pepFunk: peptide-centric functional enrichment for metaproteomic gut microbiome data

useR2020!

Caitlin Simopoulos, PhD

🐙 🐦 caitsimop

csimopou@uottawa.ca

Postdoctoral Associate
Department of Biochemistry, Microbiology and Immunology
Ottawa Institute of Systems Biology
Faculty of Medicine, University of Ottawa, Ontario, Canada

1. Brief intro to the gut microbiome
2. Challenges with metaproteomics
3. Following the pepFunk methodology with real data
4. pepFunk as a Shiny app integrated with MetaLab

The human gut microbiome

- The gut microbiome is the collection of bacteria, fungi, yeast, archaea and viruses that live in our digestive tract.
- These microorganisms perform essential functions, such as fermentation of fiber into beneficial short chain fatty acids (SCFA).
  - ▶ Our intestinal cells use SCFA for fuel.
- Humans have a close relationship with our microbiomes.

The gut microbiome is associated with both health and disease

The function (and dysfunction) of the gut microbiome can influence our health and well-being.

Some examples of diseases and conditions that have an association with the gut microbiome are:

- Immune-system-associated:
  - Inflammatory Bowel Disease (IBD)
  - Asthma
  - Multiple Sclerosis
- Metabolic disorders
  - Diabetes
  - Obesity
- Cardiovascular disease
- Mental health
  - Anxiety
  - Depression

The composition of our microbiomes can be influenced by our genetics, but are mostly influenced by our environment and lifestyles.

Geographic location

Genes

Exercise level
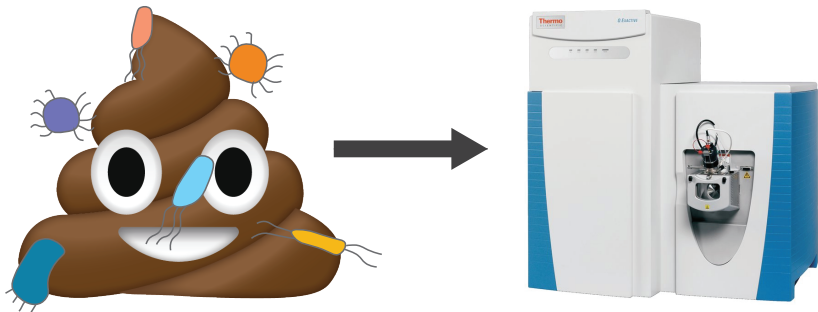
Mode of delivery

Stress

Diet

Age

Medication

[1]

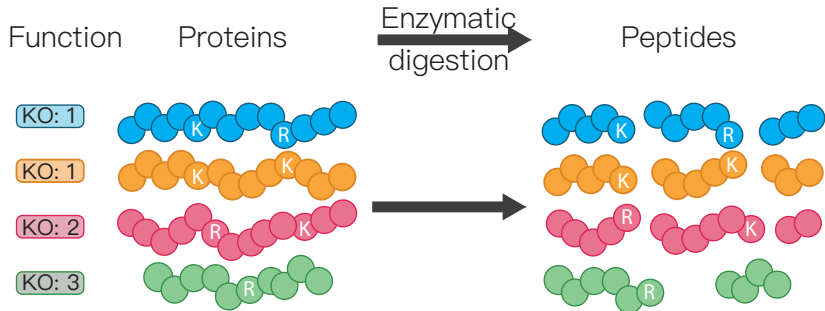# Gut microbiome research through fecal samples

- We study the gut microbiome through **fecal samples**.
- We extract microbial **proteins** out of fecal samples.
  - ▶ **Metaproteomics** = study of proteins from a community (multiple species).
- Proteins are the **functional** players of the microbiome.
- Metaproteomics allows us to look at **what gut microbes are doing**.

