Applying ’Omics to Your Research
- perspective from microbiome research

Pande Putu Erawijantari

Turku Collegium for Science, Medicine, and Technology – Postdoctoral researcher

Department of Computing, Faculty of Technology,

University of Turku,

Finland

pande.erawijantari@utu.fi

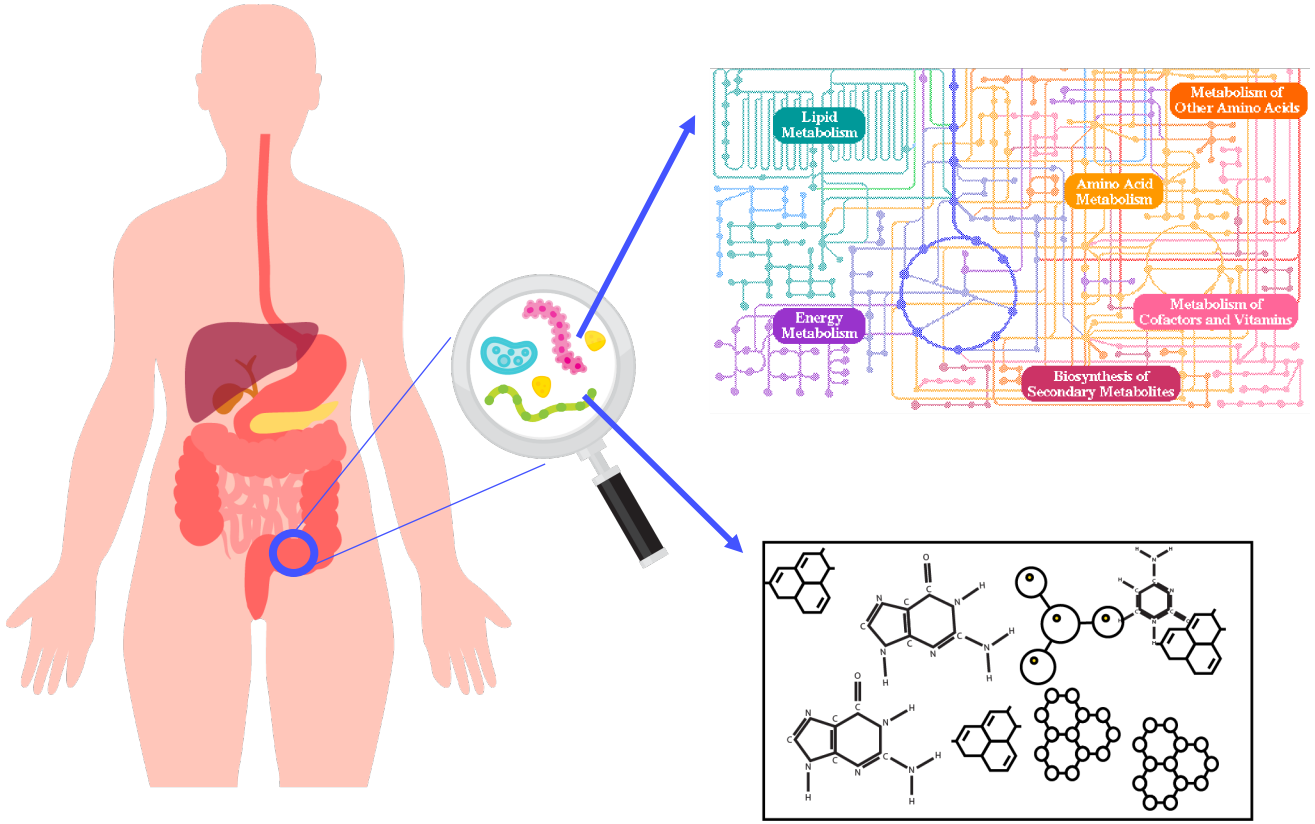
Learning objective

- Understanding the importance of characterizing each omics
  - Focus on metagenomics and/or metabolomics

Cover:
- Pipeline overview
- Example study on each omics and how it can address research questions

Not cover in the lecture: sampling and details step by step of omics measurement/instrument

