Conducting Microbiome study, a How to guide

Sam Zhu
Supervisor: Professor Margaret IP
Joint Graduate Seminar
Department of Microbiology
15 December 2015

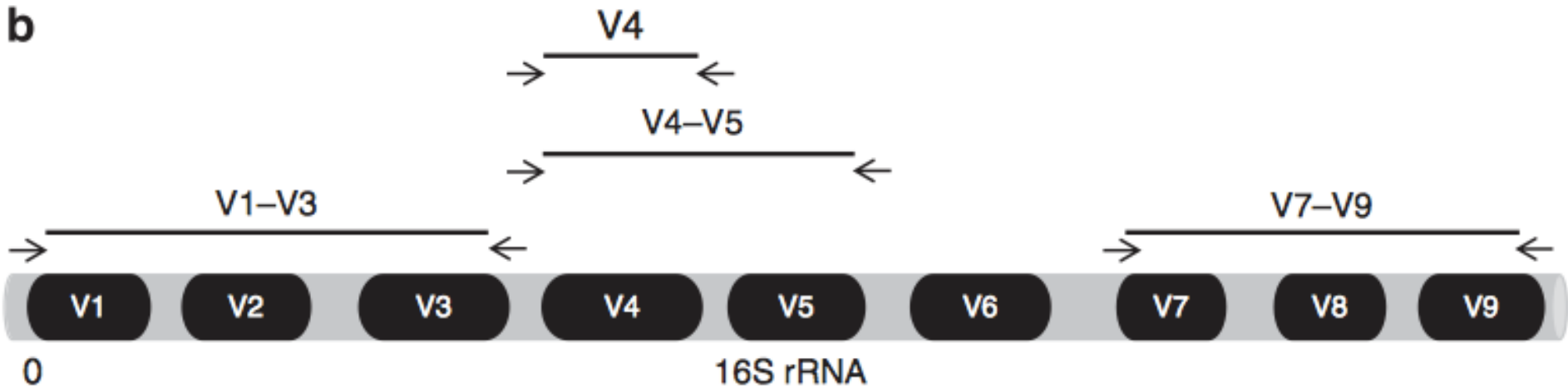
Why study Microbiome?

- ✓ Essential component, e.g. Gut microbes
- ✓ Disbiosis (microbial imbalance) & disease
  - ➢ *Obesity, diabetes, inflammatory bowl disease*
- ✓ Limitation of traditional culture methods.

16S rRNA gene

- Metagenomic Study methods based on:
  - ***16S rRNA gene*** *or Whole Genome Sequencing*
- 16S rRNA gene is widely accepted, because:
  - ✓ *9 hypervariable regions (Identification, phylogenetic)*
    + highly conserved regions (universal primers binding);
  - ✓ *Well-developed processing procedure;*
  - ✓ *Database is mature: e.g. Ribosomal Database Project (RDP)…*

16S rRNA



Black: hypervariable region

Grey: conserved region

Tyler A D, The American journal of gastroenterology, 2014, 109(7): 983-993.

16S rRNA gene

- Metagenomic Study methods based on:
  - ***16S rRNA gene*** *or Whole Genome Sequencing*
- 16S rRNA gene is widely accepted, because:
  - ✓ *9 hypervariable regions (Identification, phylogenetic)*
    + highly conserved regions (universal primers binding);
  - ✓ *Well-developed processing procedure;*
  - ✓ *Database is mature: e.g. Ribosomal Database Project (RDP)…*