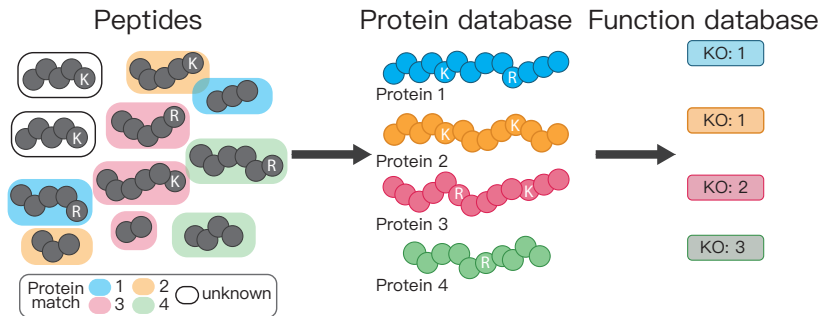# Metaproteomics and microbiome research

- Proteins are analyzed on a **mass spectrometer**.
- The **mass spectrometer** returns hundreds of thousands of spectra that we match to a database
- Using this database, we can computationally infer which protein was identified, from which microbial species, and how much of protein is in a sample.
  - ▶ We can analyze **microbiome composition** and **function**

Challenges with identifying function in metaproteomics
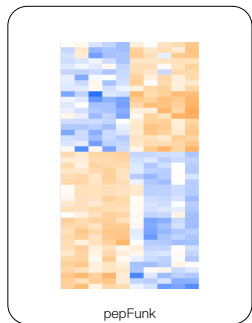

- Proteins are too large to be measured by a mass spectrometer and must be cut into smaller **peptides**.
- We use enzymes to cut proteins at **predictable sites** (K and R).
- The same peptide can be found in multiple proteins.
- **It is impossible to match some peptides back to the original protein.**

Why do we match peptides back to their parent protein?

A typical work flow matches each peptide back to a parent protein, and uses proteins for functional enrichment analysis. **What if we skipped this step and looked at functional enrichment of the identified peptides themselves?**

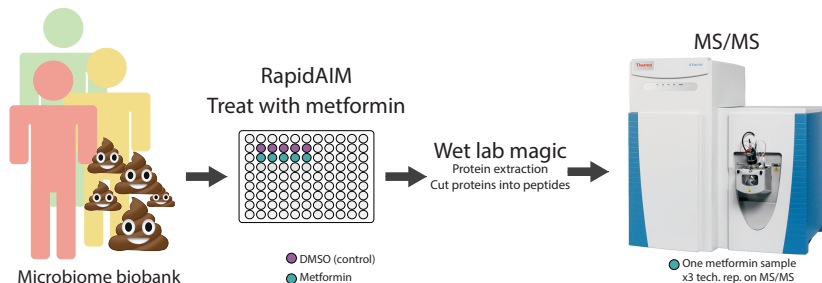# pepFunk: a peptide-centric functional enrichment methodology

pepFunk

- **pepFunk highlights molecular functions that are over- or under-represented** in microbiome samples using Gene Set Variation Analysis (GSVA) adapted for use with **peptide** intensity data
- pepFunk can handle peptides with multiple functional annotations by intensity weighting
- pepFunk is also available as a **Shiny app** for simplified analysis

1. **Subset peptides into "peptide sets"**
   ▶ KEGG pathways (gives information on functions)
2. **Estimate peptide intensity statistics** using a non-parametric kernel estimation of the cumulative density function
3. **Rank each peptide** by the expression statistic in every sample
4. **Calculate a Kolmogorov-Smirnov (KS)-like rank statistic** for each peptide set in each sample
5. **Calculate "GSVA" enrichment scores** for each peptide set[2]

I will illustrate these steps with an example from real data.

---

[2]Hanzelmann et al 2013 BMC Bioinformatics http://www.biomedcentral.com/1471-2105/14/7

RapidAIM
Treat with metformin

MS/MS

Wet lab magic
Protein extraction
Cut proteins into peptides

Microbiome biobank

● DMSO (control)
● Metformin

● One metformin sample
x3 tech. rep. on MS/MS

We treated a single microbiome with Metformin for 24 hours (using **RapidAIM**[3]). The microbial proteins were extracted and analyzed on a **tandem mass spectrometer (MS/MS)**.