Overview of human gut microbiome



Factors

- Diet
- Lifestyle
- Geography
- Age
- Medication
- Diseases
- Genetics

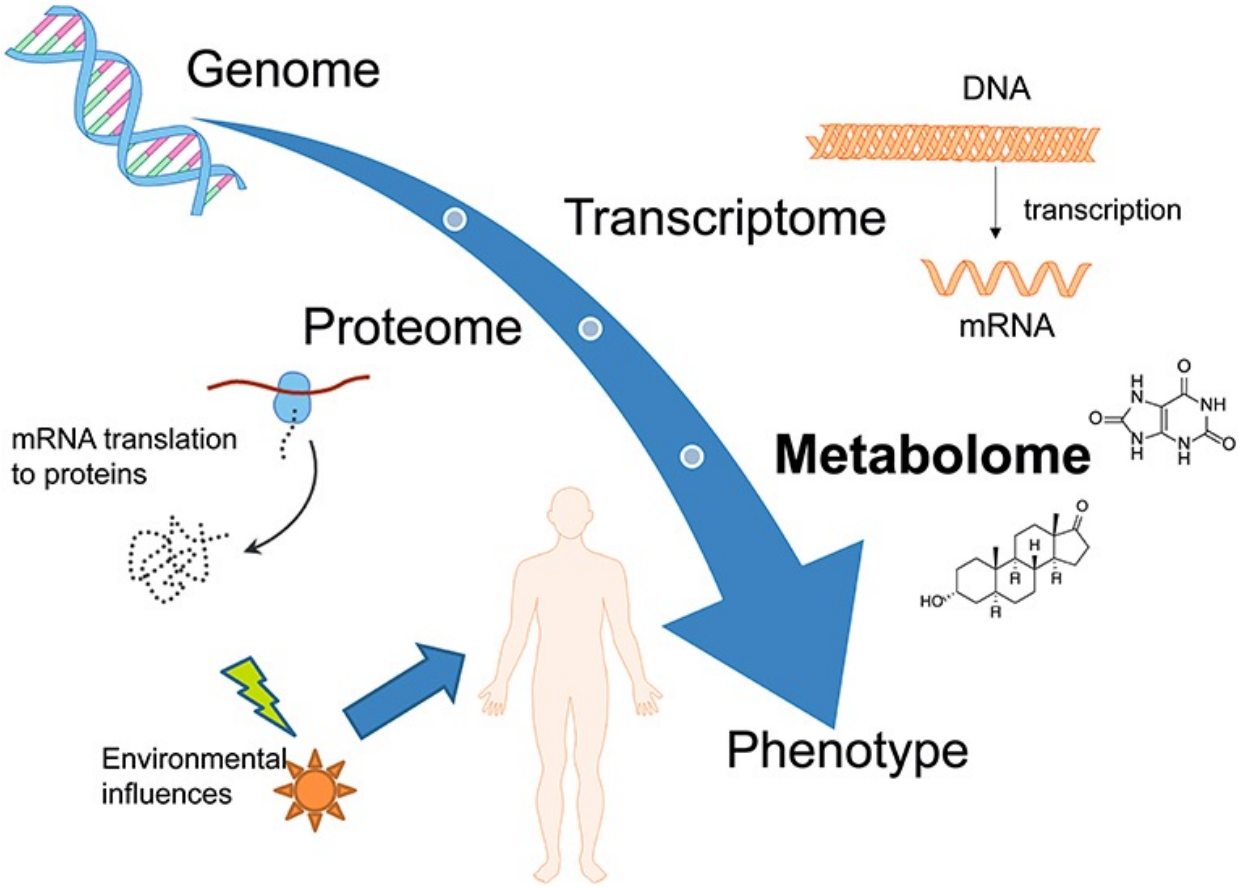
Roles

- Protecting host against pathogenic microbes
- Modulating immunity
- Regulating metabolic processes

Li et al., Nature Biotechnology, 2014, Human body picture designed by freepik.com

UNIVERSITY OF TURKU

From functional potential to activity



- **Genome** is, with few exceptions, present at the same concentration in each cell and is thereby static in nature. Genomic information allow us to perform a functional potential prediction.
- **Transcriptome** provide the information on cell and tissue specific gene expression for a better understanding of the dynamics of cellular and tissue metabolism[1].
- **Proteomic** provide more complex information regarding the cell/tissue differentiation which included the information on protein folding[1].
- **Metabolites** are the result of both biological and environmental factors and, as such provide great potential to bridge knowledge of genotype and phenotype[2].

