



Systems Biology Approaches for Microbiome Data Analysis

Translational Data Science Seminar

Jing Ma

Division of Public Health Sciences
Fred Hutchinson Cancer Research Center

23 March 2021

About Me



- Ph.D. in Statistics from U of Michigan
- Postdoc in Biostatistics from U of Pennsylvania
- Joined Fred Hutch in 2017
- Visited Texas A&M Statistics Dept during 2019-2020



REPORT

Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients

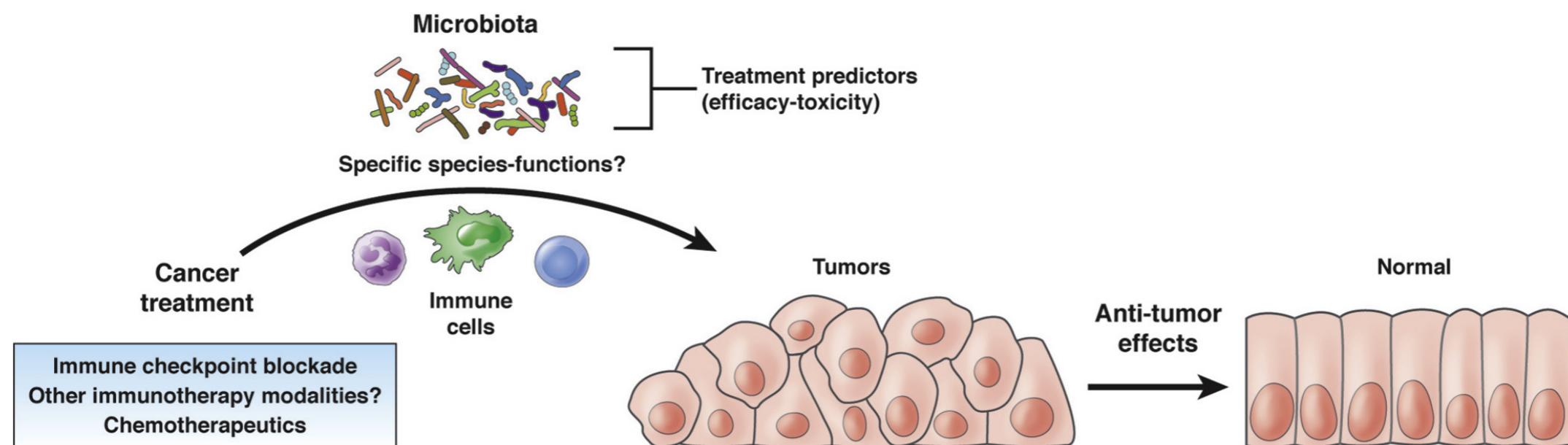
✉ Erez N. Baruch^{1,2,*†}, Ilan Youngster^{3,4}, Guy Ben-Betzel¹, Rona Ortenberg¹, Adi Lahat⁵, Lior Katz⁶, Katerina Adler⁷, Daniela Dick-Necula⁸, Stephen Raskin^{4,9}, Naamah Bloch¹⁰, Daniil Rotin⁸, Liat Anafi⁸, Camila Avivi⁸, Jenny Melnichenko¹, Yael Steinberg-Silman¹, Ronac Mamtani¹¹, Hagit Harati¹, Nethanel Asher¹, Ronnie Shapira-Frommer¹, Tal Brosh-Nissimov¹², Yael Eshet^{4,8,13}, Shira Ben-Simon¹⁰, Oren Ziv¹⁰, Md Abdul Wadud Khan¹⁴, Moran Amit¹⁵, Nadim J. Ajami¹⁴, Iris Barshack^{4,8}, Jacob Schachter^{1,4}, Jennifer A. Wargo^{14,16}, Omry Koren¹⁰, Gal Markel^{1,2,17,*‡}, Ben Boursi^{4,18,19,‡}

RESEARCH ARTICLE

Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients

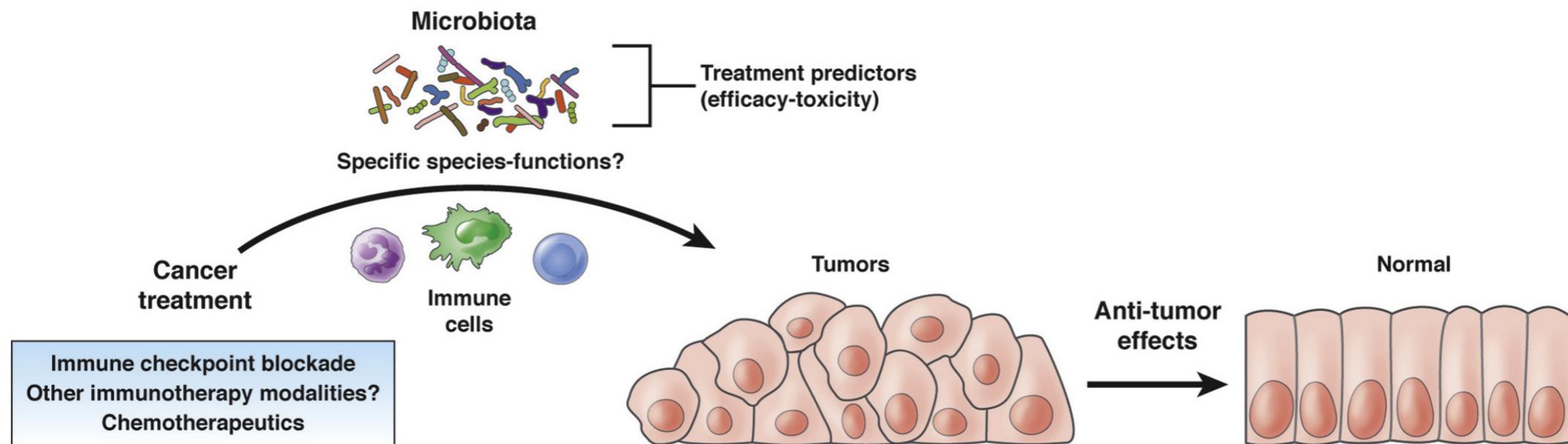
✉ Diwakar Davar^{1,*}, Amiran K. Dzutsev^{2,*}, John A. McCulloch², Richard R. Rodrigues^{2,3}, Joe-Marc Chauvin¹, Robert M. Morrison¹, Richelle N. Deblasio¹, Carmine Menna¹, Quanquan Ding¹, Ornella Pagliano¹, Bochra Zidi¹, Shuowen Zhang^{1,†}, Jonathan H. Badger², Marie Vetizou², Alicia M. Cole², Miriam R. Fernandes², Stephanie Prescott², Raquel G. F. Costa², Ascharya K. Balaji², Andrey Morgun⁴, Ivan Vujkovic-Cvijin⁵, Hong Wang⁶, Amir A. Borhani⁷, Marc B. Schwartz⁸, Howard M. Dubner⁸, Scarlett J. Ernst¹, Amy Rose¹, Yana G. Najjar¹, Yasmine Belkaid⁵, John M. Kirkwood¹, Giorgio Trinchieri^{2,‡,§}, Hassane M. Zarour^{1,9,‡,§}

Microbiome and Immunotherapy



D'Haens and Jobin (19) Gastroenterology

Microbiome and Immunotherapy



Unclear what aspects of the donor/recipient microbiota contribute to favorable response: *diversity?* *Specific taxa?* *Bacterial metabolites?*

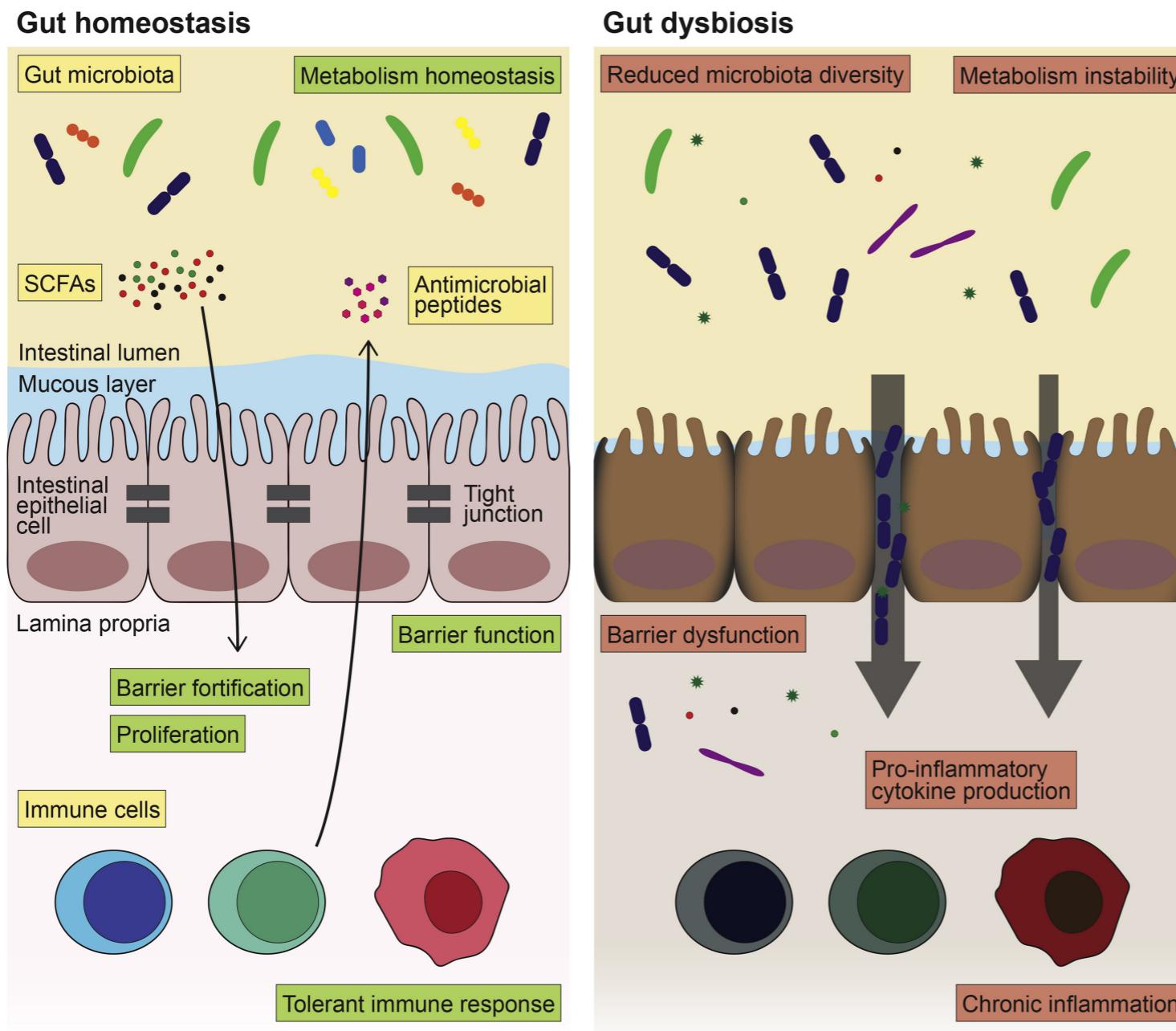
What about host attributes?

