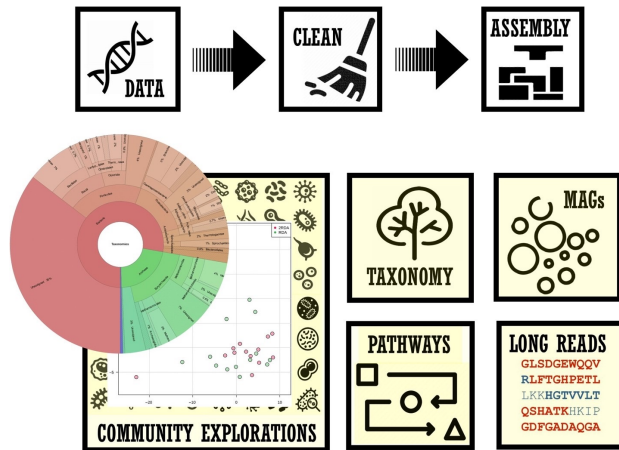
# Sample collection

- Centrifugation
  - Stool, saliva, blood, bile, plant, soil
- Vacuum filtration
  - Water
- Quenching: stop metabolism and reactions
  - Flash freeze
  - Cold solvents like methanol
- Extraction
  - Cold solvents and cold environment to prevent metabolism from occurring and preserving metabolites



# Separately analyze the metabolomics and other 'omics

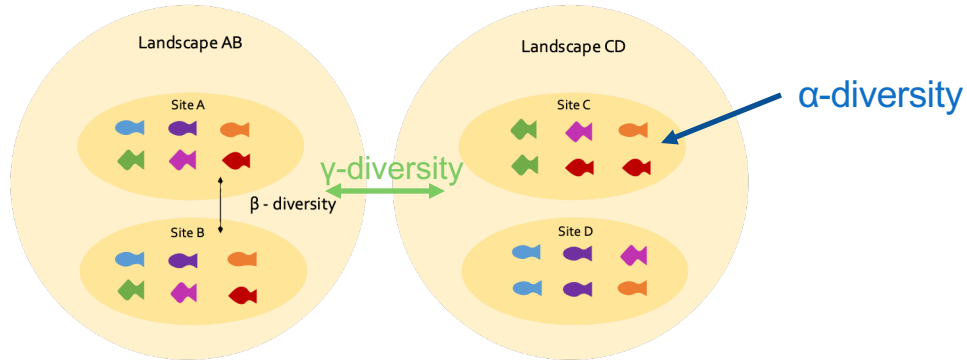
- Characterize taxonomy and optionally functional genes from metagenomic sequence data
  - Usually involves processing the data in some way and then comparing sequences against a database



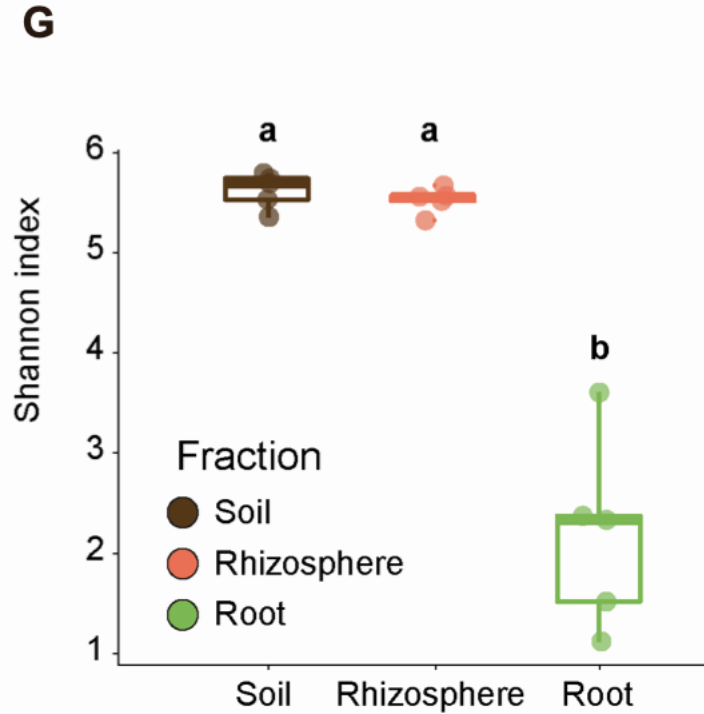
- <https://nephele.niaid.nih.gov> our application for microbiome analysis
  - WGSA2 pipeline for metagenomics Illumina
  - DADA2 for 16S Illumina
- For ONT
  - Nanopore QC
- Characterize metabolites and estimate abundances separately (as you have learned in your course)

# (Bio) Diversity

- **Species richness:** number of species in a single sample
- **Alpha diversity:** species richness weighted in some way by the abundance (count or biomass) of each species in a single sample
- **Beta diversity:** a comparison of diversity between samples



# Alpha diversity

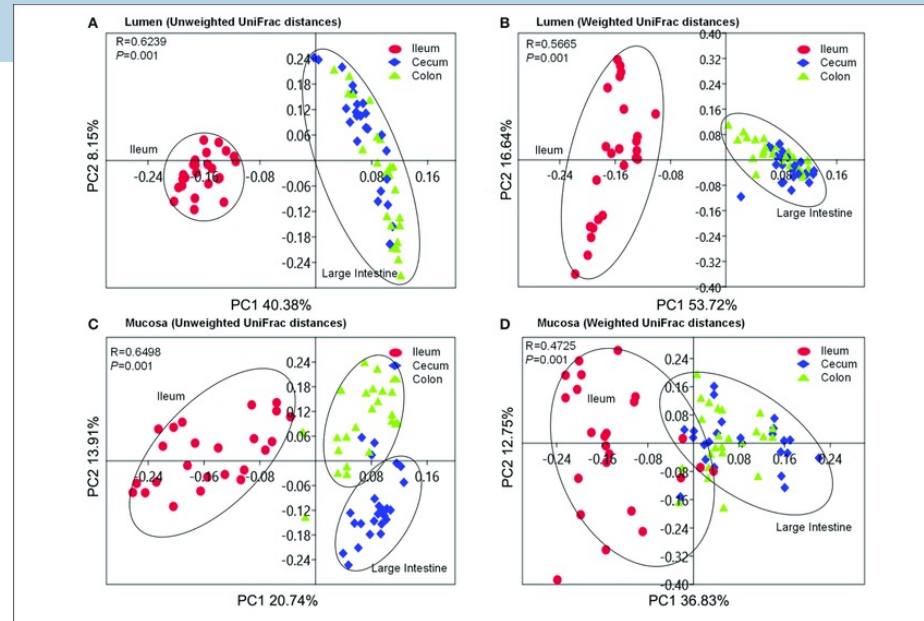


- Alpha diversity: May or may not take into account abundance
- Popular indices:
  - Shannon
  - Simpson/Inverse Simpson
  - ACE/Chao1 species richness
- [Sugar transporters spatially organize microbiota colonization along the longitudinal root axis of Arabidopsis](#)
  - Loo, Duran, et. al. Cell Host and Microbe. Mar 2024