Figure source::https://omicstutorials.com/metabolomics-in-the-past-present-and-future/

[1]Manzoni et al., Briefings in Bioinformatics, 2016
[2]Schrimpe-Rutledge, A.C *et al.*, J. Am Soc. Mass Spectrom. 2016

Difference between amplicon sequencing and metagenomics

- Amplicon sequencing = sequencing of PCR products amplified with target specific DNA primers, often refers to the amplification of the bacterial phylogenetic marker gene 16S rRNA gene
- Metagenomics or shotgun metagenomics = sequencing of random DNA from a sample without any preselection

Shotgun Metagenomics:
Taxonomic annotation in microbiome study

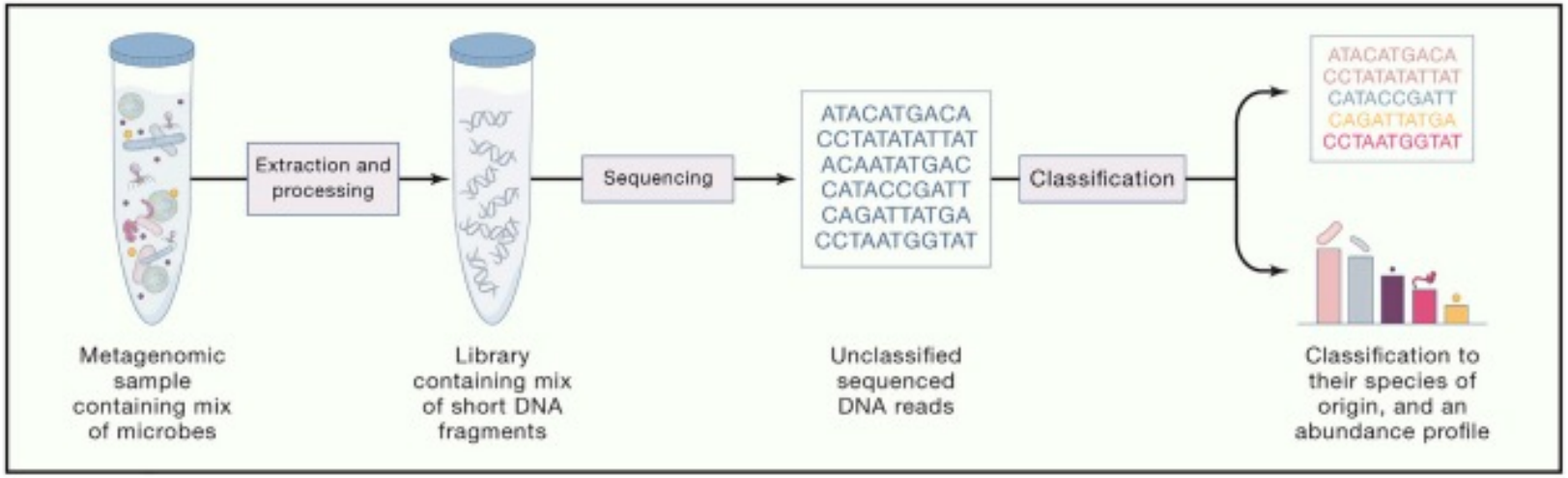


Figure 1. Processing Steps to Go from a Complex Metagenomic Sample to an Abundance Profile of Sample Content