D'Haens and Jobin (19) Gastroenterology

Microbiome and Aging



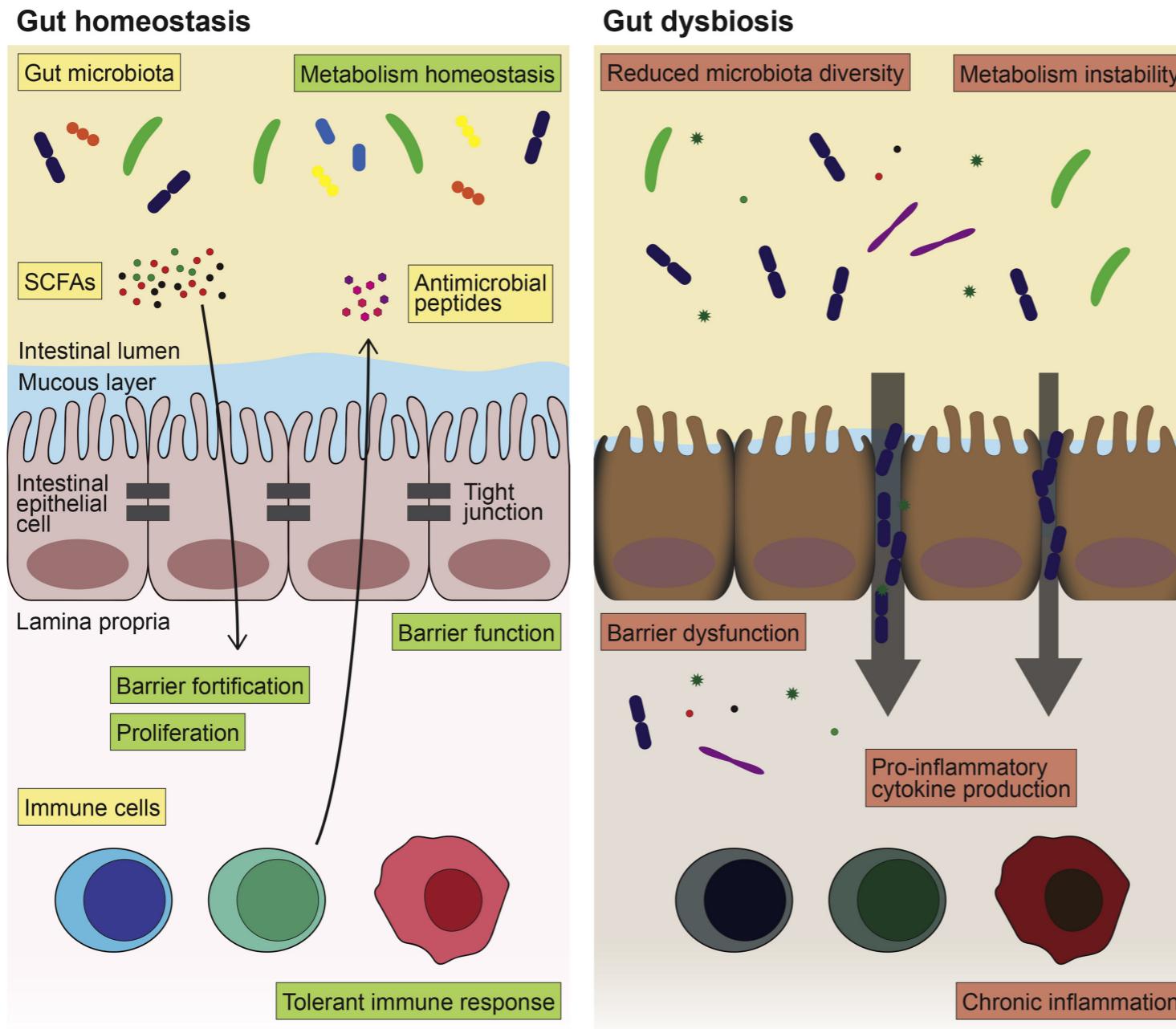
Aging \neq Chronological age



Kim and Benayoun (20) Transl. Med. of Aging

Microbiome and Aging

Aging \neq Chronological age

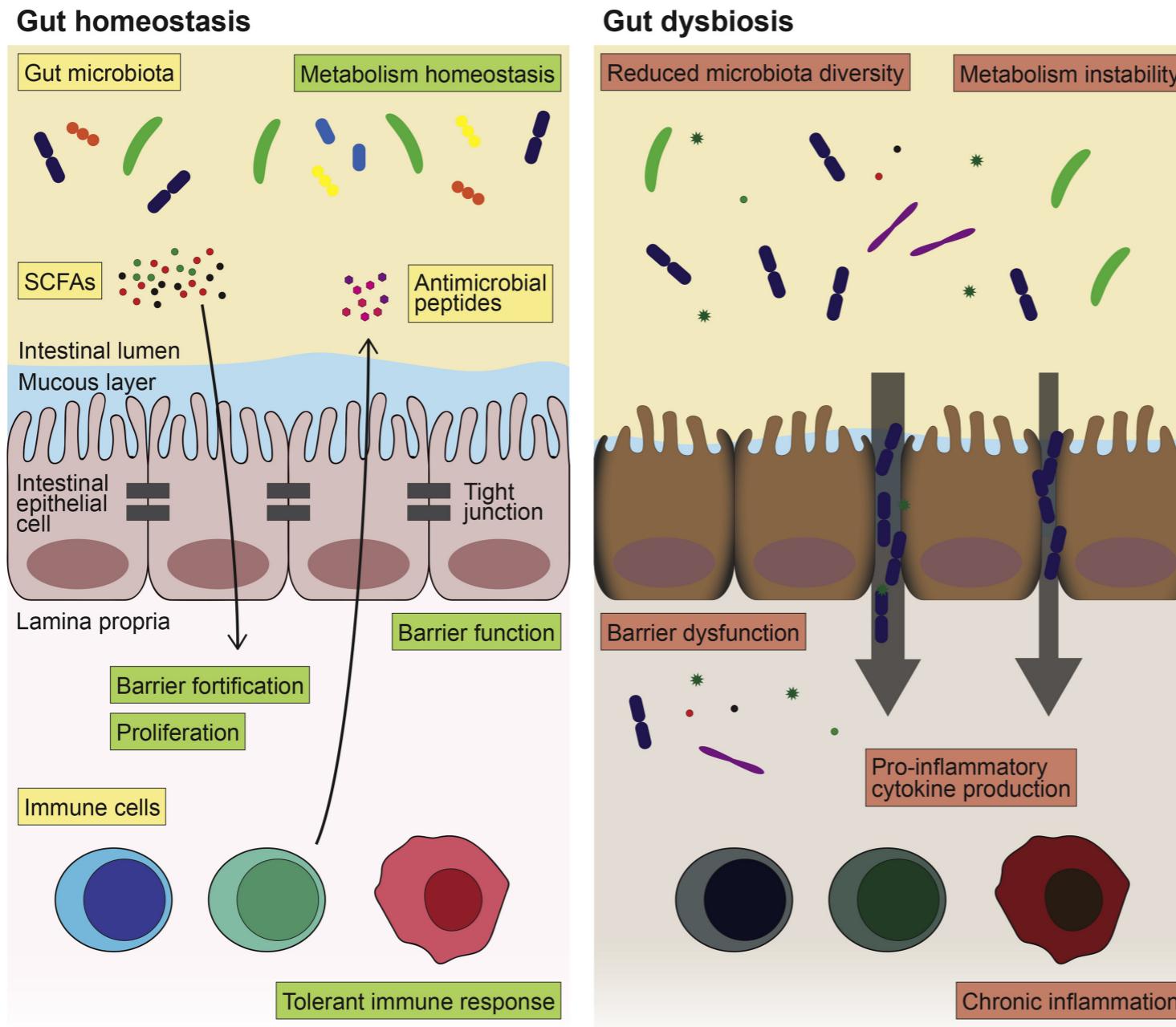


What aspects of the gut microbiome and/or their products predict healthy aging?

Kim and Benayoun (20) Transl. Med. of Aging

Microbiome and Aging

Aging \neq Chronological age



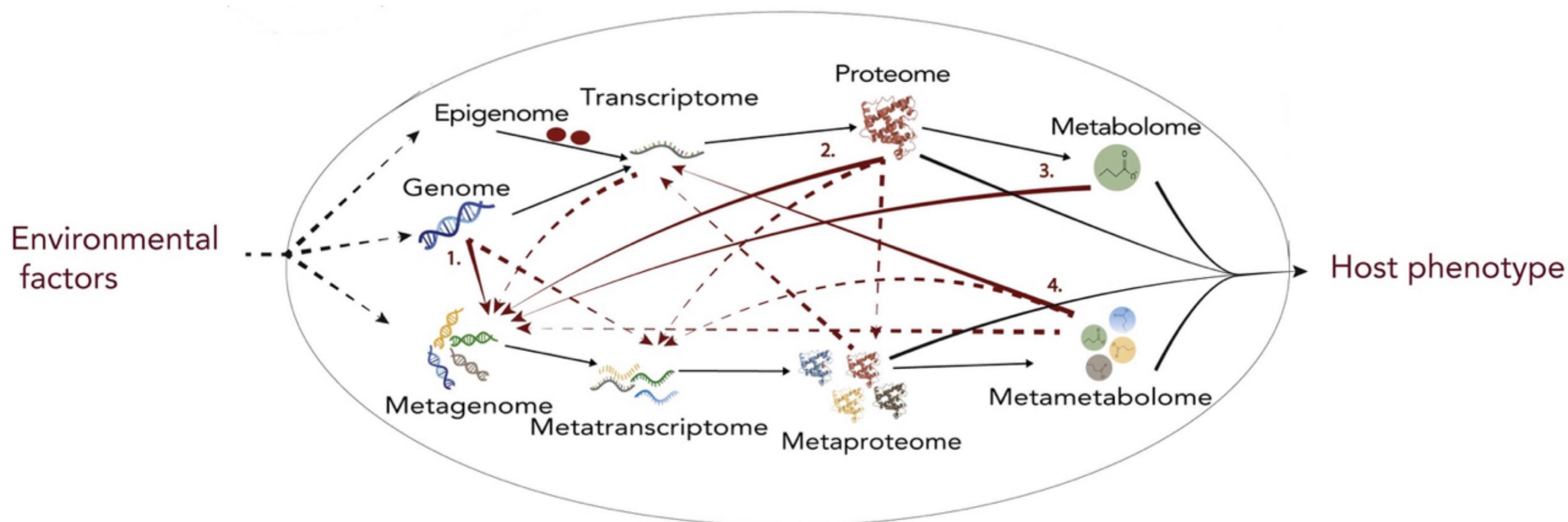
What aspects of the gut microbiome and/or their products predict healthy aging?

A multi-omics approach will provide unique insights.