# Ordination / Beta diversity

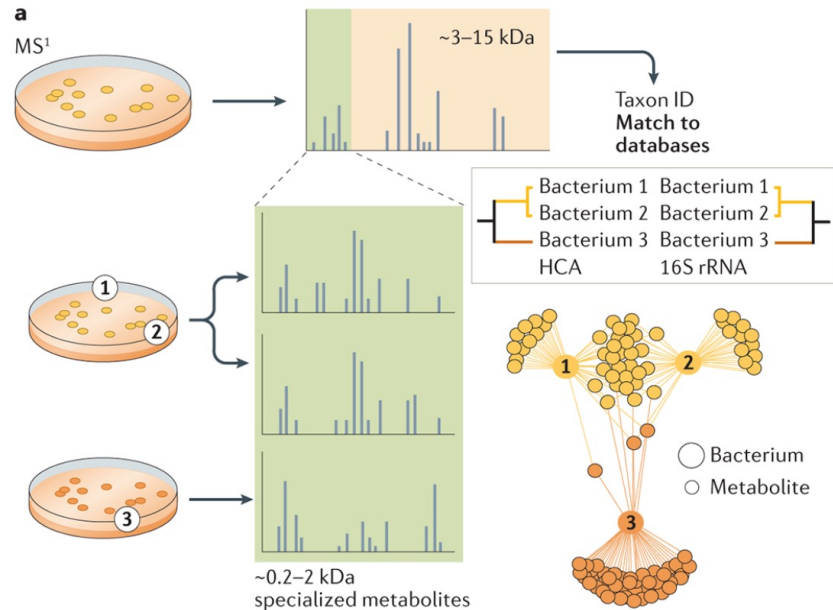
- Ordination: orders quantities in a lower dimensional space so similar samples are closer together and dissimilar ones are farther apart
  - Usually calculate a distance between samples to quantify this similarity
  - PCoA, nMDS
  - Packages for alpha & beta diversity: QIIME2 or in R using vegan/ampvis2/phyloseq libraries
  - Web-based tools:
    - Microbiomedb.org
    - <http://galaxy.biobakery.org> from the [Huttenhower Lab](#)



- PERMANOVA to look for factors (confounding, etc) that have an effect
  - R vegan adonis2 function
  - DivNet R package

# Biotyping

- Phenotypic typing based on metabolic characteristics expressed by an isolate; referred to as ‘biotypes’
- Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) biotyping
- Bauermeister, Anelize, et al. “Mass Spectrometry-Based Metabolomics in Microbiome Investigations.” *Nature Reviews. Microbiology*, vol. 20, no. 3, Mar. 2022, pp. 143–60. PubMed Central, <https://doi.org/10.1038/s41579-021-00621-9>.



# Biotyping – targeted characterization

- Culture/Isolate organisms in the lab and then profile the metabolites in order to build the database
  - Look for organism/bacterial-specific metabolites (as opposed to those produced by host)
- Pathogen detection – successful (?) especially for eukaryotic microbes
  - amr expression
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9340741/>
- One downside – a lot of these databases are proprietary
  - Made by the mass spec instrument manufacturers
- A Few open dbs that contain broad range of genera:
  - [Robert Koch Institute](#)
  - [GNPS spectral libraries](#) (down currently), [MassBank](#) (Japan, EU and North America), NIST, and [METLIN](#)

# Pathway inference

- Identify which pathway a metabolite comes from
  1. Manually search databases
    - KEGG (Kyoto Encyclopedia of Genes and Genomes)
    - <https://www.genome.jp/kegg/kegg2.html>
    - What is KEGG?



KEGG  Search Help  
» Japanese

**KEGG Home**  
[Release notes](#)  
[Current statistics](#)

**KEGG Database**  
[KEGG overview](#)  
[KEGG mapping](#)  
[Color codes](#)

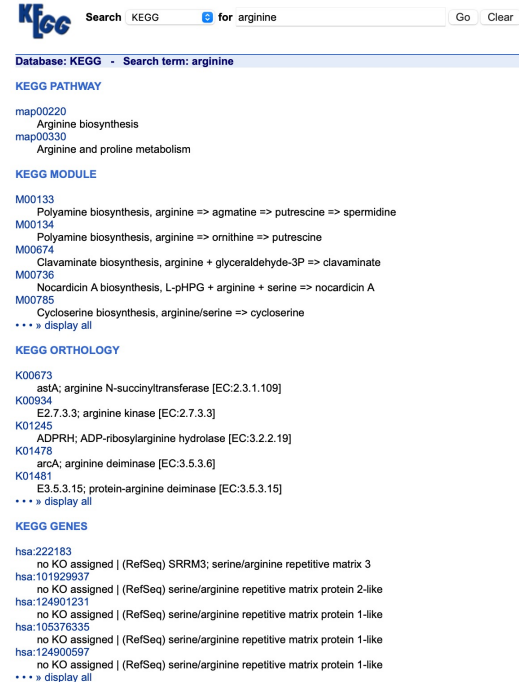
## KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG 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. See [Release notes](#) (April 1, 2024) for new and updated features.

**New feature:** [Genome alignment](#) (gene order alignment) tool is released

# Pathway inference

- KEGG (Kyoto Encyclopedia of Genes and Genomes)
  - Pathway metabolite is found
  - modules (reactions and pathways)
  - KEGG orthology, specific genes
  - Compound information
  - Enzyme information
  - And more!



The screenshot shows the KEGG database search interface. The search bar contains the text "Search KEGG for arginine" with "Go" and "Clear" buttons. Below the search bar, the results are organized into several sections:

- Database: KEGG - Search term: arginine**
- KEGG PATHWAY**
  - map00220 Arginine biosynthesis
  - map00330 Arginine and proline metabolism
- KEGG MODULE**
  - M00133 Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine
  - M00134 Polyamine biosynthesis, arginine => ornithine => putrescine
  - M00674 Clavaminat biosynthesis, arginine + glyceraldehyde-3P => clavaminat
  - M00736 Nocardicin A biosynthesis, L-pHPG + arginine + serine => nocardicin A
  - M00785 Cycloserine biosynthesis, arginine/serine => cycloserine
  - ... » display all
- KEGG ORTHOLOGY**
  - K00673 **astA**; arginine N-succinyltransferase [EC:2.3.1.109]
  - K00934 E2.7.3.3; arginine kinase [EC:2.7.3.3]
  - K01245 **ADPRH**; ADP-ribosylarginine hydrolase [EC:3.2.2.19]
  - K01478 **arcA**; arginine deiminase [EC:3.5.3.6]
  - K01481 E3.5.3.15; protein-arginine deiminase [EC:3.5.3.15]
  - ... » display all
- KEGG GENES**
  - hsa:222183 no KO assigned | (RefSeq) SRRM3; serine/arginine repetitive matrix 3
  - hsa:101929937 no KO assigned | (RefSeq) serine/arginine repetitive matrix protein 2-like
  - hsa:124901231 no KO assigned | (RefSeq) serine/arginine repetitive matrix protein 1-like
  - hsa:105376335 no KO assigned | (RefSeq) serine/arginine repetitive matrix protein 1-like
  - hsa:124900597 no KO assigned | (RefSeq) serine/arginine repetitive matrix protein 1-like
  - ... » display all

# Pathway inference with KEGG



## KEGG Mapper – Search

### KEGG Mapper

Reconstruct  
Search  
Color  
Join  
Convert ID  
Assign KO

### KEGG Web Apps

Map coloring GUI  
Map coloring URL

### KEGG Syntax

Genome alignment  
Taxonomy mapping

### KEGG Annotation

Ortholog table

### KEGG2

KEGG

### Search tool

The Search tool searches various KEGG objects, including genes, KOs, EC numbers, metabolites and drugs, against KEGG pathway maps and other network entities. Found objects are marked in red.