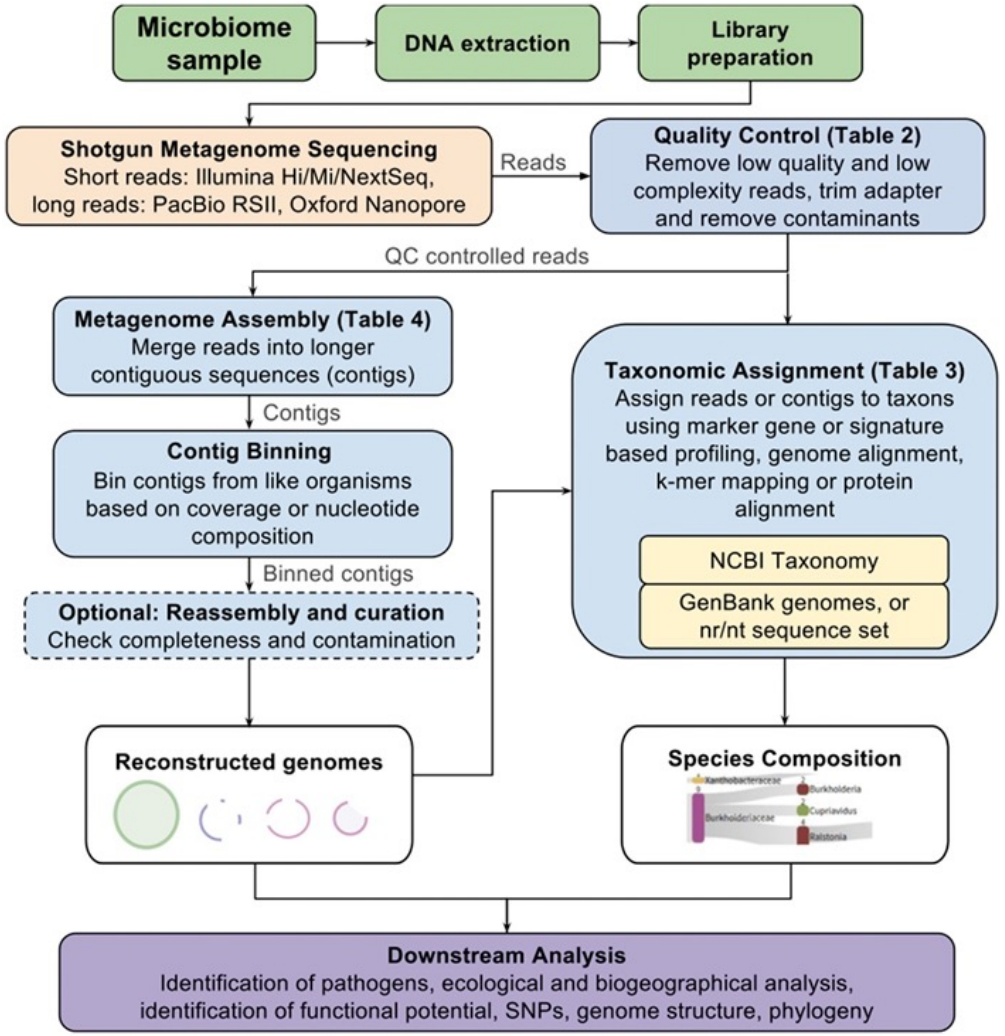
Ye et al., Cell. 2019

- Taxonomic identification is usually the first and essential step in microbiome study to identify the member of microorganism in certain sample. The information are subsequently used to infer their roles in study of interest for example in comparing the human health and diseases status.

Different Approach in taxonomic annotation

- **Metagenomic Assembly**

Approaches that involves assembling the reads, binning the contig based on similarity and identify the bins into taxon.



Common analysis flow (Breitwieser et al, 2019)

- **Reference based**

Matching the sequences typically reads or assembled contigs against a database of microbial genome to identify the taxon of each sequence.

Bioinformatic tools

- Read based: MetaPhlAn 4.0, HUMAnN 3.0, StrainPhlAn, Kraken, Centrifuge
- MetaPhlAn: 17,000 reference genomes (~13,500 bacterial and archaeal, ~3,500 viral, and ~110 eukaryotic)
- Kraken & Centrifuge: kmer based taxonomic annotation against taxonomic databases
- HUMAnN: Funtional profiling
- StrainPhlan: Strain-level identification of taxa
- Mapping: Bowtie, BWA