Ongoing Microbiome Project

NIH HUMAN MICROBIOME PROJECT

Workflow of Microbiome Study

1. Study Design

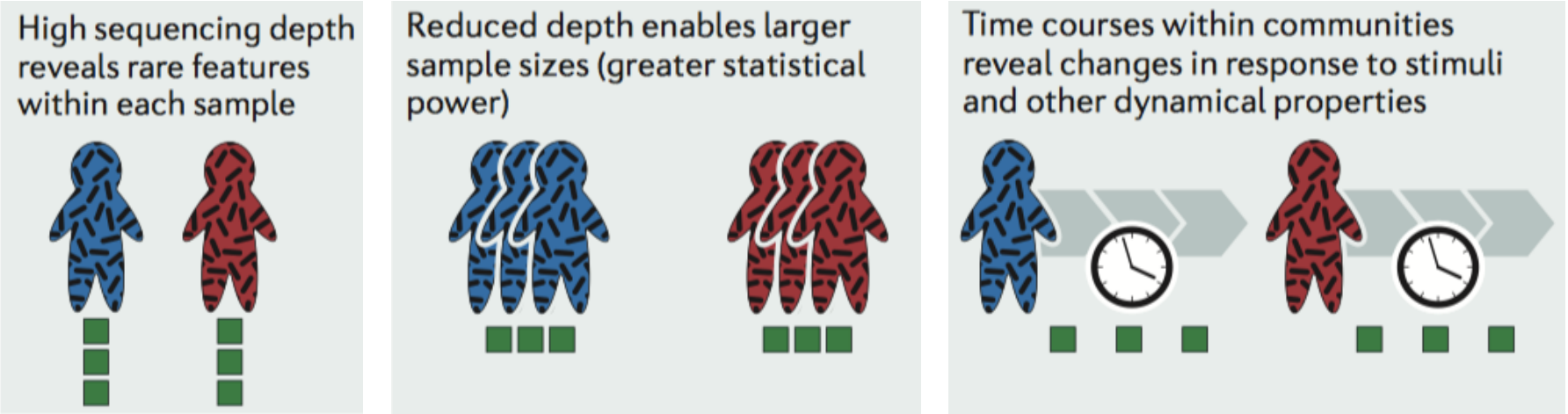
2. Sample Preparation

3. Library Preparation & Sequencing

4. Sequence Analysis

5. Statistical Analysis

Study Design



- Exclusion Criteria?
- Control group?
- Sampling methods?

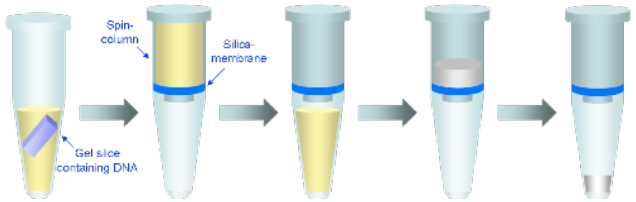
Franzosa E A, Nature Reviews Microbiology, 2015.

2. Sample preparation

1) Storing: immediate freezing, storage at -80℃.

   ➢ *Minimize freeze-thaw cycles.*

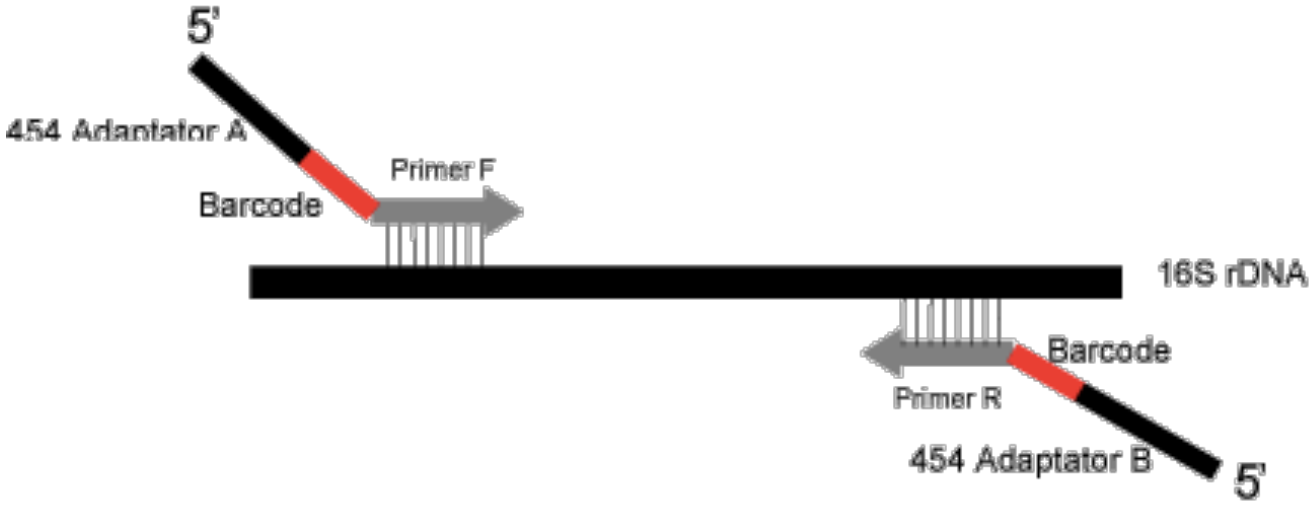
2) DNA extraction: commercially available kits specific to the sample type. (Caution: Run a blank extraction control!)



http://biotechcrunch.blogspot.hk/2011/11/9-factors-affecting-dna-extraction-from.html

3. Library preparation (PCR)

16S Primers & **Barcode & adaptor**



Fabrice A, J Comput Sci Syst Biol, 2009, 2(1): 074-92.