Systems Biology and Holo-Omics



FRED HUTCH
CURES START HERE™



Host-microbiota multi-omics

Nyholm et al. (20) iScience

Two Problems of Interest



1. Interactions between microbial taxa and other molecules
2. Individual taxon association analysis

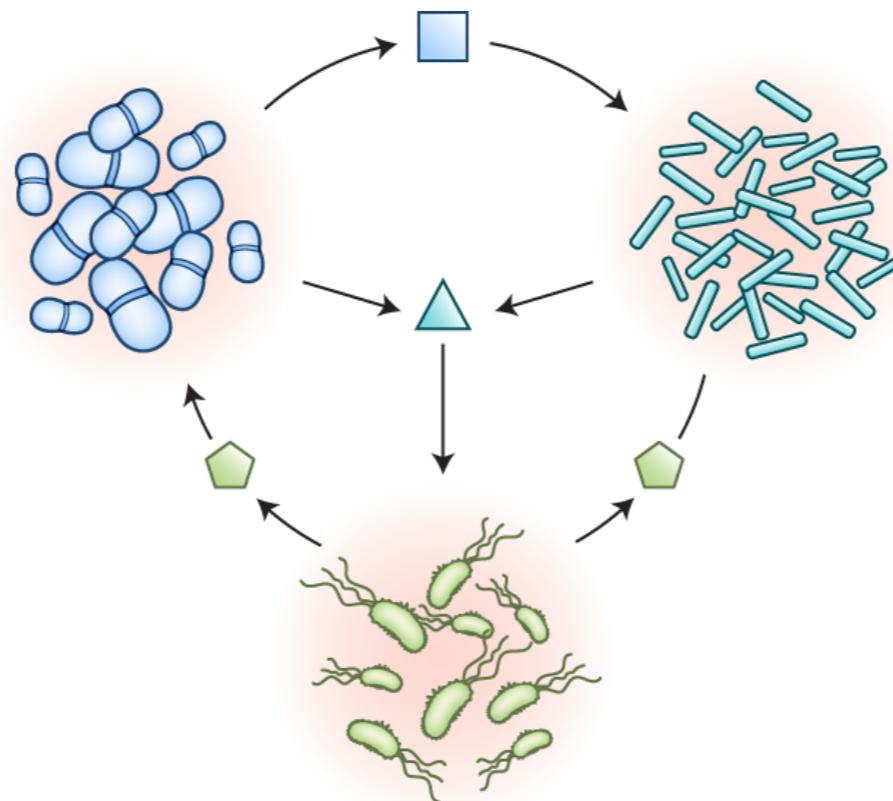
Part I

Interactions between microbial taxa and other molecules

Microbial Interactions



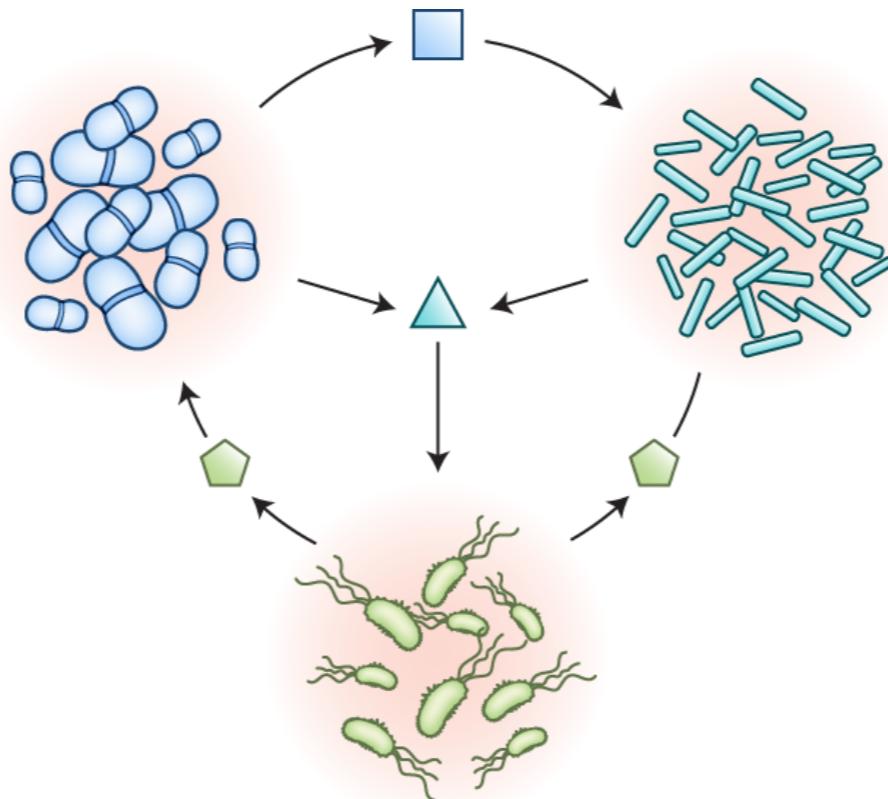
FRED HUTCH
CURES START HERE™



Tang (19) Nat. Meth.

Credit : Marina Corral Spence/Springer Nature

Microbial Interactions



Uncover community ecology and functions

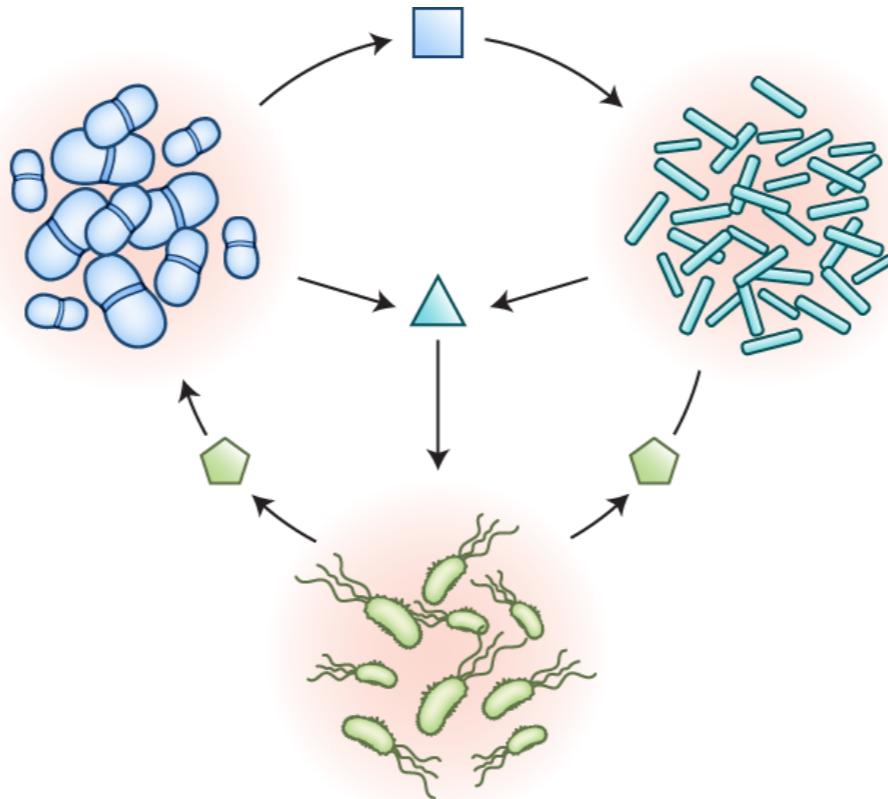
Tang (19) Nat. Meth.

Credit : Marina Corral Spence/Springer Nature

Microbial Interactions



FRED HUTCH
CURES START HERE™



Uncover community ecology and functions

Predict host phenotype and disease states

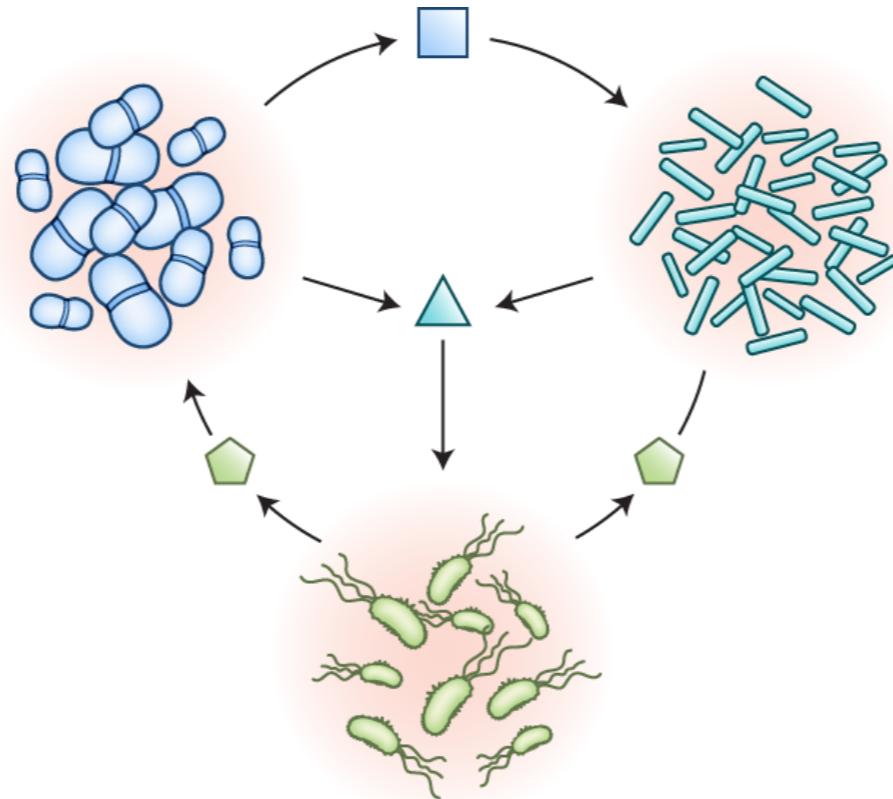
Tang (19) Nat. Meth.