[3]Li et al 2020 Microbiome https://doi.org/10.1186/s40168-020-00806-z

Peptide sets organized by function

1. Metabolism

1.0 Global and overview maps
- 01100 Metabolic pathways *Major update!*
- 01110 Biosynthesis of secondary metabolites
- 01120 Microbial metabolism in diverse environments
- 01130 Biosynthesis of antibiotics
- 01200 Carbon metabolism
- 01210 2-Oxocarboxylic acid metabolism
- 01212 Fatty acid metabolism
- 01230 Biosynthesis of amino acids
- 01220 Degradation of aromatic compounds

1.1 Carbohydrate metabolism
- 00010 Glycolysis / Gluconeogenesis
- 00020 Citrate cycle (TCA cycle)
- 00030 Pentose phosphate pathway
- 00040 Pentose and glucuronate interconversions
- 00051 Fructose and mannose metabolism
- 00052 Galactose metabolism
- 00053 Ascorbate and aldarate metabolism
- 00500 Starch and sucrose metabolism
- 00520 Amino sugar and nucleotide sugar metabolism
- 00620 Pyruvate metabolism
- 00630 Glyoxylate and dicarboxylate metabolism
- 00640 Propanoate metabolism
- 00650 Butanoate metabolism
- 00660 C5-Branched dibasic acid metabolism
- 00562 Inositol phosphate metabolism

1.2 Energy metabolism
- 00190 Oxidative phosphorylation
- 00195 Photosynthesis
- 00196 Photosynthesis - antenna proteins
- 00710 Carbon fixation in photosynthetic organisms
- 00720 Carbon fixation pathways in prokaryotes
- 00680 Methane metabolism
- 00910 Nitrogen metabolism
- 00920 Sulfur metabolism

1.3 Lipid metabolism
- 00061 Fatty acid biosynthesis
- 00062 Fatty acid elongation
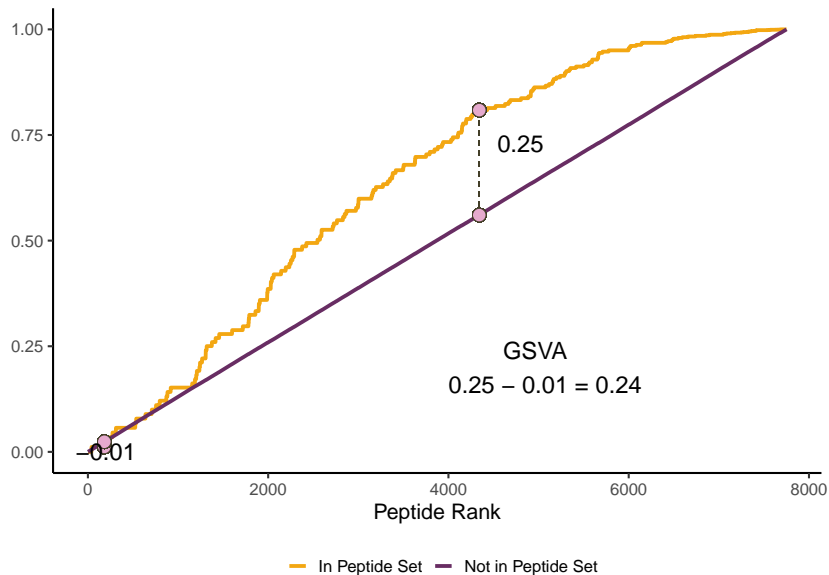- 00071 Fatty acid degradation

- We created peptide sets according to KEGG terms
  - ▶ Describes function
- Peptides are assigned to functions according to our curated database
- Peptides with multiple functions are weighted according to confidence in functional assignment
- Calculate potential enrichment of functions in each sample