3. Sequencing

- **Platform:** Illumina MiSeq & Ion torrent

Platform	Cost per run	Read Length	Run time	Throughput
Ion torrent (318 chip)	$625	100 bases	3h	1000Mb
MiSeq	$750	150*2 bases	27h	1500Mb

Ion torrent

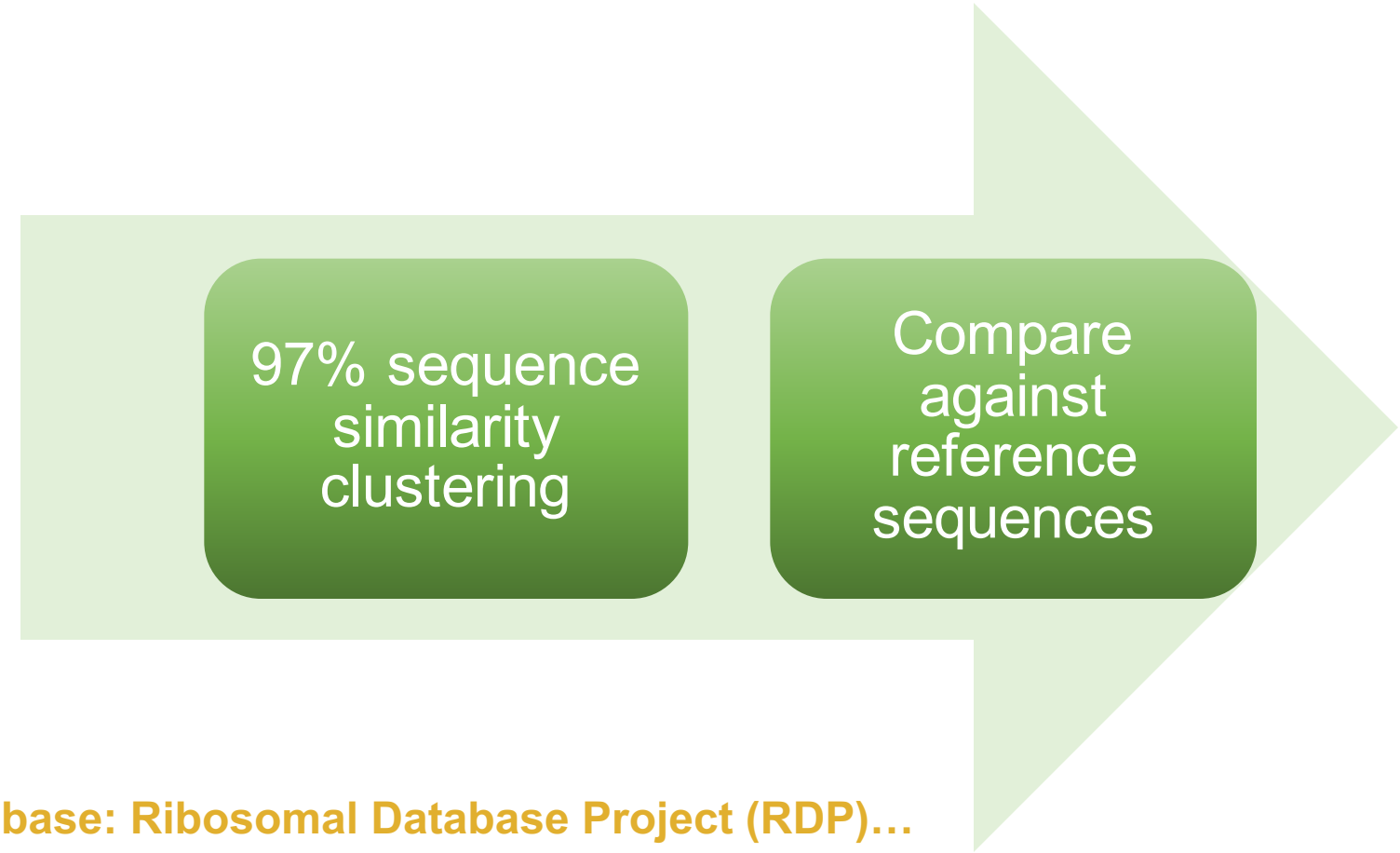
MiSeq

Loman N. Nature biotechnology, 2012, 30(5): 434-439.