Credit : Marina Corral Spence/Springer Nature

Microbial Interactions



FRED HUTCH
CURES START HERE™



Uncover community ecology and functions

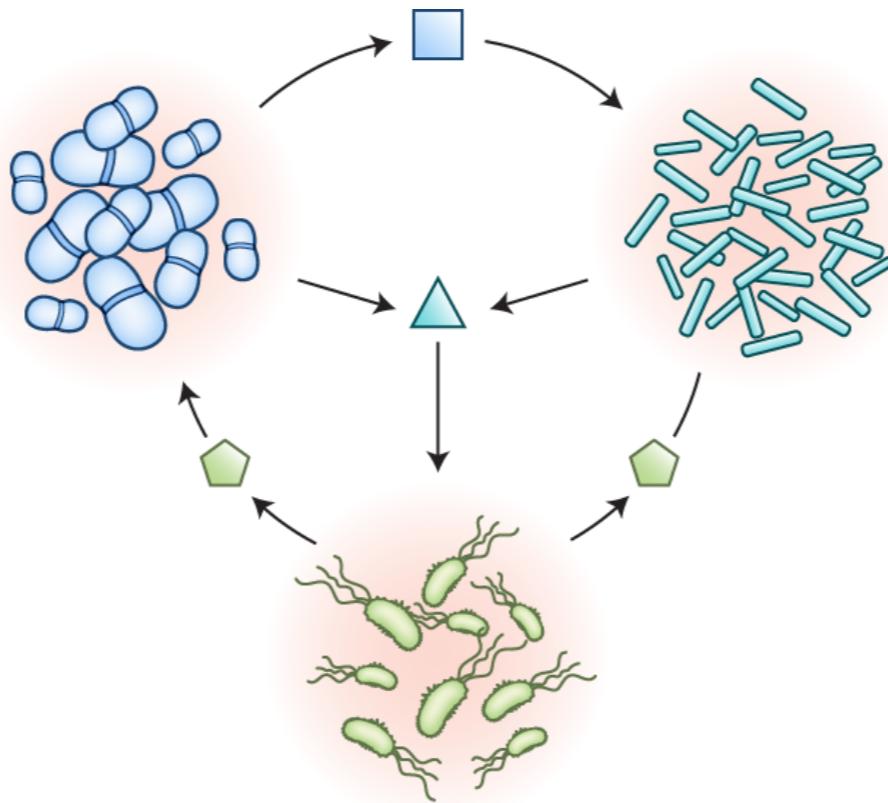
Predict host phenotype and disease states

Biomarker discovery by mapping differential interactions

Tang (19) Nat. Meth.

Credit : Marina Corral Spence/Springer Nature

Learn Microbial Interactions

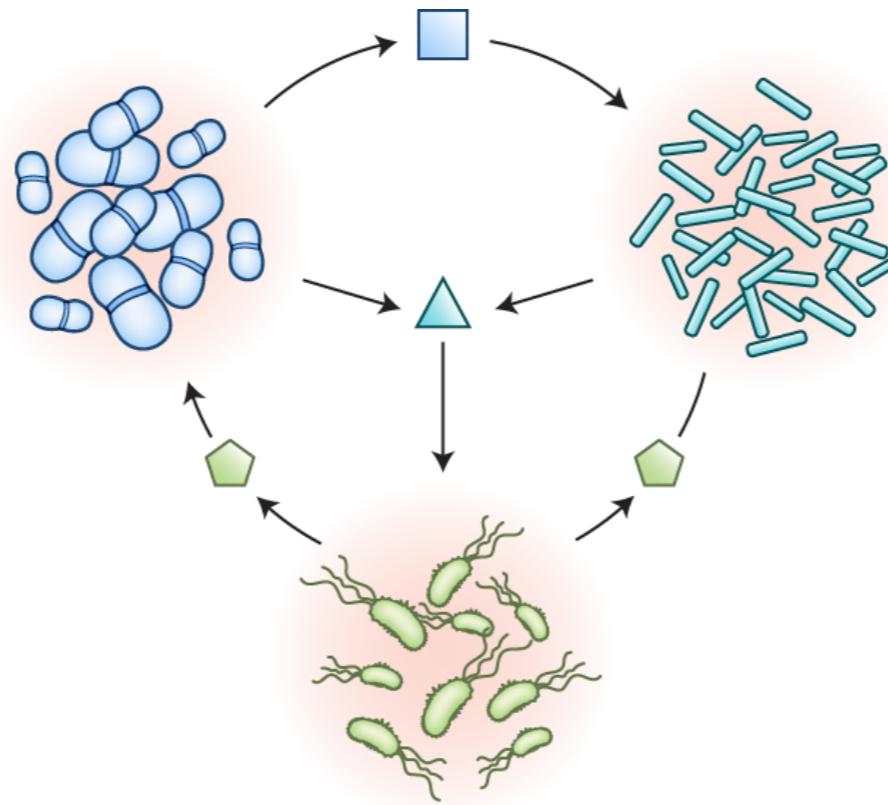


Challenge: majority of microbial species can't be cultured.

Tang (19) Nat. Meth.

Credit : Marina Corral Spence/Springer Nature

Learn Microbial Interactions



Challenge: majority of microbial species can't be cultured.

Solution: computational methods help understand how microbes interact with each other, and with other molecules.

Tang (19) Nat. Meth.

Credit : Marina Corral Spence/Springer Nature

Probabilistic co-abundance networks

- Uncover nonrandom associations.
- Predict novel ecological interactions, compared to mechanistic models.
- Detect altered co-abundances associated with an outcome.

Co-abundance Networks



Marginal correlations (ReBoot, SparCC, etc.)

- Easy to infer.
- Can't distinguish direct vs indirect interactions.



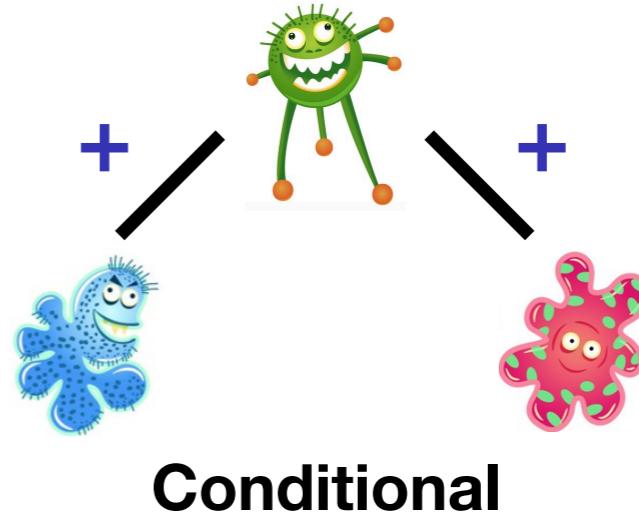
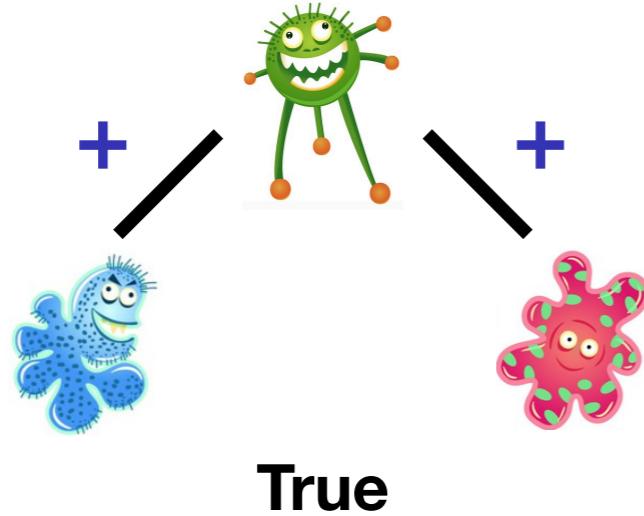
Faust and Huttenhower (12), Friedman and Alm (12). PLoS Comp Bio.

Co-abundance Networks



Conditional correlations (SpiecEasi, BAnOCC, etc.)

- Can distinguish direct vs indirect interactions.
- Typically require stronger model assumptions.

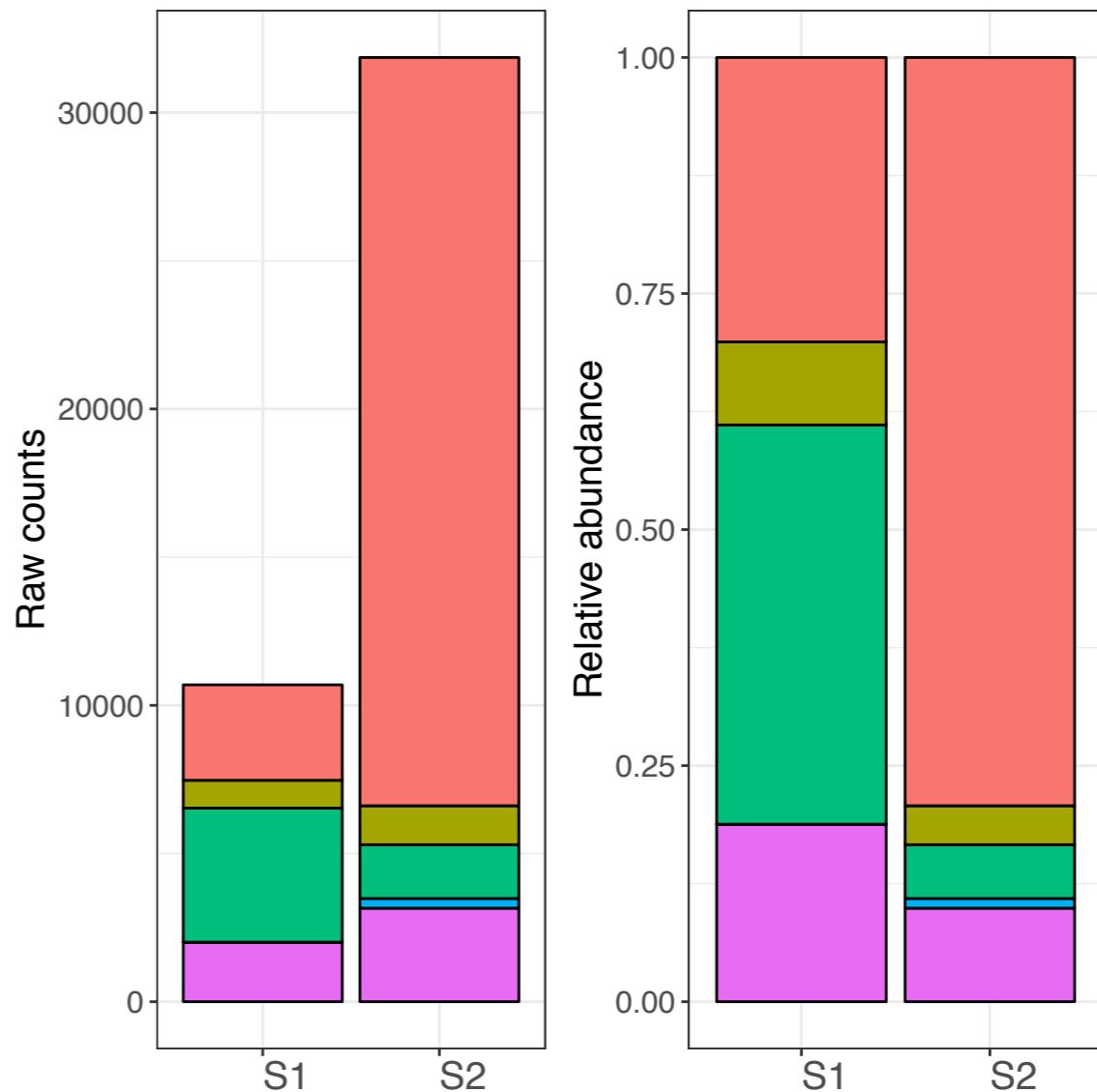


Kurtz et al. (15), Schwager et al. (17). PLoS Comp Bio.