**Search mode:**  Reference  hsa  other org

### Enter query KEGG identifiers

C00062

Examples:

Select

hsa mode

diseasegene.txt (human disease genes)  
drugtarget.txt (drug targets)  
mabtarget.txt (drug targets, MAB only)

Or upload file:  no file selected

(to extract K/C/G/D/R/RC numbers)

- Include "same as" objects  
 Include aliases (for hsa and other org modes)

## KEGG Mapper Search Result

Pathway (15)

Brite (0)

Module (6)

Network (2)

Disease (0)

Sort by the pathway list

Show matched objects

- hsa00220 Arginine biosynthesis - Homo sapiens (human) (1)
- hsa05230 Central carbon metabolism in cancer - Homo sapiens (human) (1)
- hsa01100 Metabolic pathways - Homo sapiens (human) (1)
- hsa05146 Amoebiasis - Homo sapiens (human) (1)
- hsa05022 Pathways of neurodegeneration - multiple diseases - Homo sapiens (human) (1)
- hsa00330 Arginine and proline metabolism - Homo sapiens (human) (1)
- hsa04974 Protein digestion and absorption - Homo sapiens (human) (1)
- hsa01230 Biosynthesis of amino acids - Homo sapiens (human) (1)
- hsa00970 Aminoacyl-tRNA biosynthesis - Homo sapiens (human) (1)
- hsa00470 D-Amino acid metabolism - Homo sapiens (human) (1)
- hsa02010 ABC transporters - Homo sapiens (human) (1)
- hsa05014 Amyotrophic lateral sclerosis - Homo sapiens (human) (1)
- hsa04148 Efferocytosis - Homo sapiens (human) (1)
- hsa04150 mTOR signaling pathway - Homo sapiens (human) (1)
- hsa05142 Chagas disease - Homo sapiens (human) (1)

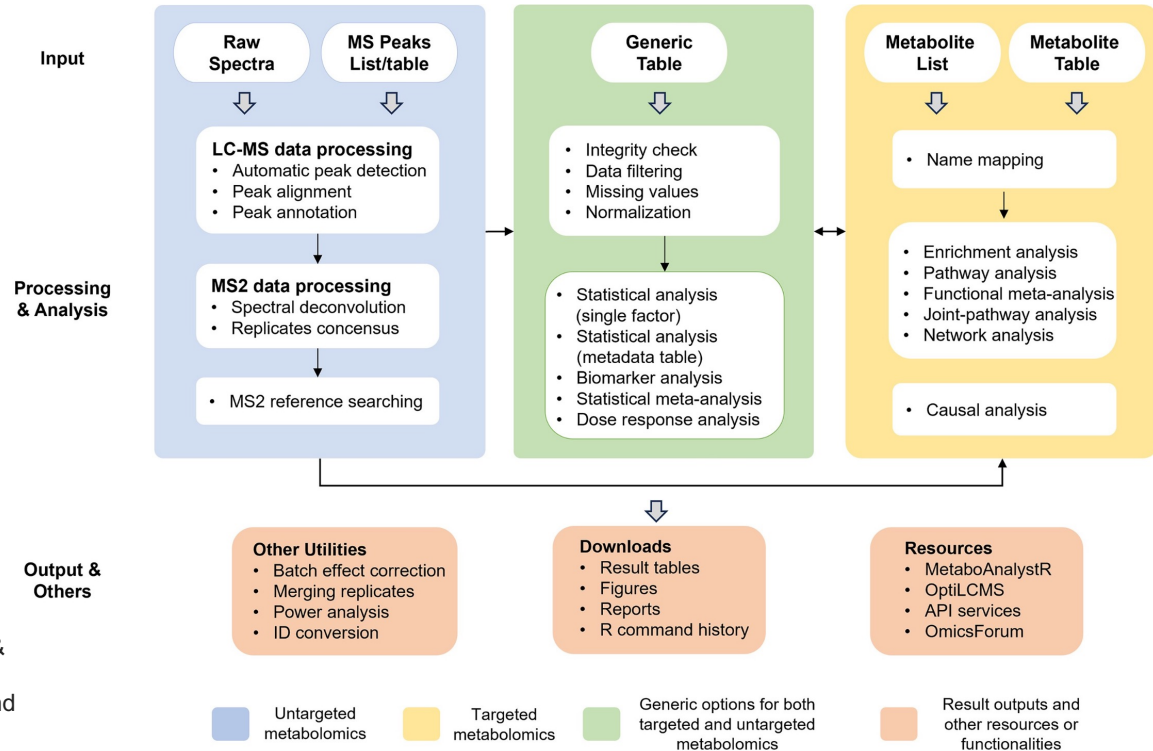
# Pathway inference

- Other databases:
  - Human metabolome project
  - LIPID MAPS
  - The gut microbiome-metabolome dataset collection
  - <https://reactome.org>
- Repositories and databases
  - Metabolights (EMBL-EBI)
  - Metabolomics Workbench (NCBI)

# Pathway inference

Tool for automated pathway inference

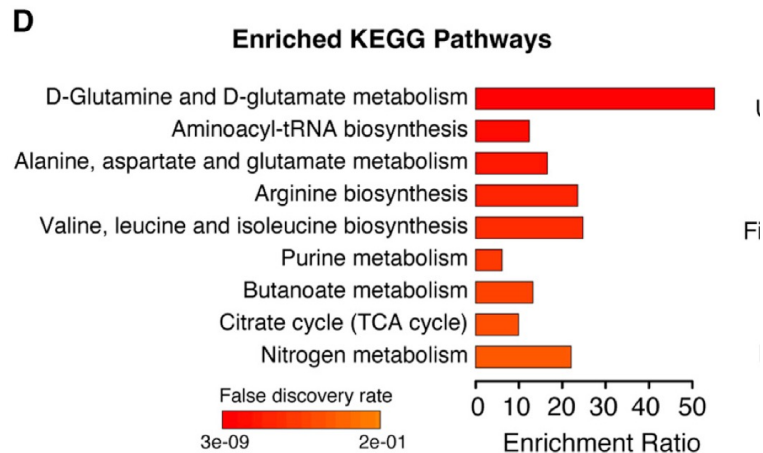
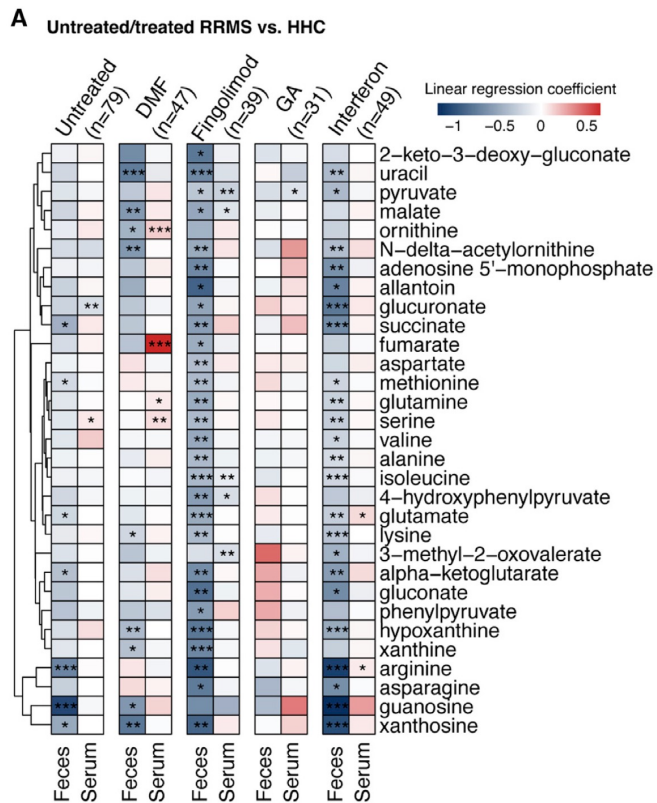
- Metaboanalyst (<https://www.metaboanalyst.ca>)



Pang, Z., Lu, Y., Zhou, G., Hui, F., Xu, L., Viau, C., ... & Xia, J. (2024). MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Research*, gkae253.