4. Sequence Analysis

- **Software packages:** QIIME, mothur
- **Steps:**
    1. Sequence Quality Control
        - ✓ *Trimming adaptors*
        - ✓ *Filtering low quality sequence*
        - ✓ *Filtering Chimeras (5%~45%)*
    2. Taxonomic identification
        - ➢ Operational Taxonomic Units (OTUs) Analysis

Operational Taxonomic Units (OTUs) Analysis



*** Database: Ribosomal Database Project (RDP)…**

5. Statistical Analysis: Measure of Diversity

Relative abundance:

- Quantitative measure
- Frequency of a specific taxon

Alpha diversity:

- Measure of diversity within a sample
- Calculated by algorithms

Beta diversity:

- Comparison of diversity between samples
- Displayed as Principle Coordinate Analysis (**PCoA**) plot

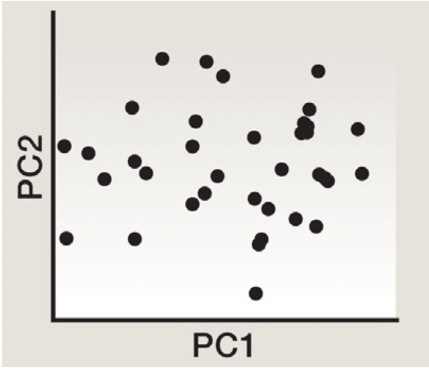
Principle Coordinate Analysis (PCoA) plot

- Microbiome data:

  Multi-dimension, each species = one dimension

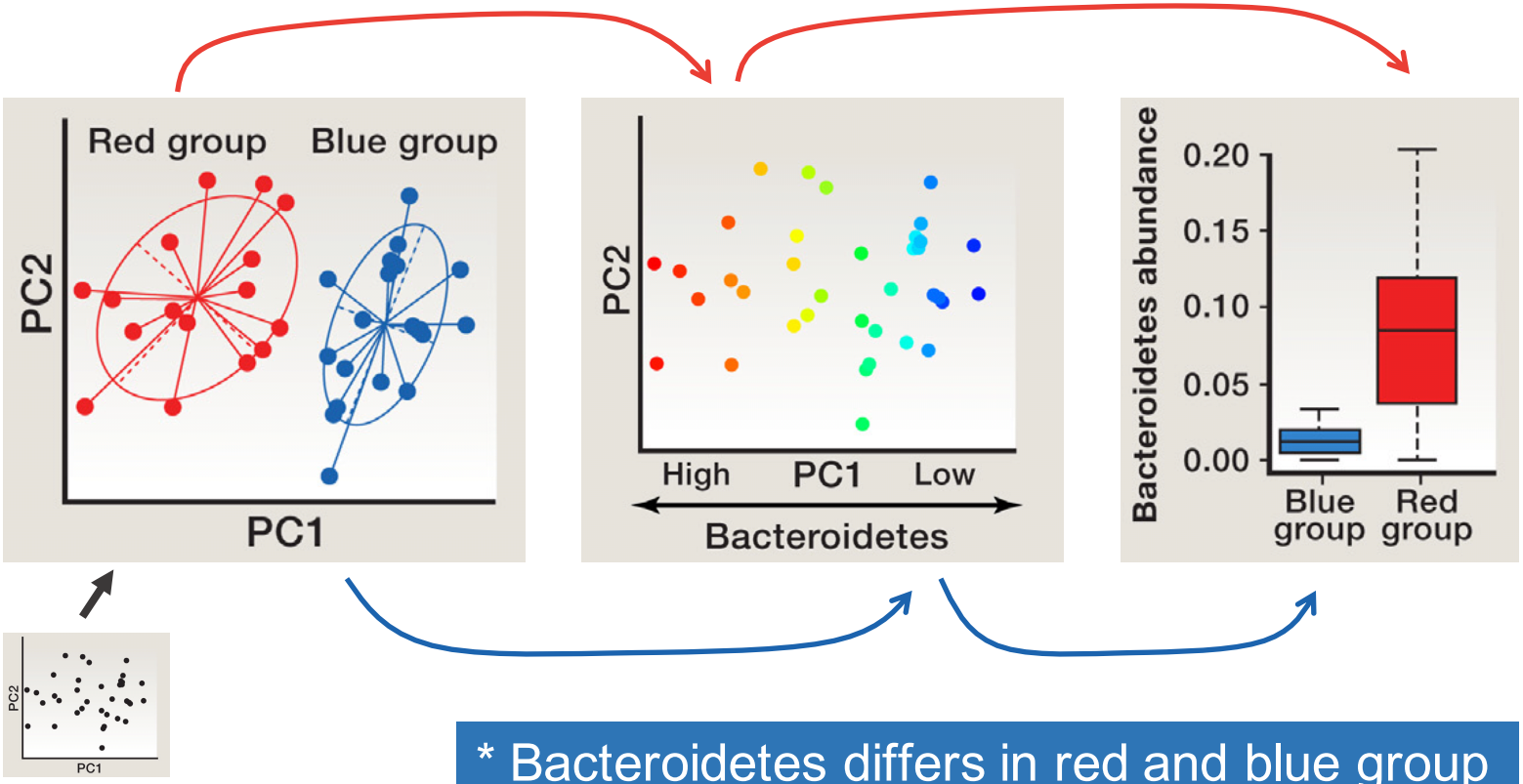
- Mathematically reduce them to two/three dimensions:

  Principle Coordinates (PC)



Data visualization: PCoA plot
*each dot= a sample

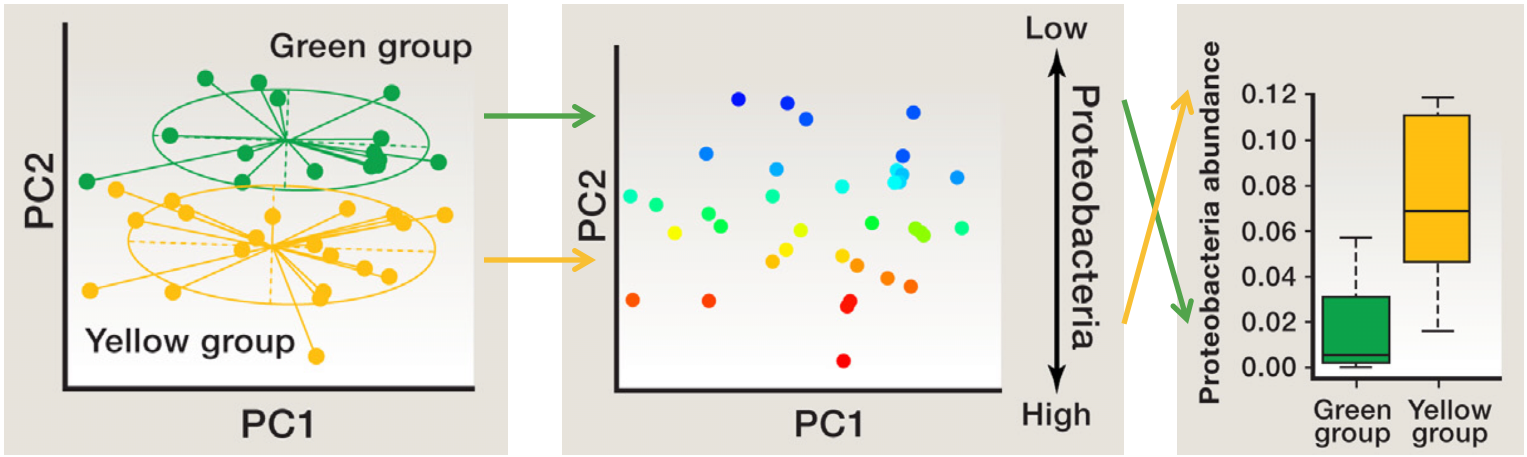
Principle Coordinate Analysis (PCoA) plot

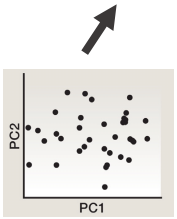


* Bacteroidetes differs in red and blue group

Goodrich J K, Cell, 2014, 158(2): 250-262.

Principle Coordinate Analysis (PCoA) plot





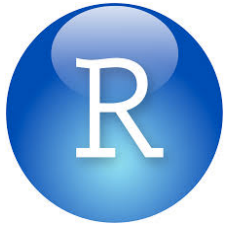
* Proteobacteria differs in green and yellow group

Goodrich J K, Cell, 2014, 158(2): 250-262.