Microbiome Data are Compositional



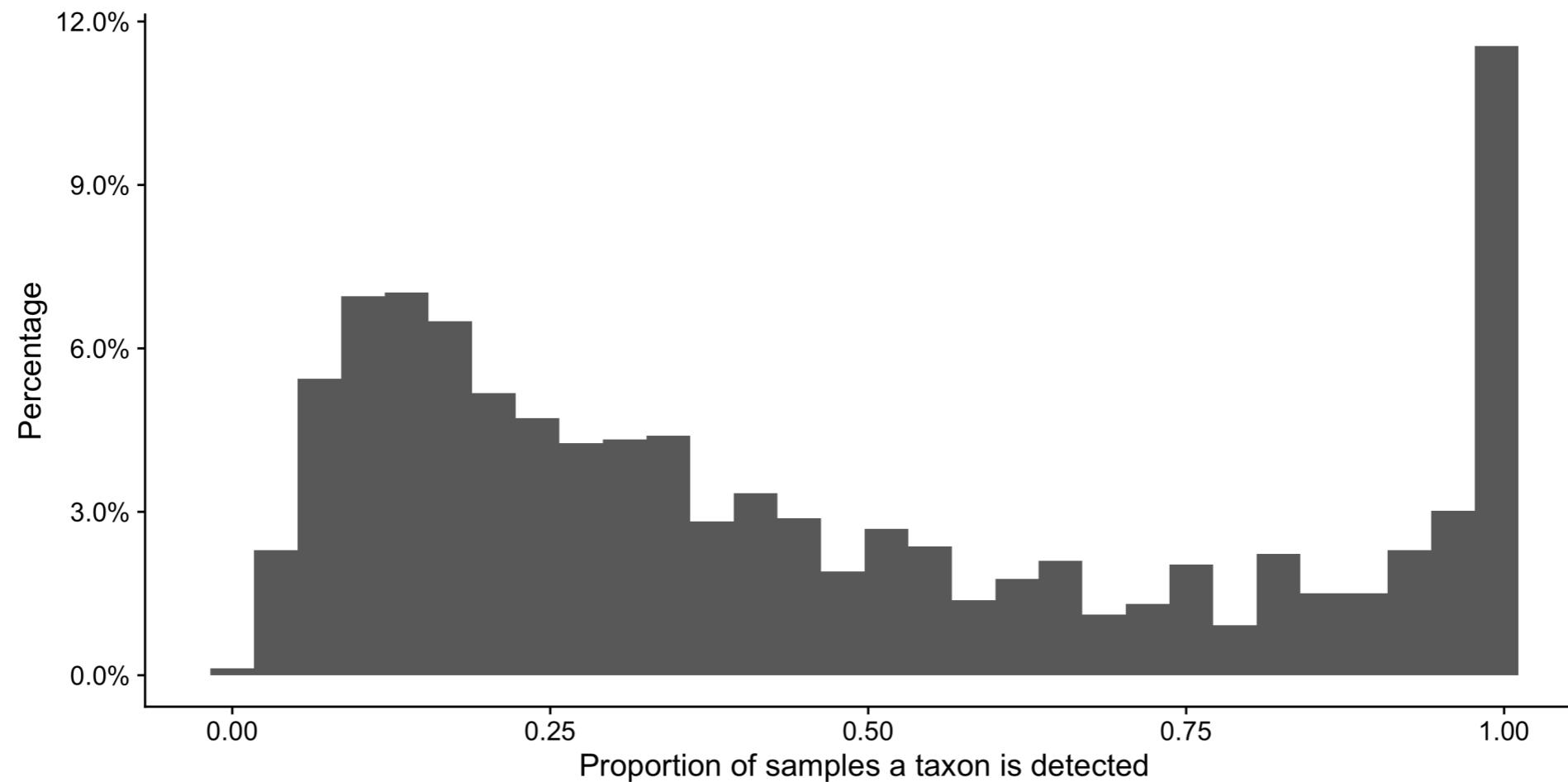
- Hard to control total number of reads in one experiment
- Transform counts to proportions: **percent composition**
- Relative abundances are also used in metabolomics.



Microbiome Data are Sparse



Observation based on cancer microbiome data $p = 1500$.

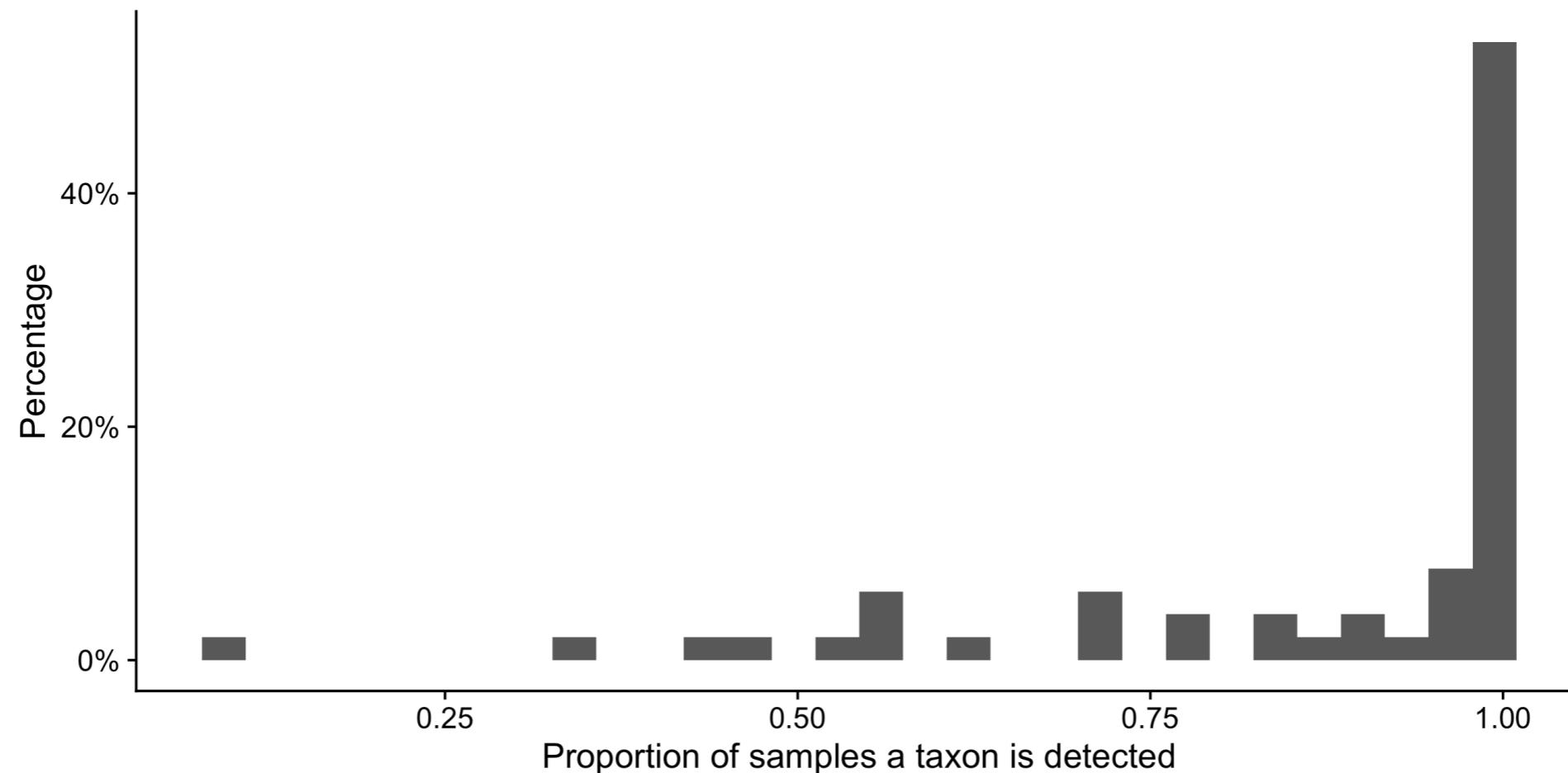


Poore et al. (20) Nature

Microbiome Data are Sparse



Observation based on vaginal microbiome data $p = 51$.

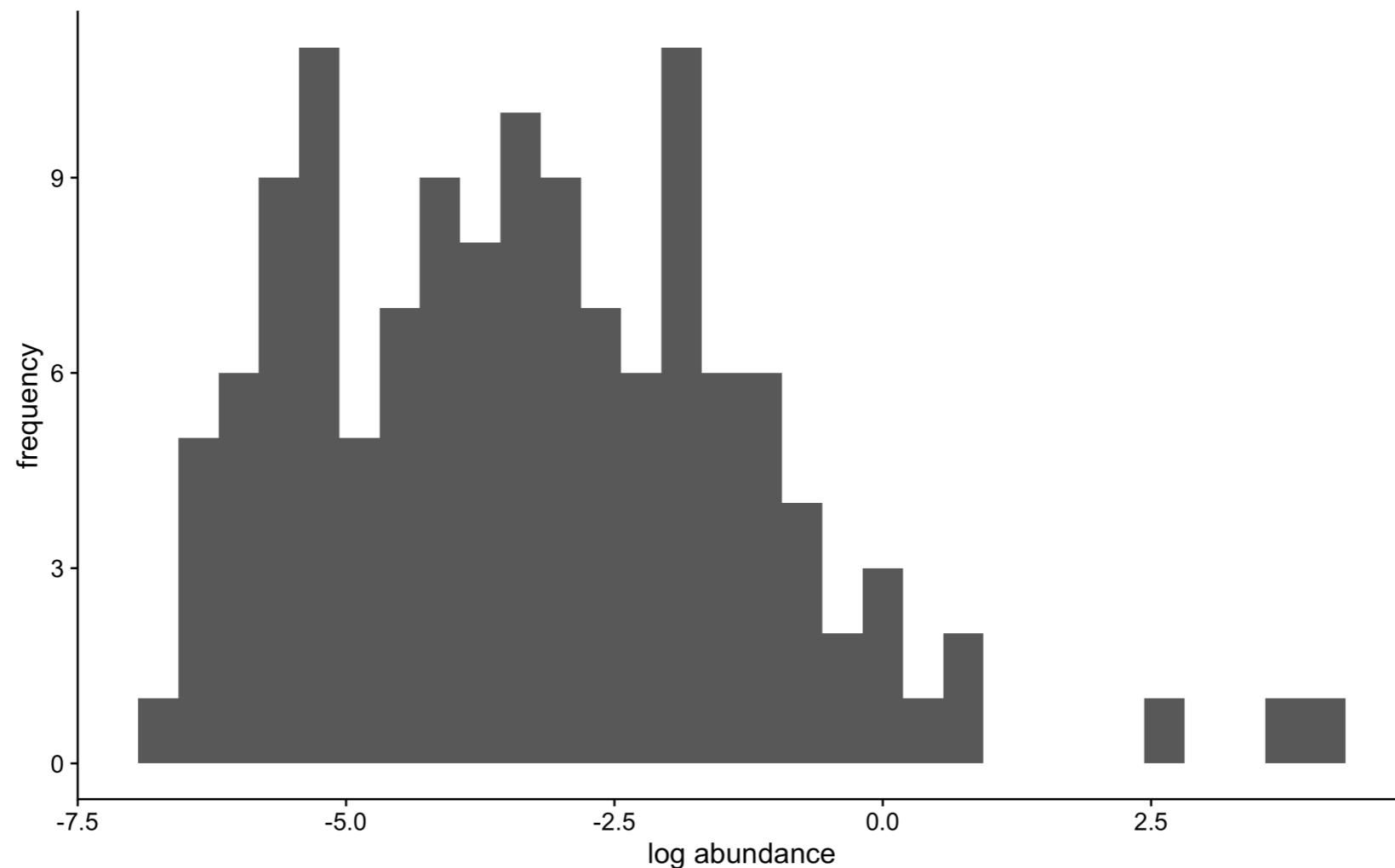


McMillian et al. (15) Scientific Reports

Centered Log-Ratio Transformation



Marginal distribution of the genus *Porphyromonadaceae* based on 131 observations in McMillian et al. (15).