# Example: Metabolic changes in the microbiome associated with relapsing-remitting multiple sclerosis treatments



Zhou, X., Baumann, R., Gao, X., Mendoza, M., Singh, S., Sand, I. K., ... & Baranzini, S. E. (2022). Gut microbiome of multiple sclerosis patients and paired household healthy controls reveal associations with disease risk and course. *Cell*, 185(19), 3467-3486.

# Pathway inference and data integration

- Use information from other -omics to link pathway to taxa or group of taxa in the microbiome
  - Metagenomics: which microbes have the gene or pathway for a specific reaction
  - ★ Metatranscriptomics: Which microbes are expressing genes for a specific reaction
  - ★ Proteomics: Which microbes are expressing proteins for a specific reaction

# Are the metabolites microbial or host derived?

- Often host contamination in samples
  - Having sequence data from host is helpful to increase confidence
  - Tools that help to distinguish source or contribution of metabolite
    - MIMOSA: Model-based Integration of Metabolite Observations and Species Abundances
    - AMON: annotation of metabolite origins via networks

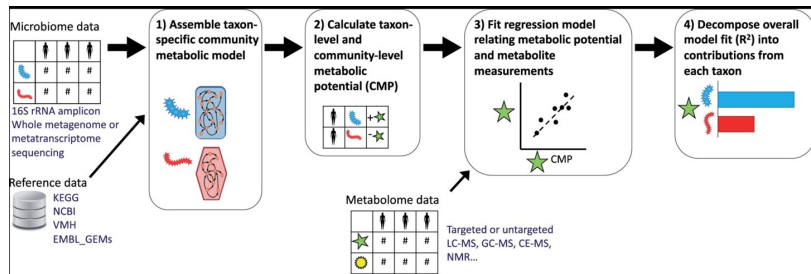
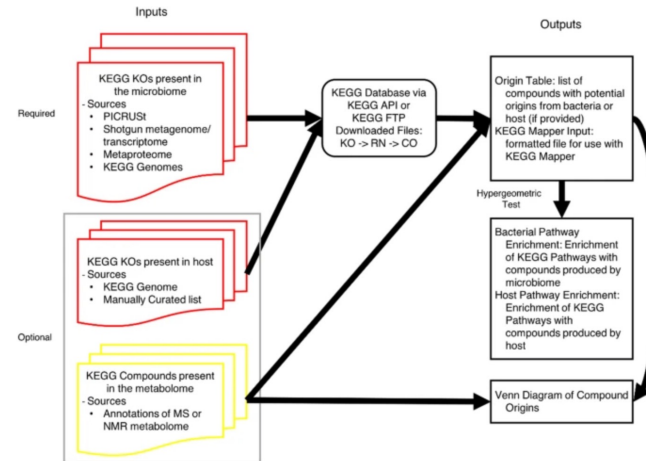


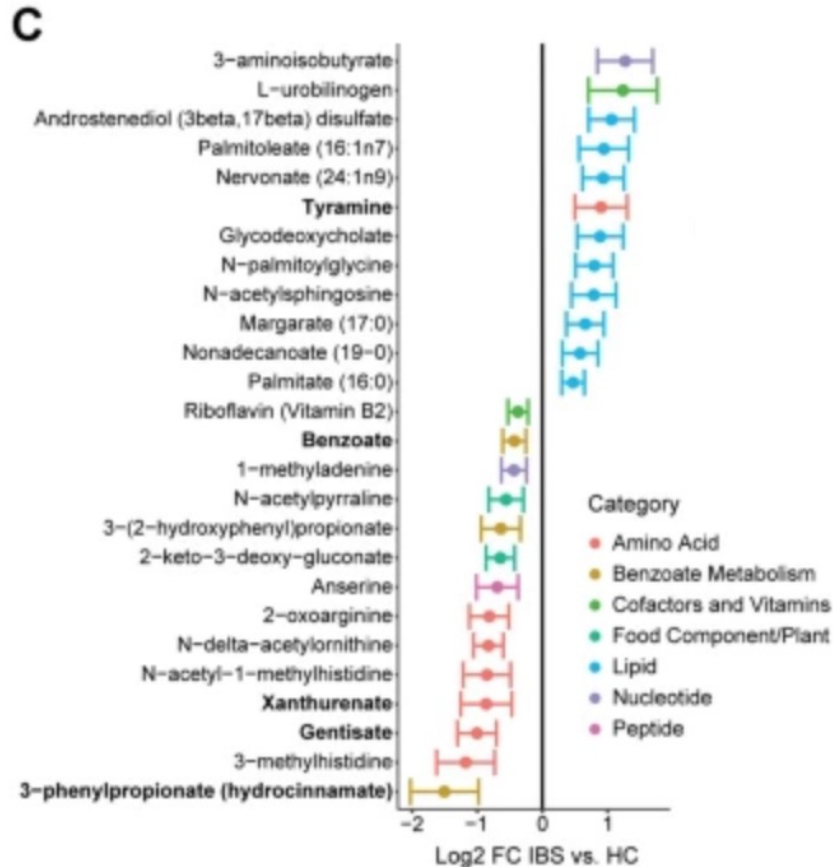
Fig. 1



Noecker, C., Eng, A., Muller, E., & Borenstein, E. (2022). MIMOSA2: a metabolic network-based tool for inferring mechanism-supported relationships in microbiome-metabolome data. *Bioinformatics*, 38(6), 1615-1623.

Shaffer, M., Thurimella, K., Quinn, K., Doenges, K., Zhang, X., Bokatzian, S., ... & Lozupone, C. A. (2019). AMON: annotation of metabolite origins via networks to integrate microbiome and metabolome data. *BMC bioinformatics*, 20, 1-11.

# Example



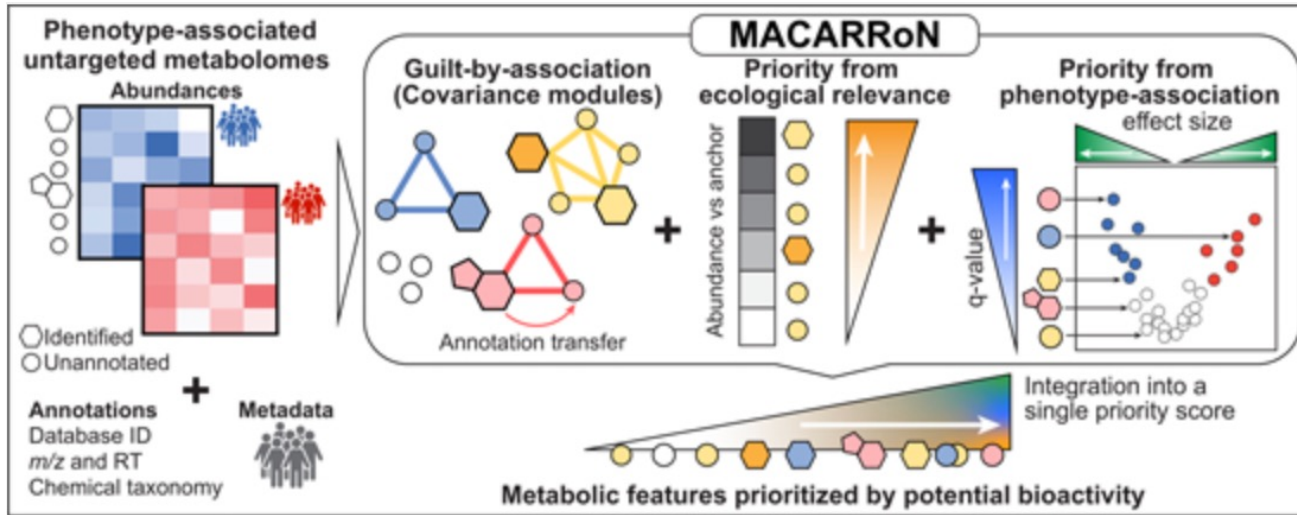
Metabolites abundances were identified associated with microbial function potential in IBS patients compared to healthy controls using MIMOSA2 using “community wide metabolic potential score” (relative capacity for the microbes to perform that function)

Jacobs, J. P., Lagishetty, V., Hauer, M. C., Labus, J. S., Dong, T. S., Toma, R., ... & Mayer, E. A. (2023). Multi-omics profiles of the intestinal microbiome in irritable bowel syndrome and its bowel habit subtypes. *Microbiome*, 11(1), 5.