Assembly based

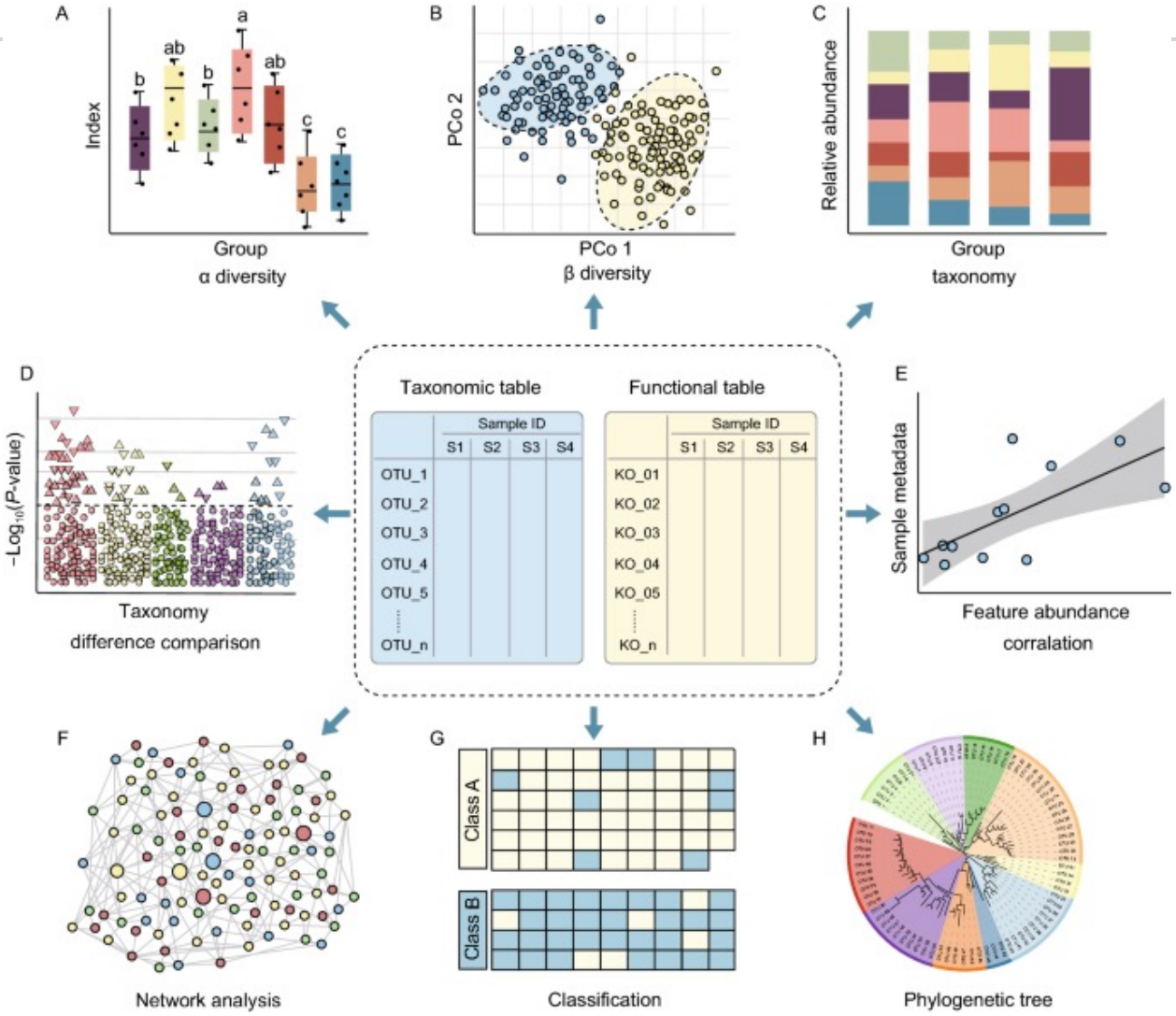
- Assemblers: Megahit, MetaSpades, IDBA-UD, SOAPdenovo, MetaVelvet
- Computational binning: MetaBat, Groopm, Autometa, MetaWrap
- DAS Tool is an automated method that integrates the results of a flexible number of binning algorithms to calculate an optimized, non - redundant set of bins from a single assembly
- CheckM: Assessing the quality of metagenome bins
- Manual binning: Anvi'o
- Gene prediction: Prodigal
- Gene annotation: HMMER, Blast, Prokka,..

Which approach will work best for my datasets?

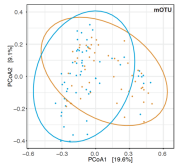
Unfortunately, there are no universal answers to that, depends on

- Sequencing depth
- Study design and research questions

Typical downstream analysis from different omics

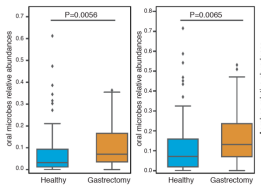


Liu et al., 2020



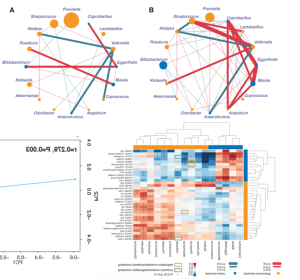
- **Multidimentional reduction**

- Beta diversity such as using Principal Coordinate Analysis (PCoA)
Followed by PERMANOVA



Measuring population diversity

- Differential abundance analysis



- Network analysis

Microbes-microbes correlations

- Omic integration

Data analysis workflow- Result intepretation

Figure:Statistical Analysis for two group comparison

Source code: https://github.com/yamada-lab/GastrectomyGC_Microbiome
Paper: Erawijantari et al., Gut, 2020

Example also available in OMA exercises

Dealing with high-dimensional data in microbiome study: multidimensional reduction