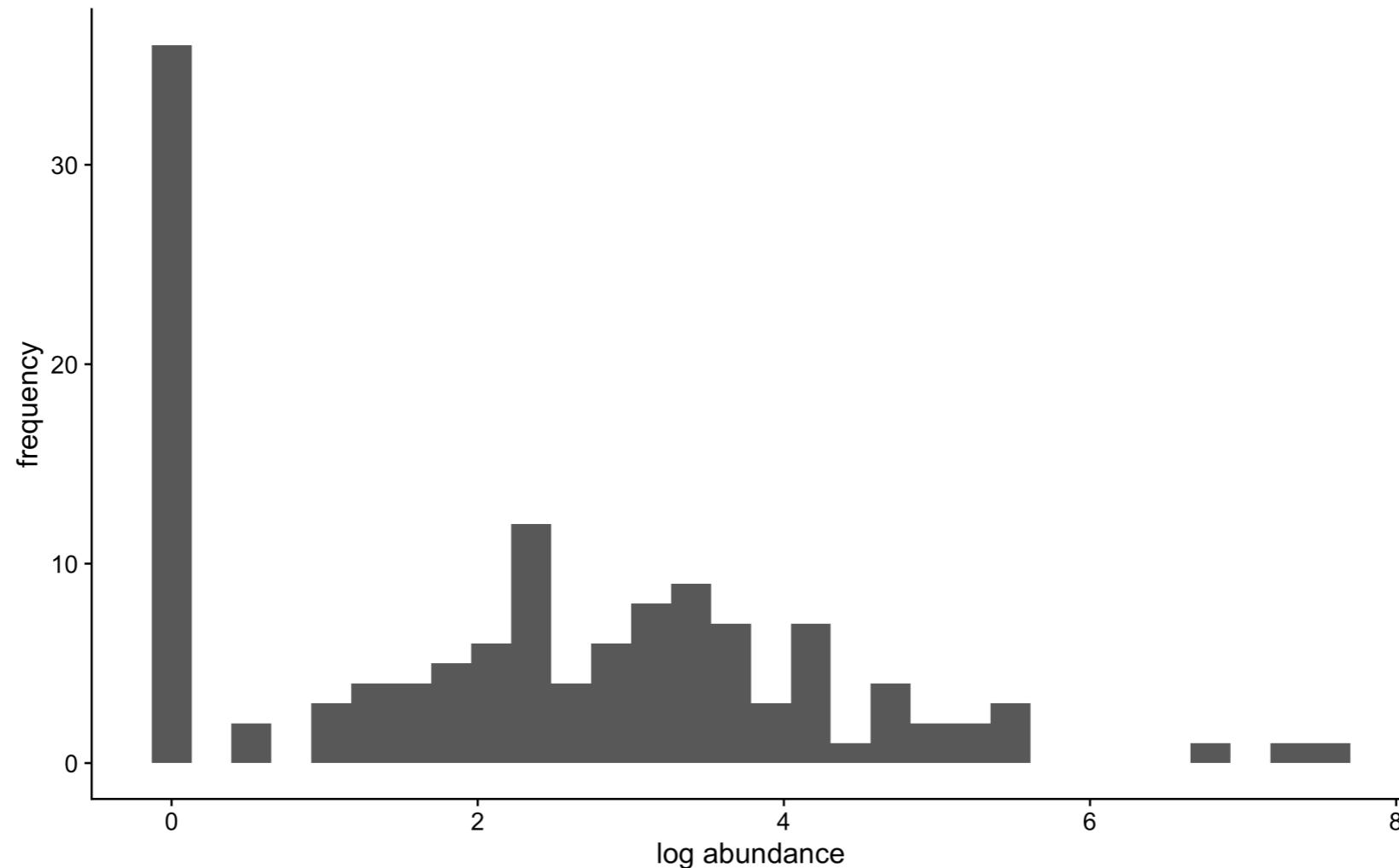


Inappropriate handling of zeros leads to extremely non-normal distribution.

Robust CLR Transformation



Marginal distribution of the genus *Porphyromonadaceae* based on 131 observations in McMillian et al. (15).



Robust CLR transforms **positive counts** only.

Zeros are assumed to be due to undersampling (censoring).

$$X^o = X^*(X^* > 0)$$


Observed Latent truth

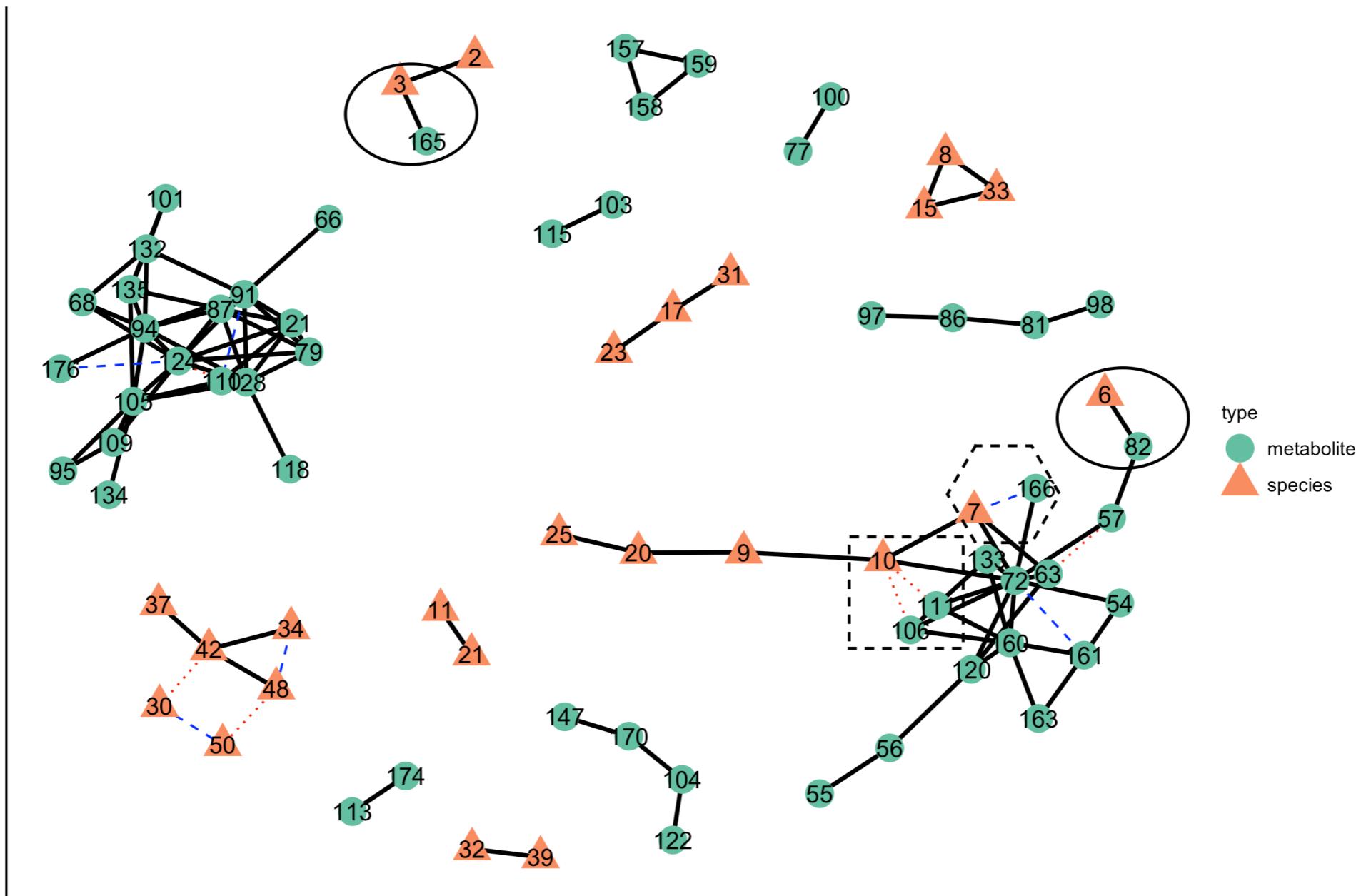
metaMint

- Learn associations among latent true log abundances from observed data
- Apply regularization to deal with high-dimensionality.

metaMint Facilitates Data Integration



Results based on BV microbiome and metabolomic data in McMillian et al. (15).
Dashed blue edges are unique to metaMint.



Part II

Individual taxon association analysis

Selection of Individual Taxon



Outcome

Taxa



Scientific Question: which taxa predict the outcome?

Selection of Individual Taxon



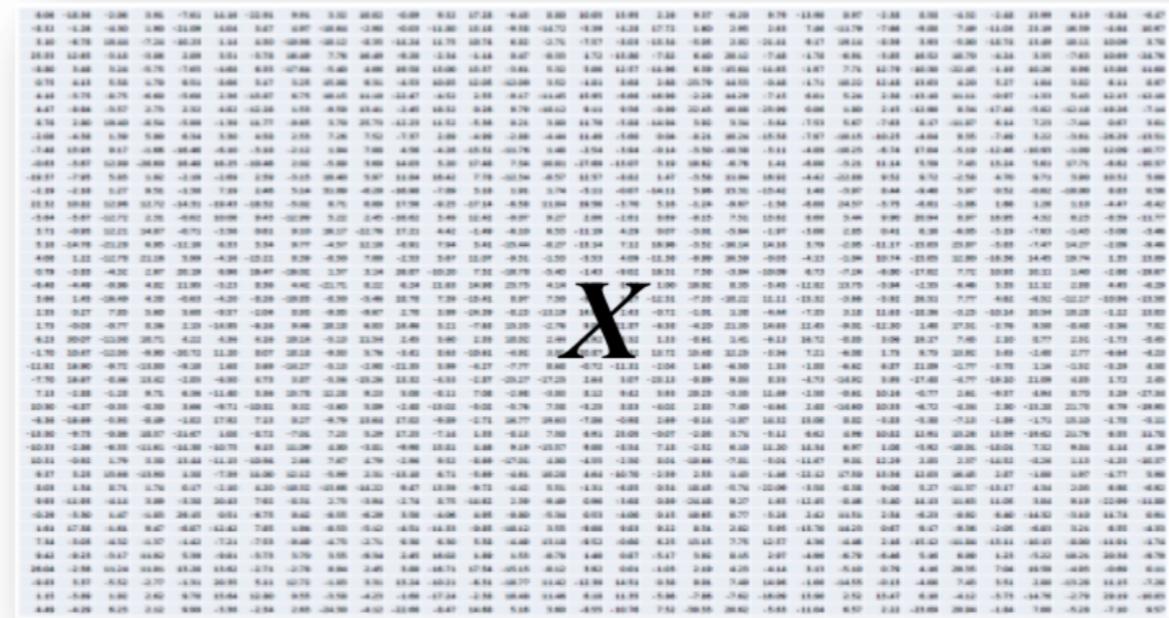
Outcome



?