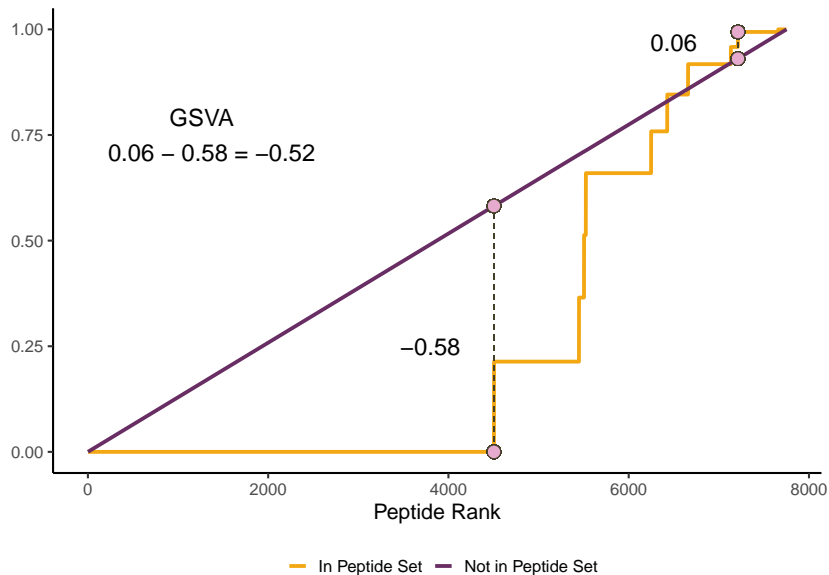# We calculate a KS-like statistic for **each** peptide set in **each** sample

We use a **random walk** along the peptide rankings (from highest to lowest intensity) to determine if the peptide set is **more highly or lowly ranked than expected**. If a peptide is in a peptide set, **we rise according to its previously calculated peptide expression statistic**. We do the same for all peptides not in our set of interest.
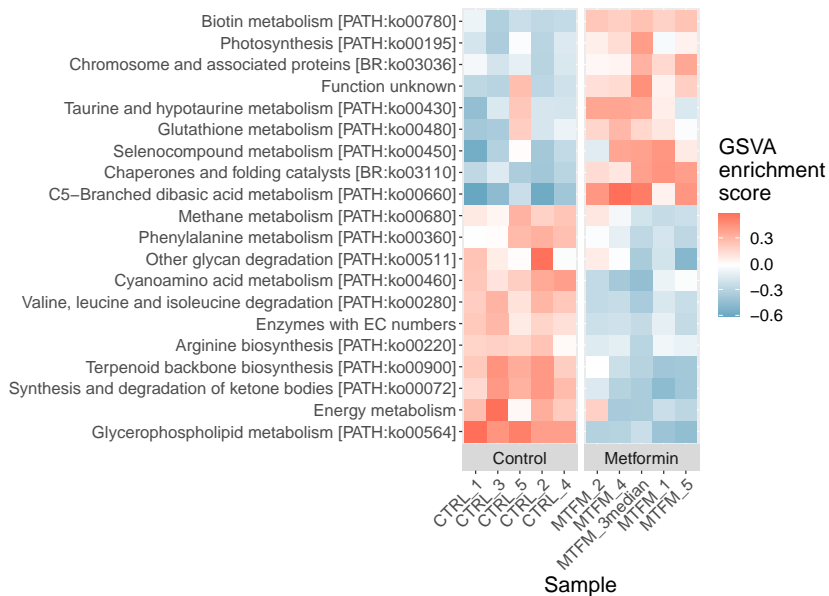
Positive GSVA score, higher intensity than expected

Negative GSVA score, lower intensity than expected

pepFunk as part of MetaLab suite

- MetaLab/iMetLab[4] is an entire platform dedicated to metaproteomic data analysis
- pepFunk is included in the iMetaLab platform along with other Shiny apps for metaproteomic data analysis

[4]Cheng et al 2020 J. Am. Soc. Mass Spectrom doi:10.1021/jasms.0c00083

The pepFunk Shiny app

Applications of pepFunk

- We can use pepFunk to look at functional changes of a microbiome:
  - ▶ after drug treatment, better understanding of side effects
  - ▶ because of diet
  - ▶ because of disease
  - ▶ other lifestyle changes
- pepFunk is **not limited to gut microbiomes**
  - ▶ **soil microbes** (agriculture)
  - ▶ **water microbes** (water treatment)
  - ▶ . . . your sourdough starter? (the ultimate sourdough bread)

Acknowledgements

https://github.com/northomics/pepFunk

- **Dr Daniel Figeys**
- **Dr Mathieu Lavallée-Adam**
- Dr Zhibin Ning
- Dr Xu Zhang
- Dr Leyuan Li
- Krystal Walker

- Patrick Smyth

Link to Shiny app
Link to publication

TECHNOMISE

Natural Sciences and Engineering
Research Council of Canada

Conseil de recherches en sciences
naturelles et en génie du Canada

uOttawa

Canada