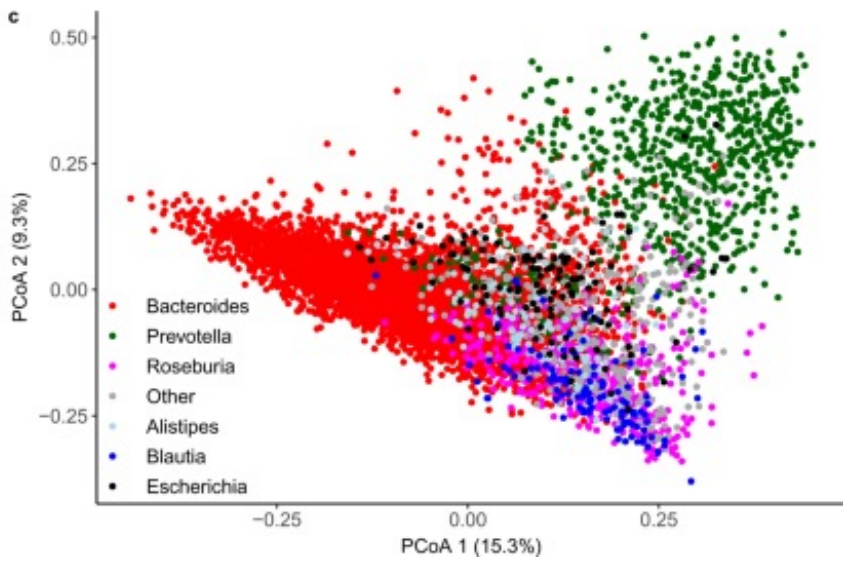
Taxon table

	Control 1	Control 2	Case 3	Case 4	.....
Species 1	0.05127856	0.579605386	0.360309588	0.08239625	0.360309588
Species 2	0.065766244	0.579605386	0.354275032	0.102283303	0.354275032
Species 3	0.065766244	0.579605386	0.08239625	0.579605386	0.08239625
.....	0.102283303	0.579605386	0.130373517	0.664904935	0.130373517
.....	.....	.....	.....	.....	.....

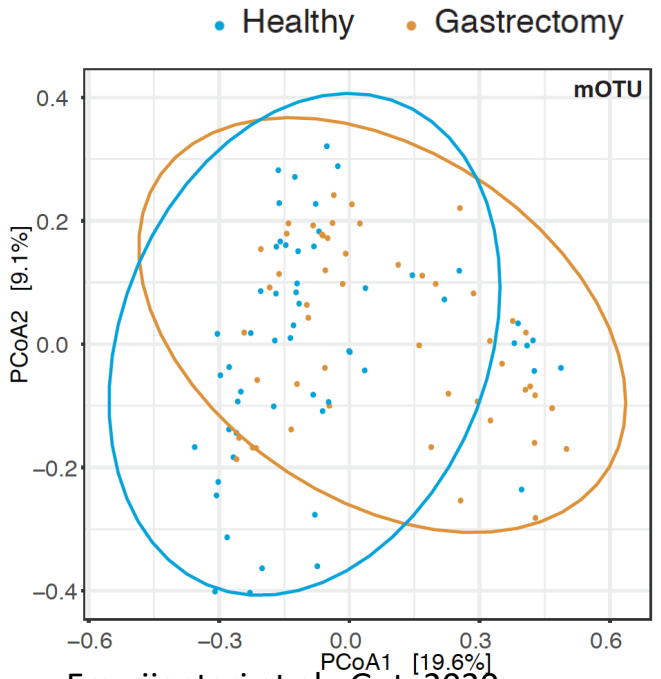
+

Subject metadata

	Age	Gender	BMI	Smoking status	.....
Control 1	60	0.579605386	0.360309588	0.08239625	0.360309588
Control 2	50	0.579605386	0.354275032	0.102283303	0.354275032
Case 3	65	0.579605386	0.08239625	0.579605386	0.08239625
Case 4	60	0.579605386	0.102283303	0.579605386	0.102283303
.....	.....	.....	.....	.....	.....



Salosensaari et al., Nat Comm, 2021



Erawijantari et al., Gut, 2020

[OMA example](#)

Importance:

- Exploratory phase
- Visualize data pattern in 2 dimension
- Visualize the beta diversity (e.g PCoA)

UNIVERSITY OF TURKU

Measuring population diversity

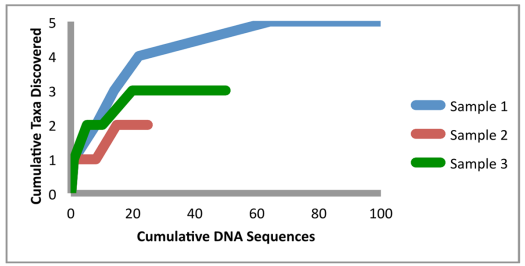
A) Sequence Abundance

OTU	Sample 1	Sample 2	Sample 3
A	60	0	35
B	24	5	5
C	10	0	0
D	5	0	0
E	1	0	0
F	0	20	10
Total	100	25	50