← n samples →

Taxa



Scientific Question: which taxa predict the outcome?

Common Challenge: more predictors than observations.

Selection of Individual Taxon



Common Challenge: more predictors than observations.

Solutions:

- Dimension reduction on predictors

Selection of Individual Taxon



Common Challenge: more predictors than observations.

Solutions:

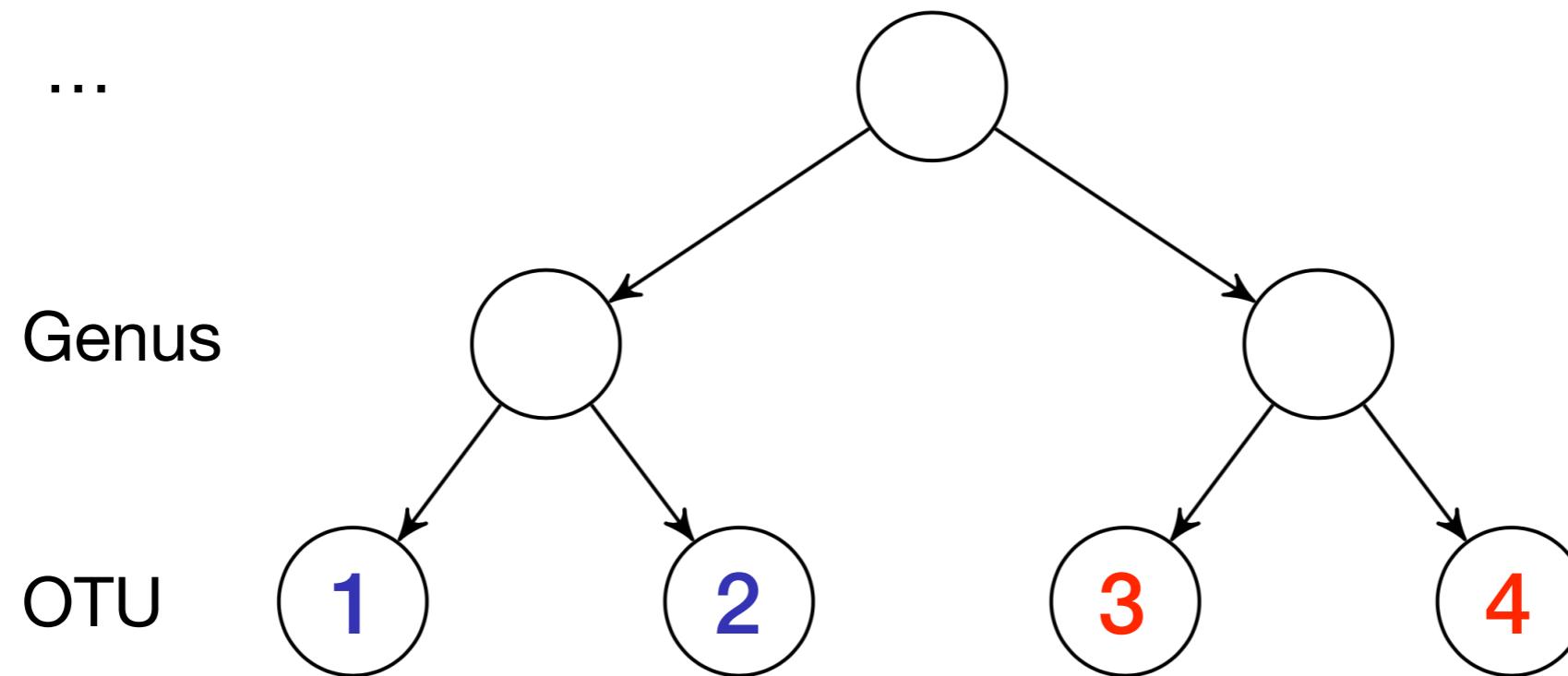
- Dimension reduction on predictors.
- Assume effects are sparse: a small number of taxa have nonzero effects on the outcome.

Common Challenge: more predictors than observations.

Solutions:

- Dimension reduction on predictors
- Assume effects are sparse: a small number of taxa have nonzero effects on the outcome.
- Assume effects are smooth with respect to prior structures (e.g. the phylogenetic tree)

Informative Prior Structure

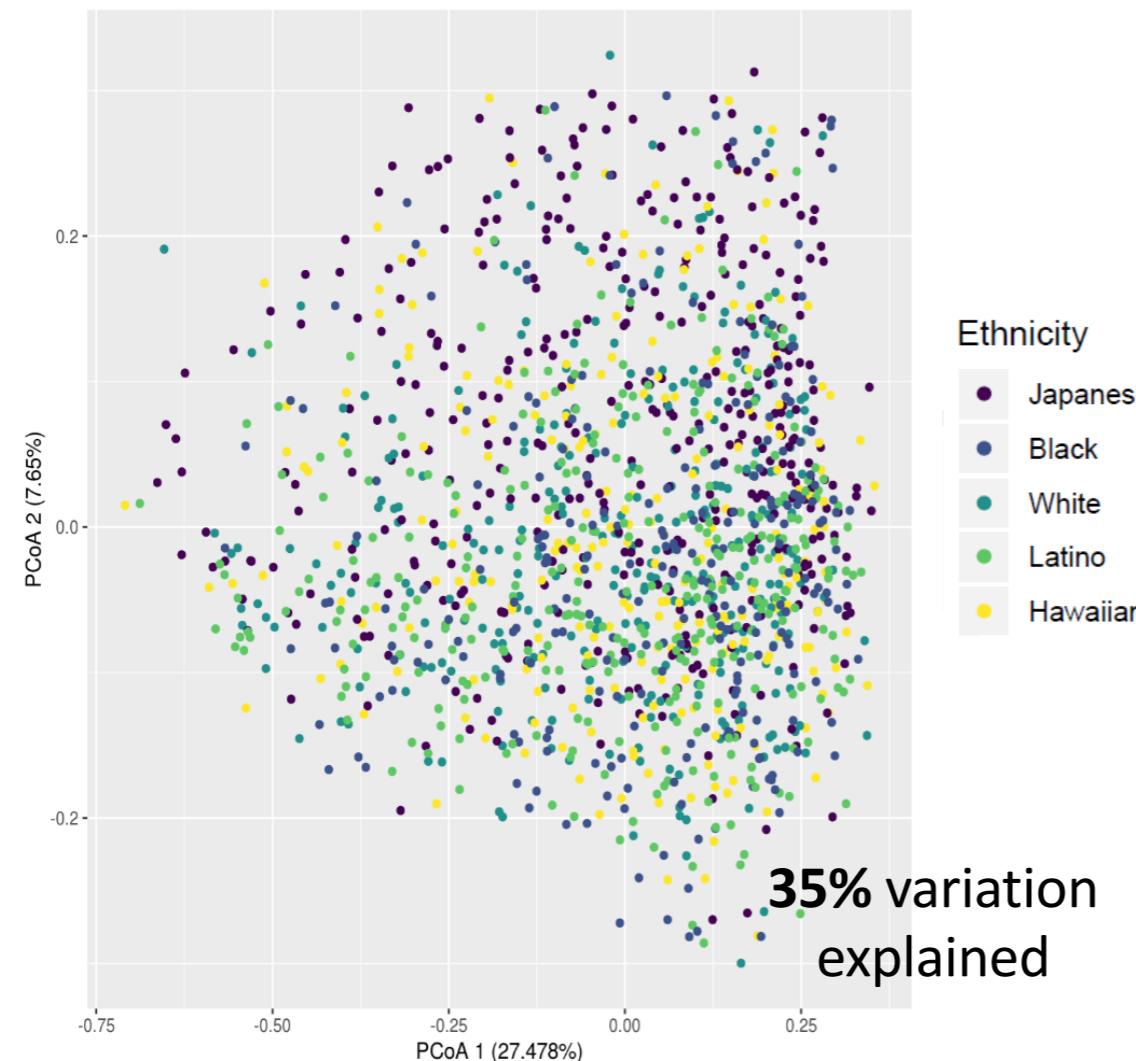


If the tree is informative, then it is reasonable to think that **taxa closely related on the tree have similar effects on the outcome.**