# Exploring uncharacterized metabolites as biomarkers

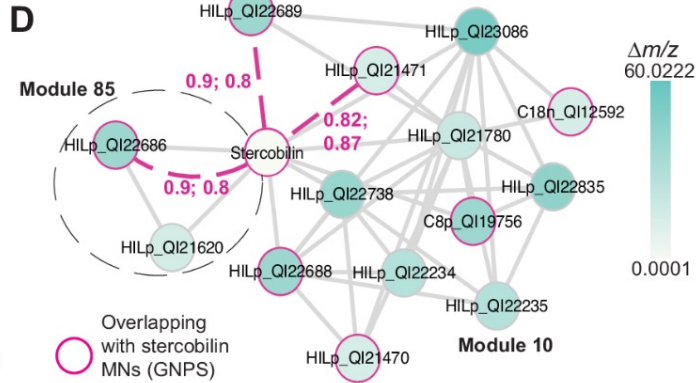
## ■ MACARRoN

(<https://huttenhower.sph.harvard.edu/macarron/>)

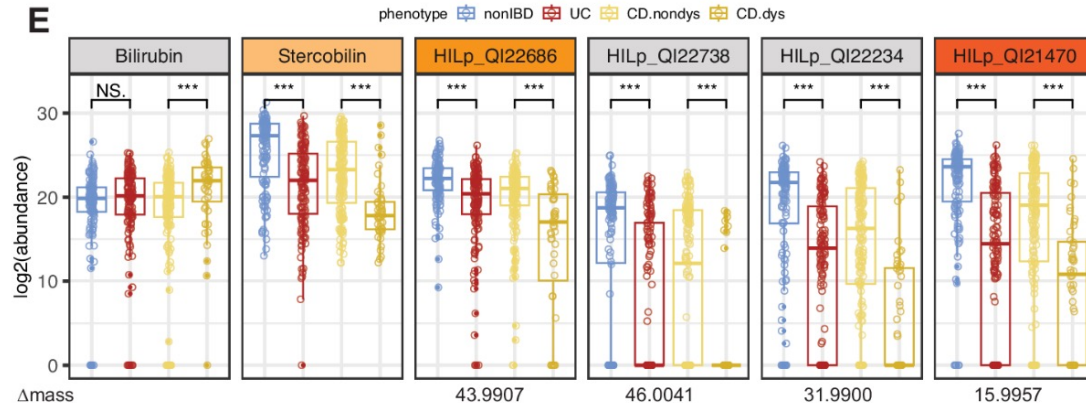


Bhosle, A., Bae, S., Zhang, Y., Chun, E., Avila-Pacheco, J., Geistlinger, L., ... & Huttenhower, C. (2024). Integrated annotation prioritizes metabolites with bioactivity in inflammatory bowel disease. *Molecular Systems Biology*, 20(4), 338-361.

# Identified novel metabolites depleted in IBD



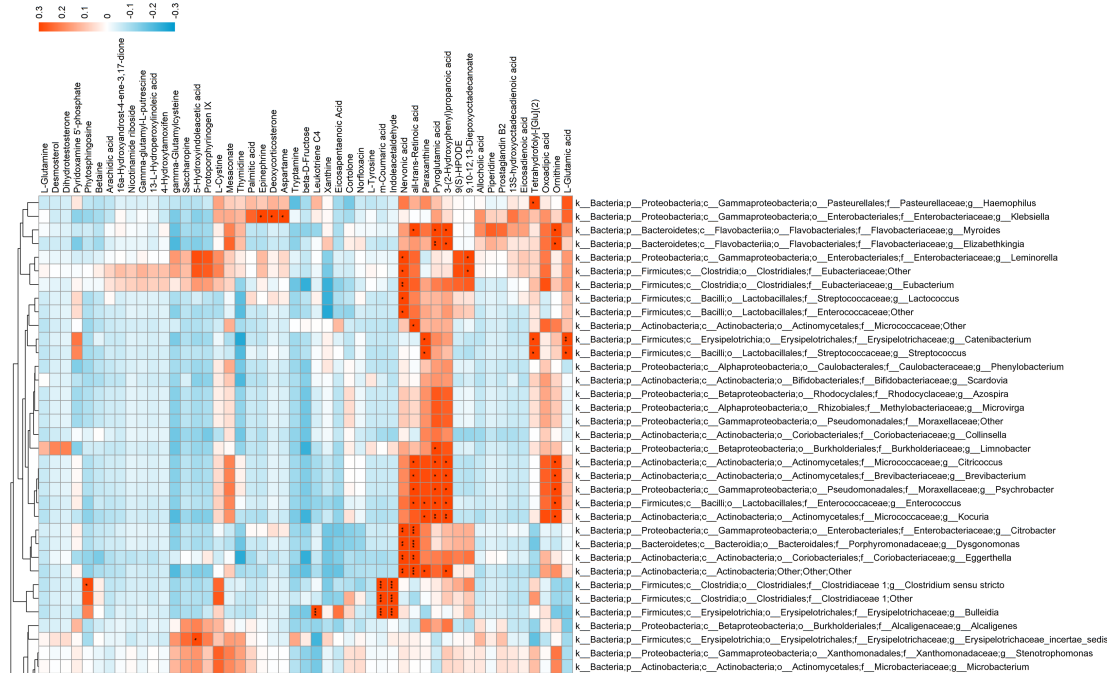
Unidentified metabolic covariates with stercobilin



# Data Integration: Correlations

- Statistical correlations between abundances/presence-absence of metabolites and taxa
  - Correlations between single variables – single taxa and single metabolite – **one value for every pair of variables**
- Non-parametric correlations (and correlation tests)
  - Spearman, Kendall
  - `cor` & `cor.test` functions
  - Good for microbiome and metabolome produced from community
- Parametric correlations – normally distributed data
  - Pearson
  - Good for metabolomics from single organism
- Visualization
  - Correlation matrix heatmap or network or clustered tree
  - biplot