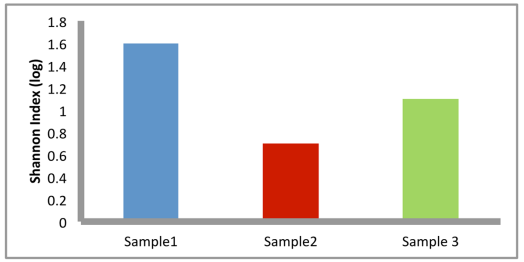
B) Sequence Relative Abundance

OTU	Sample 1	Sample 2	Sample 3
A	0.60	0	0.70
B	0.24	0.20	0.10
C	0.10	0	0
D	0.05	0	0
E	0.01	0	0
F	0	0.80	0.20
Total	1.0	1.0	1.0

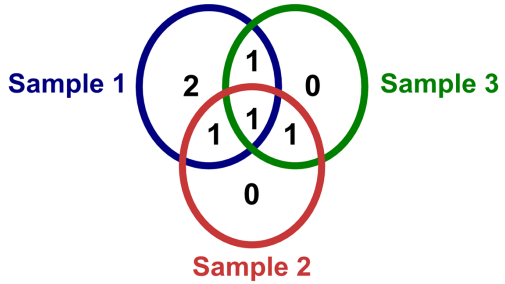
C) Collector's Curve of Sample Richness



D) Within-Sample Alpha Diversity



E) Between-Sample Beta Diversity



Ecological representations of microbial communities

- collector's curves for species richness estimated using e.g using Chao1, ACE index
- Alpha diversity (both richness and eveness) estimated using e.g Shannon or Simpson index
- Beta diversity: between sample differences

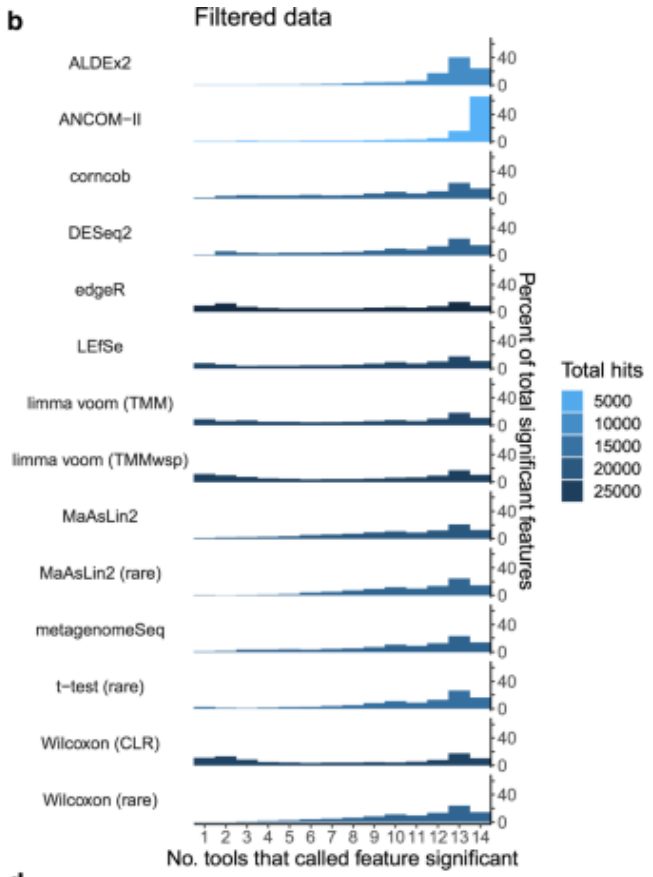
Morgan et al., PLOS Comp Biol, 2012

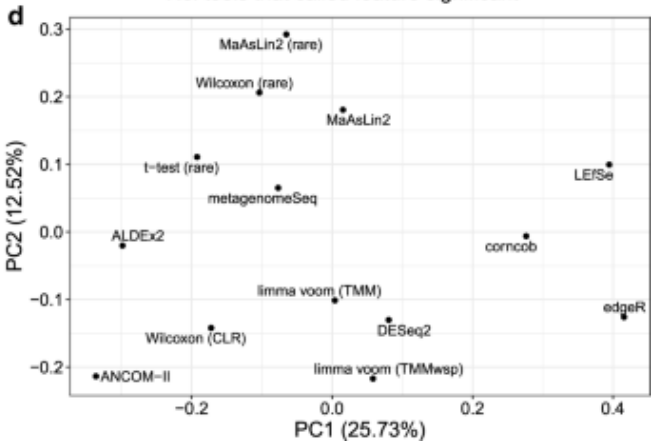
Why this matter?:

- Alpha diversity and species richness has been reported to be associated with human health status
- Beta diversity can be measured to see how difference the group tested within population