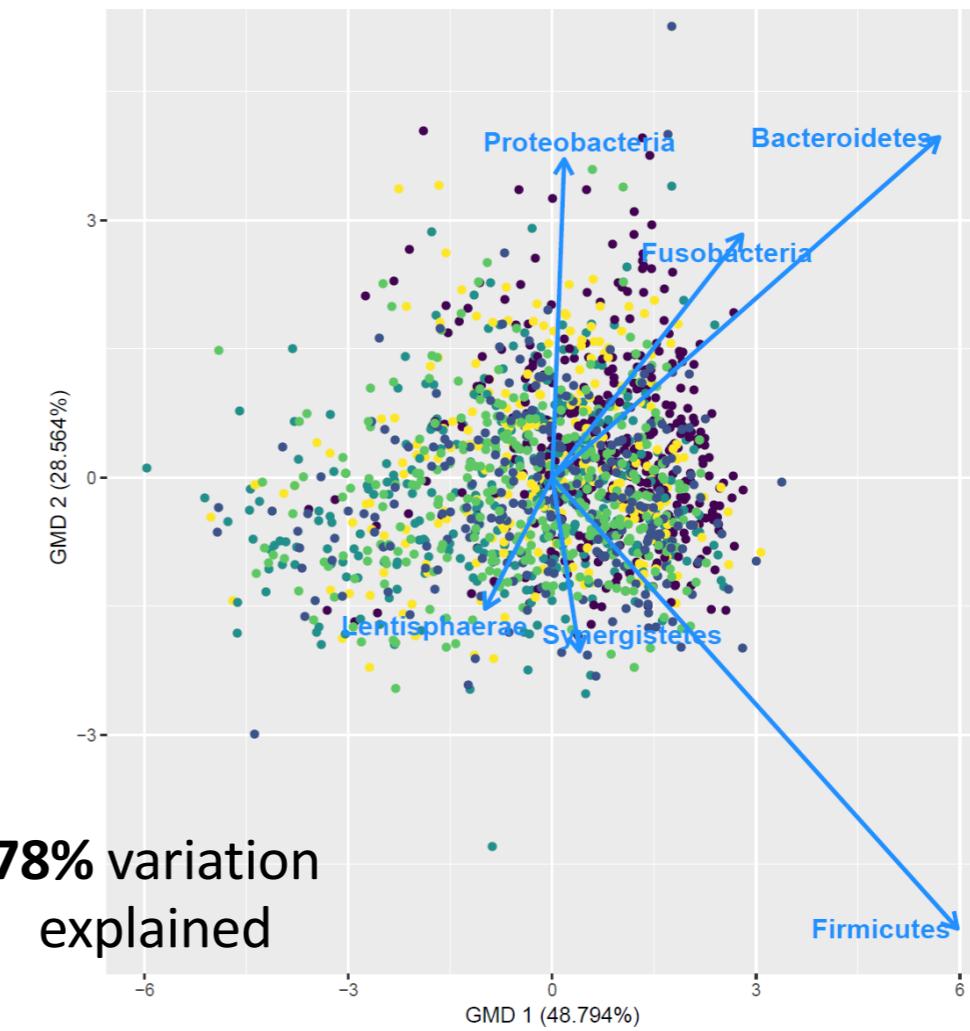
GMD-biplot



PCoA using UniFrac Δ



PCoA using X and UniFrac Δ ; phyla (arrows) from GMD biplot



Plots courtesy to Tim Randolph

Informative Two-way Structures



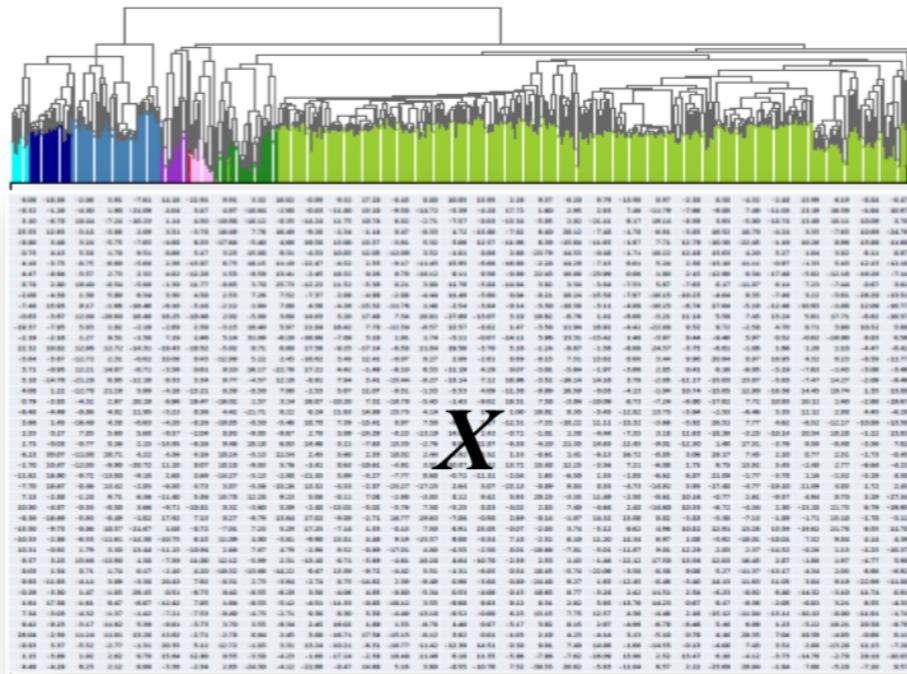
FRED HUTCH
CURES START HERE™

Outcome $Q =$

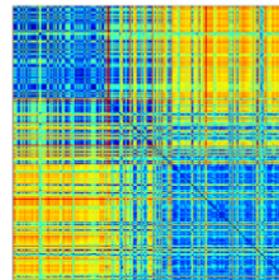


\mathbf{z}

$\leftarrow n \text{ samples} \rightarrow$



$(n \times n)$



Similarity Kernel $\mathbf{H} = -\frac{1}{2} \mathbf{J} \Delta \mathbf{J}$

Informative Two-way Structures



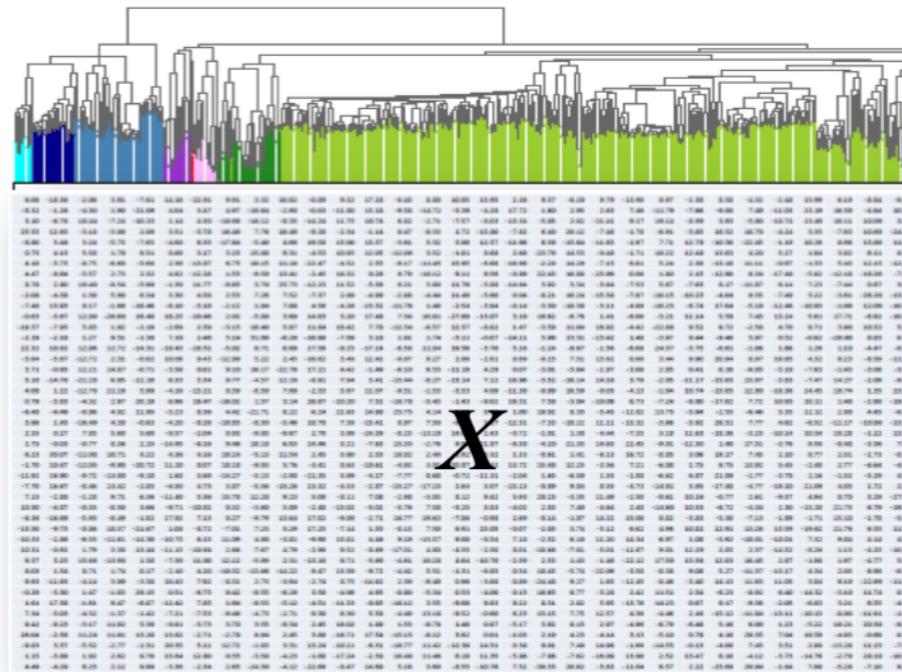
FRED HUTCH
CURES START HERE™

Outcome $Q =$



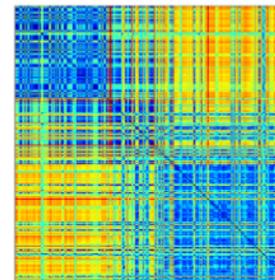
\mathbf{z}

← n samples →



($n \times n$)

Similarity Kernel $\mathbf{H} = -\frac{1}{2} \mathbf{J} \Delta \mathbf{J}$



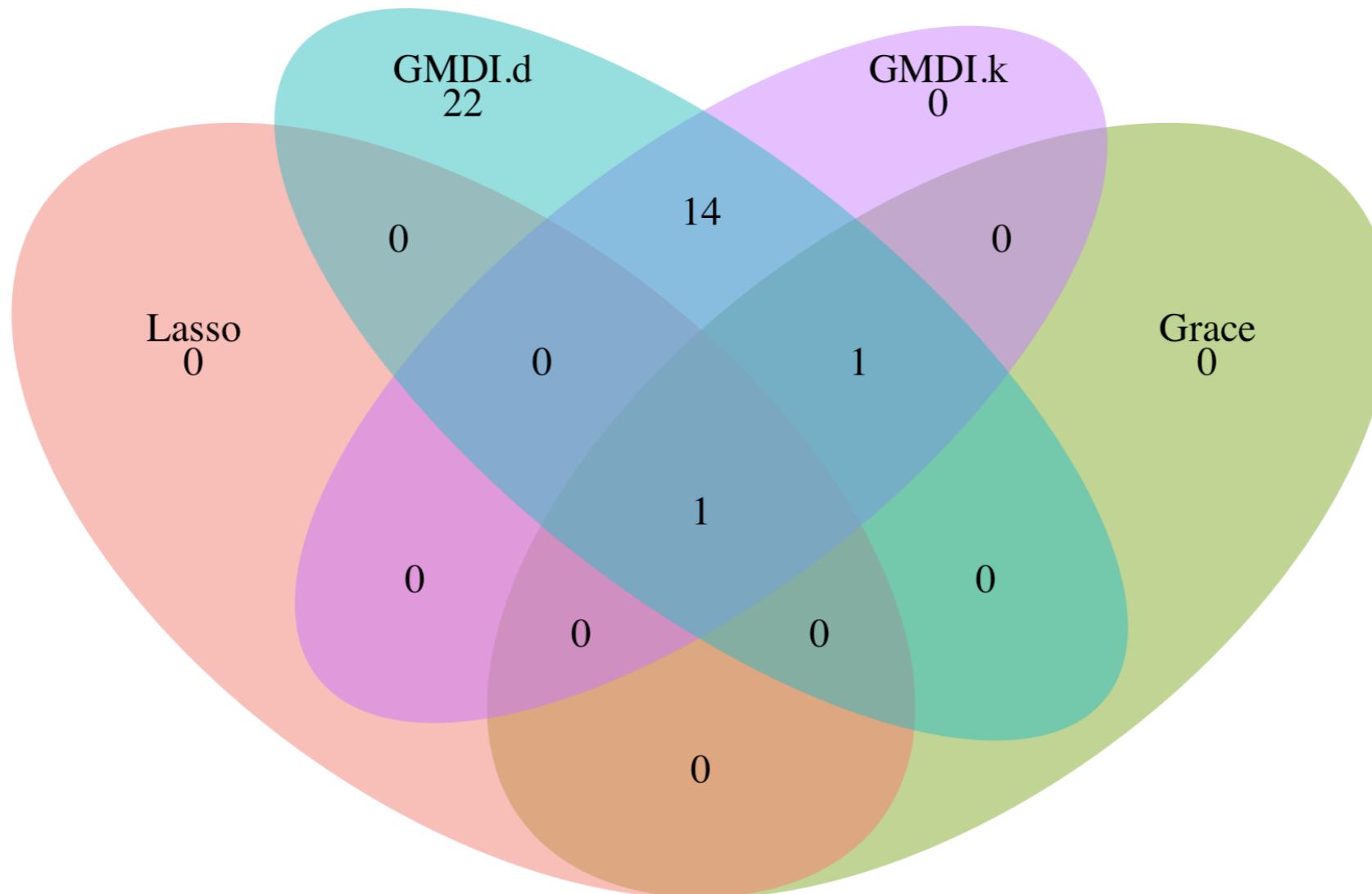
Similarity Kernel can come from

- Longitudinal or family design
- Another data view (e.g. metabolomic data)

GMDI Association Analysis



Which bacteria are associated with age?



**Number of sig. associations (FDR = 0.1) based on data from
Yatsunenko et al. (12) Nature: n = 100, p = 149.**

Summary



Accounting for correlations among features and observations yield more powerful inferences.

- [metaMint](#) identifies novel microbe-metabolite interactions.
- [GMD-biplot](#) provides more interpretative visualization.
- [GMDI](#) is more powerful in detecting taxa-outcome associations.

Collaboration



- Systems biology
- Multivariate association analysis
- Network analysis
- Statistical learning

Collaboration



- Systems biology
- Multivariate association analysis
- Network analysis
- Statistical learning

Large sample size is great, but not necessary!

Acknowledgement



Yue Wang



Ali Shojaie



Tim Randolph

- The GMD-biplot and its application to microbiome data. *mSystems*. 2019
- Joint microbial and metabolite network estimation with the censored Gaussian graphical model. *Statistics in Biosciences*. 2020
- Generalized matrix decomposition: estimation and inference for two-way structured data. 2021+