# Correlation heatmap



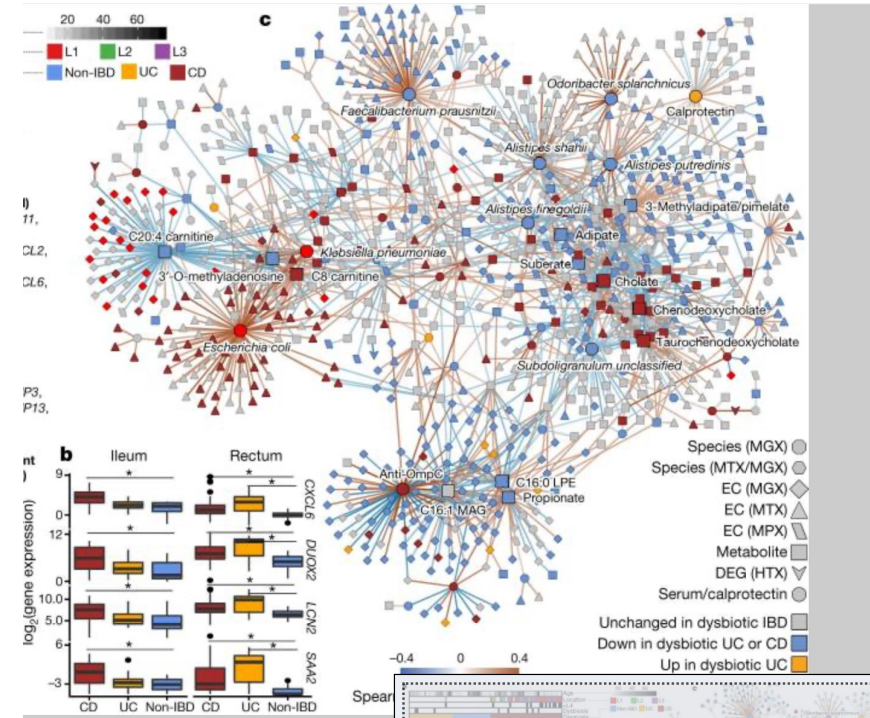
- This is a crop of a big heatmap correlating all bacteria metabolites
- Could also correlate only differentially abundant variables

Qian, Xiang, et al. "Integrated Microbiome, Metabolome, and Proteome Analysis Identifies a Novel Interplay among Commensal Bacteria, Metabolites and Candidate Targets in Non-small Cell Lung Cancer." *Clinical and Translational Medicine*, vol. 12, no. 6, June 2022, p. e947. *PubMed Central*, <https://doi.org/10.1002/ctm2.947> .



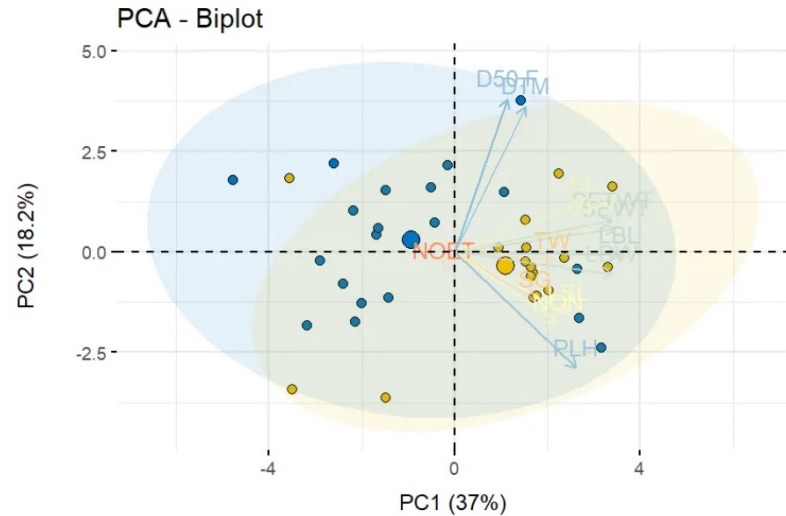
# Correlation network

- Structure in the network
  - Look for clusters/cliques
- Tools for making networks and doing analysis on those networks
  - Cytoscape
  - Omicsnet
  - ggnet R package
- Lloyd-Price, Jason, et al. “Multi-Omics of the Gut Microbial Ecosystem in Inflammatory Bowel Diseases.” *Nature*, vol. 569, no. 7758, 2019, pp. 655–62. PubMed Central, <https://doi.org/10.1038/s41586-019-1237-9>.



# Correlation biplot

- PCoA of the "main" dataset
- And then correlation distances and angles with the secondary dataset
- Different methods
  - CCA – Canonical Correlation Analysis
- Tools/R packages
  - miXomics
  - ggbiplot
  - princomp
- <https://medium.com/@victorallan/an-intuitive-guide-to-pca-in-r-a-step-by-step-tutorial-with-beautiful-visualization-examples-73bce5ee02e5>

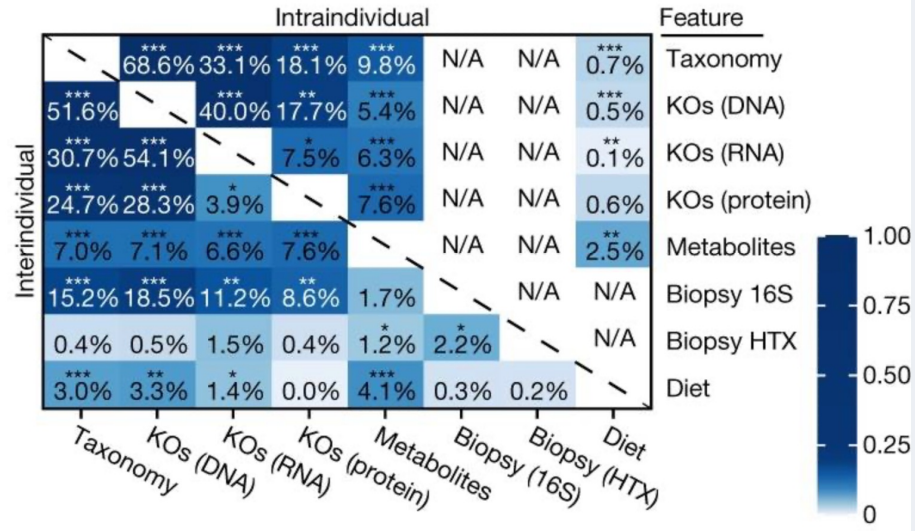


# Mantel test for correlations

- Correlation test between 2 matrices of variables (as opposed to 2 variables)
- Often we use 2 distance matrices to see if the relationships/similarity of **microbiome taxa** between samples correlate to the similarity between **metabolites** of samples
  - Do samples with similar **taxon profiles** also have similar **metabolomics profiles**?
- Ape R package function [`mantel.test`](#)
  - Alternative: ade4 package function [`mantel.rtest`](#)

# Mantel test: example

- Lloyd-Price paper  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6650278/>
- Bray-Curtis distance used for all data types
- Tutorial:  
<https://jkzorz.github.io/2019/07/08/mantel-test.html>



# Other methods of data integration

## Procrustes

Similar to the Mantel test  
Dimensionality reduction

[microbiomeanalyst.ca](https://microbiomeanalyst.ca)