OMA example

UNIVERSITY OF TURKU





Nearing et al., Nat Comm, 2022

Differential abundance analysis

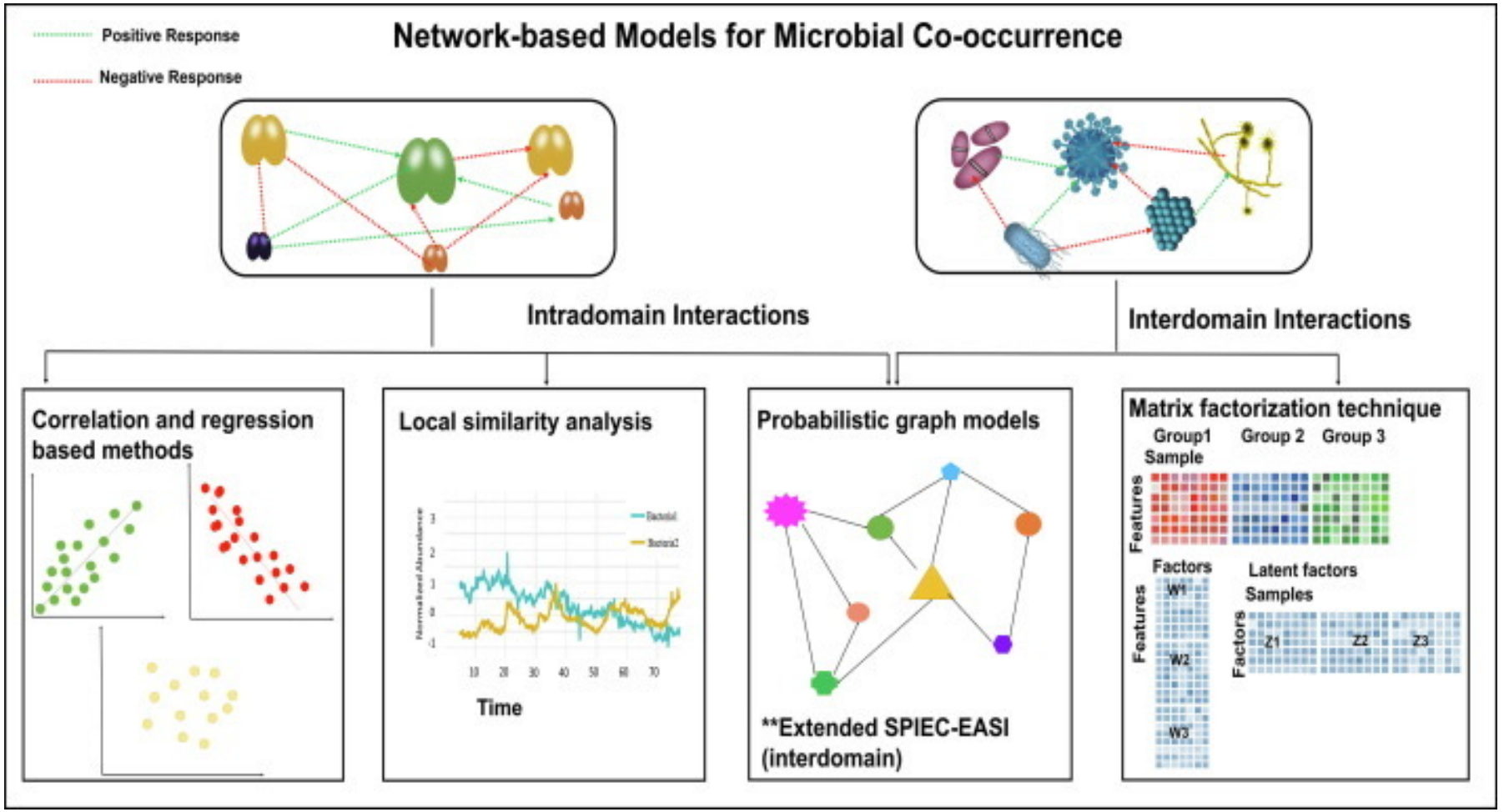
- This is the common first step in control-case study design to compare the two or more groups using statistical methods.
- Classic statistics: Mann-Whitney U test, t-test, wilcoxon test follow by false positive correction
- Other tools: see Nearing et al., Nat Comm, 2022
  - Different tools produce substantially different results: **always remember to check the asumption**
  - Other strategy: employed different methods to see the robustness of the results

OMA example

UNIVERSITY OF TURKU

Community level analysis through network

Aim: inferring inter and intra kingdom interaction in microbial community
For example to estimate the keystone species, community stability/disturbance
Tools example: SparCC, CoNet, SPIEC-EASI



Matchado et al., 2021

UNIVERSITY OF TURKU