

# Analyzing patterns of co-occurrence between metabolomics and metagenomics features of human microbiome

Egor Shcherbin

Saint Petersburg Academic University

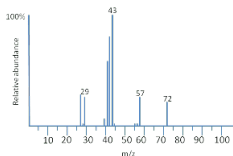
20 February 2018

# Human microbiome



# Metabolomics and metagenomics features

- ▶ Metabolomics features
  - ▶ MS1 or MS/MS features clustered using MS-Cluster



- ▶ Metagenomics features
  - ▶ Operational Taxonomic Units (OTUs)
  - ▶ Biosynthetic Gene Clusters (BGCs)



# Problems

1. Finding associations between features
  - ▶ Molecular networks (Watrous et. al., 2012)
2. Discovering and identifying microbial compounds
  - ▶ Dereplicator+ (Mohimani et. al., submitted)
  - ▶ VarQuest (Gurevich et. al, 2018)

## Patterns of co-occurrence

	Samples															P-value
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Compound X OTU A	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0.43
Compound X OTU B	1	1	0	0	0	0	1	0	0	1	1	0	0	1	1	0.04

- ▶ Two features are associated if they have similar patterns of occurrence in the samples
- ▶ Use one-tailed Fisher's exact test to evaluate similarity

# Pipeline

1. Identify the samples in which each MS/OTU/BGC feature is present
2. Search for pairs of features with statistically significant similar patterns of co-occurrence
3. Construct a network of significant pairs of features
4. Simplify the network by keeping top  $k$  edges for each node ( $k = 5$ )
5. Extract connected components



# Dataset overview

## 1. AGP

- ▶ 2125 samples
- ▶ 74913 MS/MS features
- ▶ 11265 OTU features
- ▶ 7712 significant matches of metabolomics features to OTU features at  $1e-10$  threshold

## 2. iHMP

- ▶ 388 samples
- ▶ 59916 MS1 features
- ▶ 29846 nodes, 37562 edges ( $p$ -value  $< 1e-10$ ), and 4296 connected components
- ▶ 512 out of 59916 MS1 features annotated by Human Metabolome Database (HMDB). Among them, 152 associated with known and novel metabolites, including carnitine, triglyceride, and cholate.



## PQS

Dereplicator+ search of the AGP dataset revealed a list of 68 identified compounds which had at least one significant match with an OTU feature. Particularly,

- ▶ OTU features from the *Pseudomonas* genus clustered with pseudomonas quinolone signal (PQS) BGC, and several identified compounds including phenazine carboxylic acid ( $m/z$  225.06), rhamnolipid ( $m/z$  673.39), and 4 compounds annotated as PQS with  $m/z$  258.14,  $m/z$  260.16,  $m/z$  270.18, and  $m/z$  288.19

## *F. prausnitzii*

- ▶ Significant correlations observed between Stercobilin,  $m/z$  595.35, Urobilin,  $m/z$  592.35, and microbial strain *Faecalibacterium prausnitzii*.
  - ▶ It is well known that Stercobilin and Urobilin are end-products of heme catabolism by colon micro-organisms, but the specific organism responsible for this biotransformation has not been reported previously.



## Flaxseed

Moreover, we searched iHMP metabolomics features for masses of two flaxseed orbitides, Cyclolinopeptide A and Cyclolinopeptide E.

- ▶ They were found in a cluster containing 6 other metabolomics features, with  $m/z$  1058.61,  $m/z$  1082.51,  $m/z$  511.26,  $m/z$  999.53,  $m/z$  1015.50, and  $m/z$  1086.52
- ▶ First two, with  $m/z$  1058.61 and  $m/z$  1082.51, are likely to be Cyclolinopeptide B and Cyclolinopeptide H, and other are potentially novel compounds associated with flaxseed
- ▶ Molecular networking failed to capture the association between cyclolinopeptides, due to their distinct sequences.

## ASMS annual conference 2018

**Shcherbin, E.**, Cao, L., Jarmusch, A. K., Dorrestein, P. C. & Mohimani, H.  
“Association between metabolomics and metagenomics features of human microbiome revealed by co-occurrence pattern analysis”, *abstract submitted*