## Statistical regression

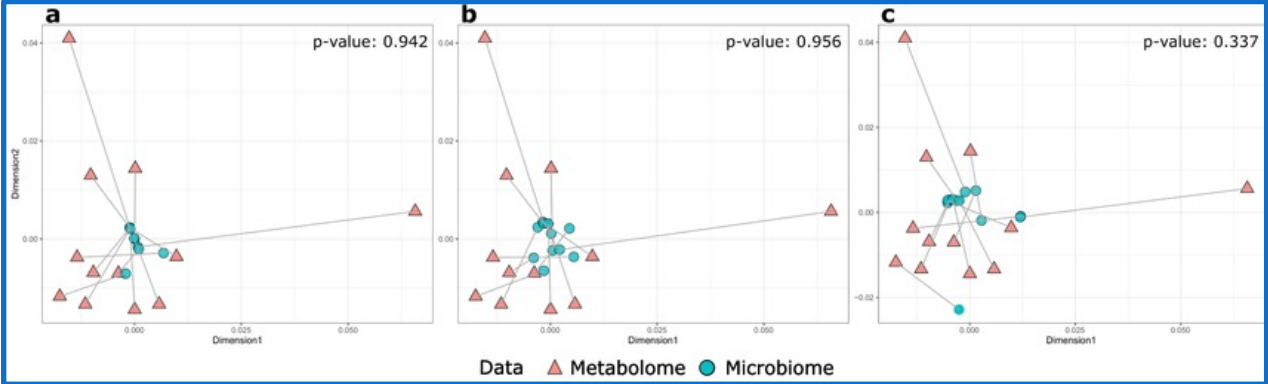
random forests,

## gradient methods

## Machine learning

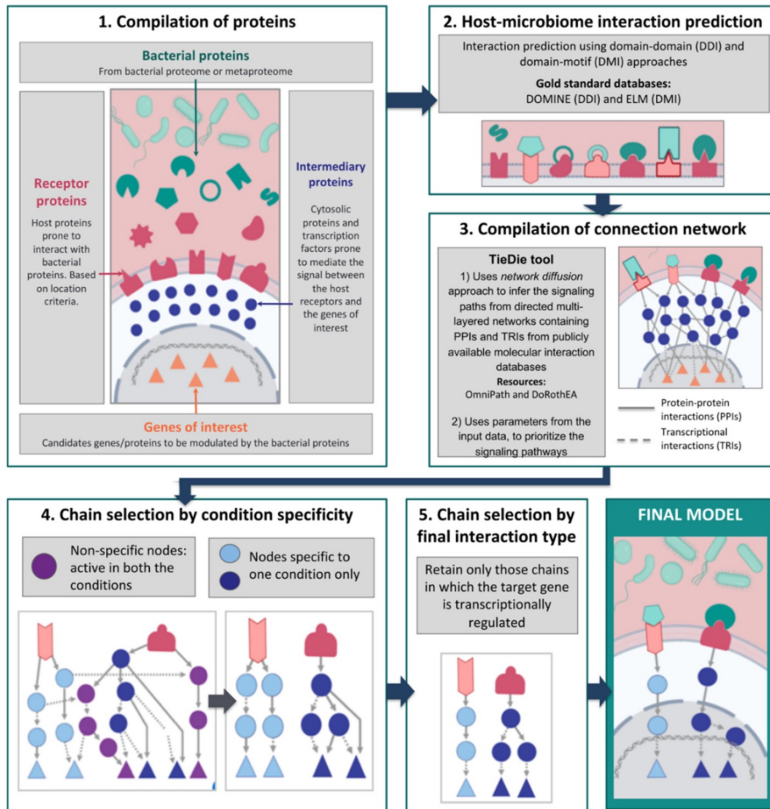
Need a lot of data

Deep learning



Kim, Younjung, et al. "In-Depth Characterisation of the Urine Metabolome in Cats with and without Urinary Tract Diseases." *Metabolomics*, vol. 18, no. 4, Mar. 2022, p. 19. Springer Link, <https://doi.org/10.1007/s11306-022-01877-9>

# Microbiolink: cross kingdom analysis



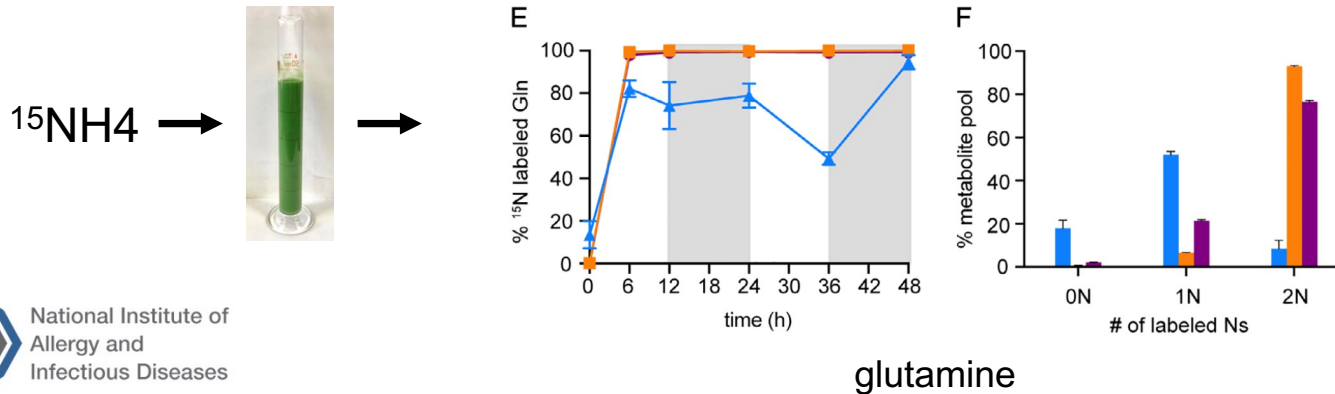
1. Infer reaction and gene(s) producing metabolite
2. Collect gene/protein annotations. Identify genes of interest.
3. Look up domain-domain interactions in a database.
4. Network analysis to infer signaling pathways using protein-protein interaction databases.
5. Remove nodes in common between conditions (this is specific to this particular study).
6. Keep the chains that include the target genes of interest.

# Stable labeled isotope probing

- Stable isotope-resolved metabolomics
  - Based on the assimilation of stable labeled isotopes ( $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled substrates) into newly synthesized metabolites/biomass
  - enables dynamic tracking of individual atoms through metabolic networks that exist in the microbiome
  - Can be more informative about metabolism than metabolomics “snapshot”

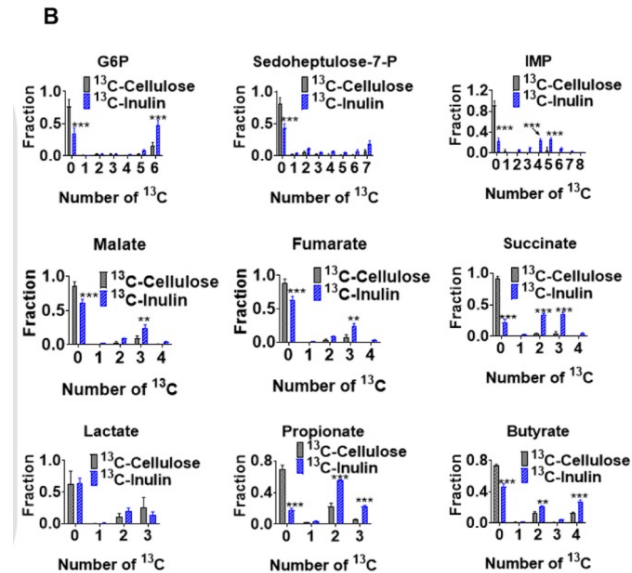
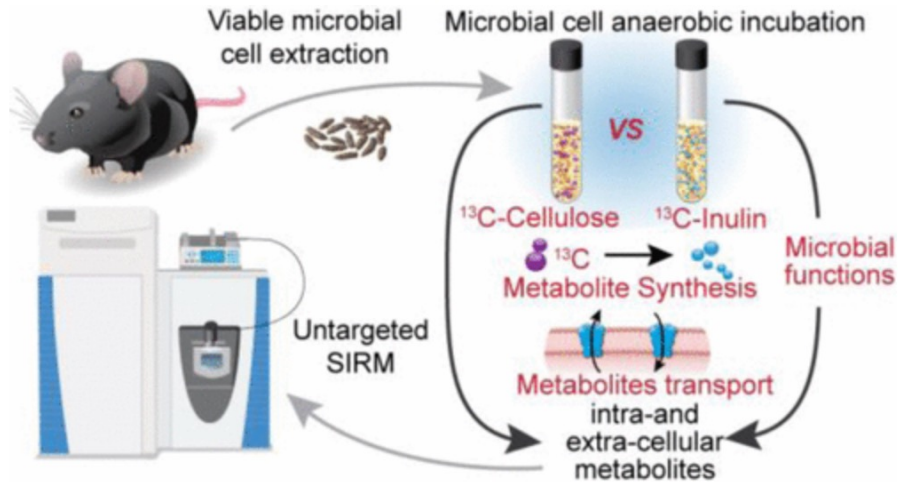
# Stable labeled isotope probing