5. Statistical analysis: Regression Analysis

- **Response variables**: alpha diversity, PCs, the abundances of OTUs
- **Statistical models**: Multiple Logistic Regression…
- **Control cofounding factors:**

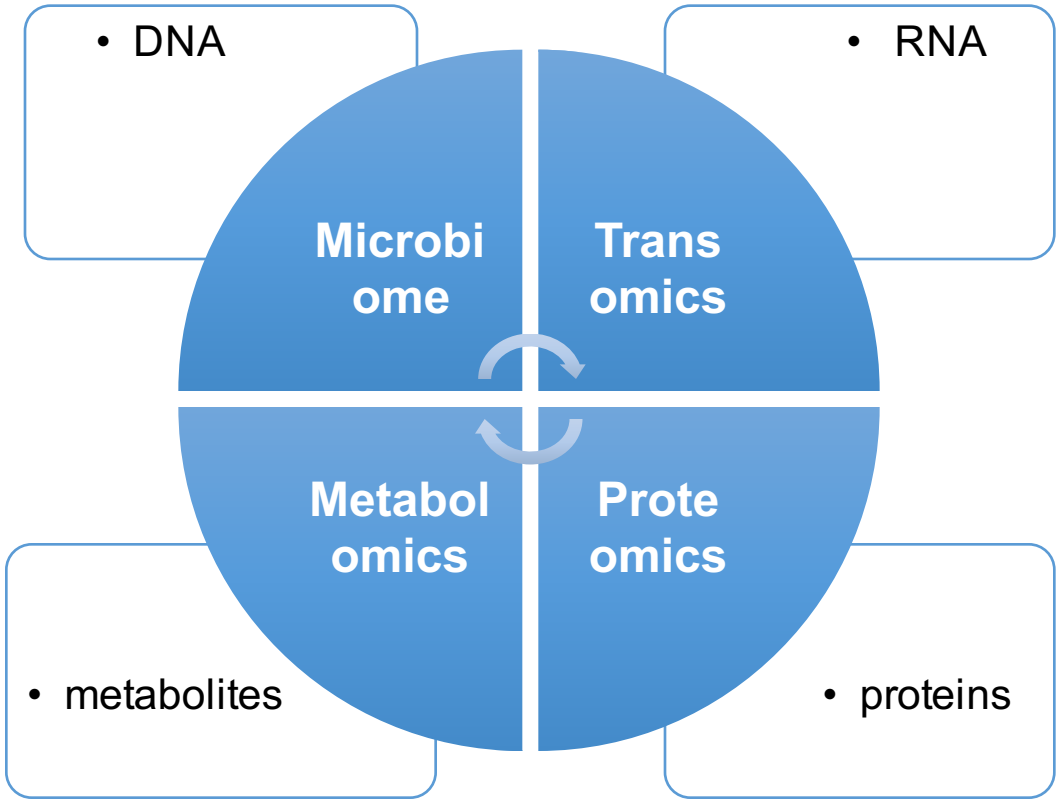
  maternal effects, gender, age, sex, technical sources of variation, season…...



Future

- **Limitations of 16s rRNA gene based methods:**
  - Species level
  - Can’t observe dynamic changes of different conditions
- **Future:**
  - **Multi-omic data** to measure functional activity under given condition

Multi-Omic data



Reference

1. Goodrich J K, Di Rienzi S C, Poole A C, et al. Conducting a microbiome study[J]. Cell, 2014, 158(2): 250-262.
2. Woo P C Y, Lau S K P, Teng J L L, et al. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories[J]. Clinical Microbiology and Infection, 2008, 14(10): 908-934.
3. Tyler A D, Smith M I, Silverberg M S. Analyzing the human microbiome: a “how to” guide for physicians[J]. The American journal of gastroenterology, 2014, 109(7): 983-993.
4. Franzosa E A, Hsu T, Sirota-Madi A, et al. Sequencing and beyond: integrating molecular'omics' for microbial community profiling[J]. Nature Reviews Microbiology, 2015.
5. Caporaso J G, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data[J]. Nature methods, 2010, 7(5): 335-336.
6. Loman N J, Misra R V, Dallman T J, et al. Performance comparison of benchtop high-throughput sequencing platforms[J]. Nature biotechnology, 2012, 30(5): 434-439.
7. Larsen N, Vogensen FK, van den Berg FW et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS One 2010;5:e9085.
8. Bahl MI, Bergstrom A, Licht TR. Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by down- stream quantitative PCR analysis. FEMS Microbiol Lett 2012;329:193–7.

Thank you!
Q&A