- *in vitro*
  - Add labeled substrate to *in vitro* microbiome
    - Gut reactor, mesocosm experiment with water
- *in vivo*
  - Feed mouse or organism diet or substrate labeled with  $^{13}\text{C}$ ,  $^{15}\text{N}$
  - Collect stool or biopsy





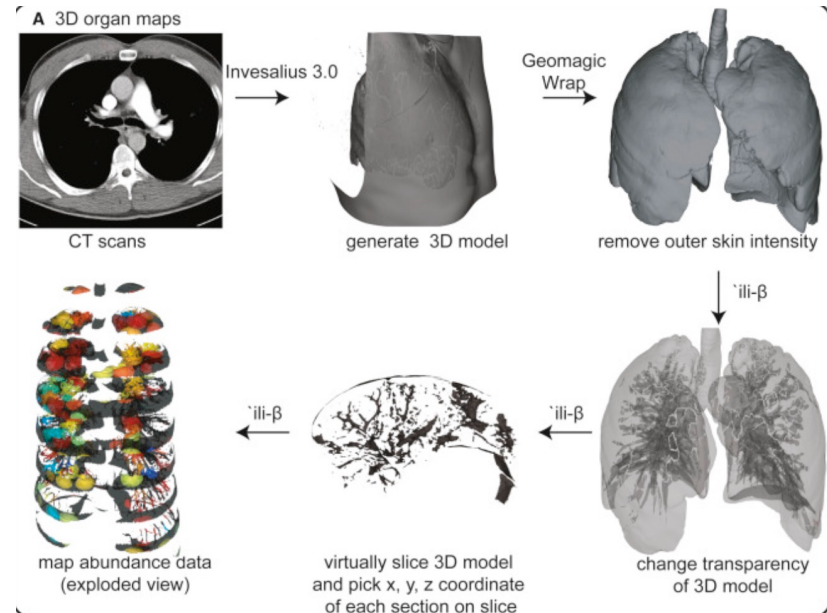
# Example: Untargeted stable isotope probing of the gut microbiota metabolome using $^{13}\text{C}$ -labeled dietary fibers.



Deng, P., Valentino, T., Flythe, M. D., Moseley, H. N., Leachman, J. R., Morris, A. J., & Hennig, B. (2021). Untargeted stable isotope probing of the gut microbiota metabolome using  $^{13}\text{C}$ -labeled dietary fibers. *Journal of proteome research*, 20(5), 2904-2913.

# 3D cartography: Spatial heterogeneity in metabolites and microbiome

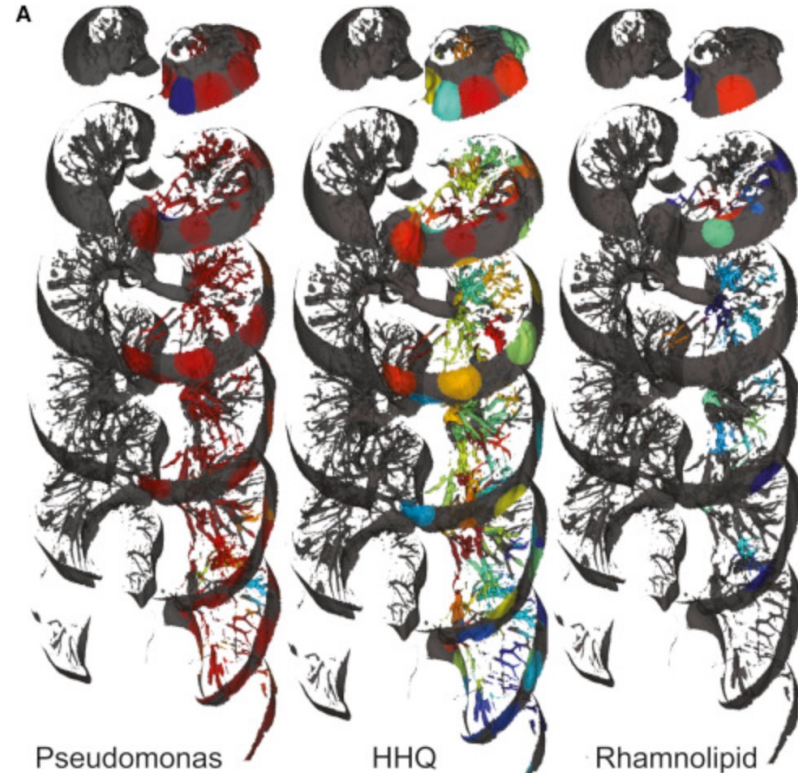
Use 3D imaging along with spatially distinct microbiome and metabolomic sample collection



# Example

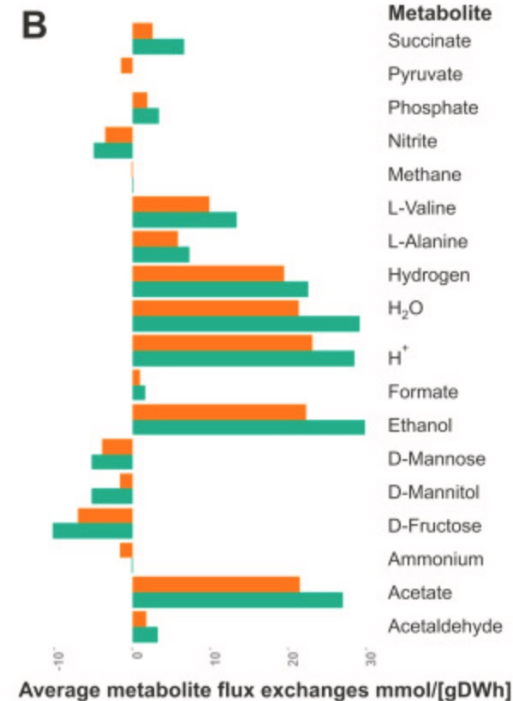
Identifying spatial relationships between *Pseudomonas* in cystic fibrosis patients and *Pseudomonas*-derived metabolites

Garg, N., Wang, M., Hyde, E., da Silva, R. R., Melnik, A. V., Protsyuk, I., ... & Dorrestein, P. C. (2017). Three-dimensional microbiome and metabolome cartography of a diseased human lung. *Cell host & microbe*, 22(5), 705-716.



# Flux balance analysis

- No metabolomics data? Consider FBA!
  - Reconstruct metabolism using genomic information from microbial community
  - Various metabolic models and tools available



Basile, A., Campanaro, S., Kovalovszki, A., Zampieri, G., Rossi, A., Angelidaki, I., ... & Treu, L. (2020). Revealing metabolic mechanisms of interaction in the anaerobic digestion microbiome by flux balance analysis. *Metabolic Engineering*, 62, 138-149.

# Thank you!!

Science Support: Metagenomics/Microbiome

Bioinformatics and Computational Biosciences Branch (BCBB)

National Institute of Allergies and Infectious Diseases (NIAID)

National Institute of Health (NIH)

Bethesda, MD, USA

**Find out more about what is offered at the BCBB:**

<https://www.niaid.nih.gov/research/metagenomics-bioinformatics-and-computational-biosciences-branch>

Contact: [bioinformatics@niaid.nih.gov](mailto:bioinformatics@niaid.nih.gov)