Integrative and multi'omic analyses of microbial communities and the human microbiome

Curtis Huttenhower

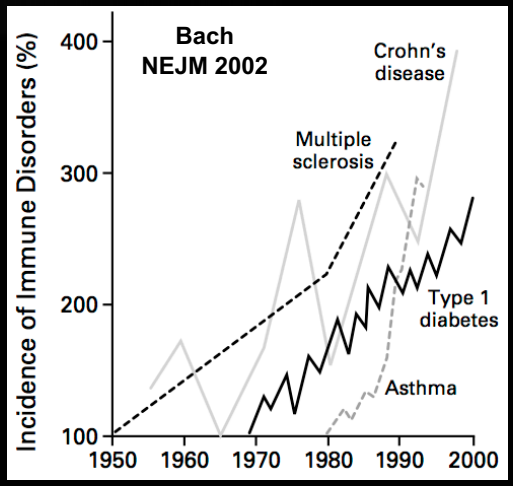
Harvard T.H. Chan School of Public Health
Department of Biostatistics

07-31-19

BROAD INSTITUTE

The microbiome in population health

bioBakery
A virtual environment for meta'omic analysis



Interdisciplinarity

- Microbiology
- Computer science
- Immunology
- Biostatistics
- Epidemiology
- Infectious disease
- Ecology
- Women's health
- Genetics
- …

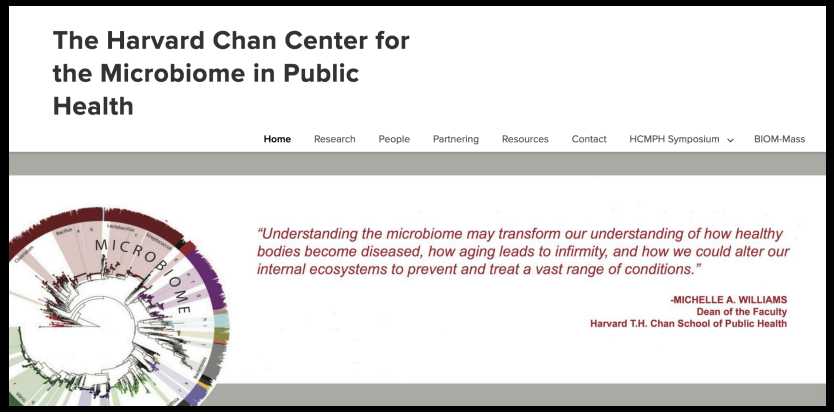
Basic science

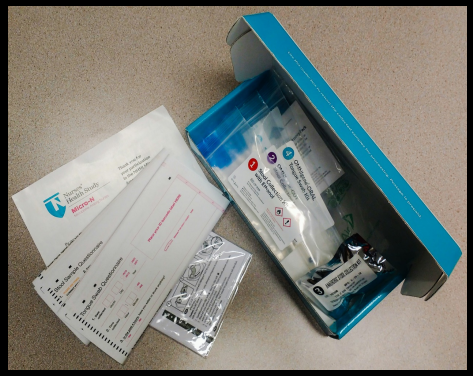
- New bugs, new genes, new molecules
- Unique statistical and machine learning challenges
- Technology and biomedical engineering development

Impact

- Entirely new ways to influence and measure health
  - Microbial biomarkers
  - Novel bioactives
  - FMTs / engineered organisms / communities
- New translational routes in almost every infectious, immune, cardiometabolic condition
- Industry and entrepreneurship opportunities

The microbiome in population health





- HCMPH: Program for academic collaboration, industry partnerships, technology transfer, faculty and student recruiting…
  - Gnotobiotic and analysis service cores, partnership with Broad for 'omics
  - BIOM-Mass: Biobank for Microbiome Research in Massachusetts
- Standardized room-temperature, USPS-compatible kit for:
  - Stool and oral samples.
  - Metagenomics, metatranscriptomics, metabolomics, and culture / gnotobiotics.
- Automation for ~3,000,000 sample aliquotting and -80C storage.
- Initial "flagship" MICRO-N collection targeting 25,000 women from NHSII.


Wendy Garrett

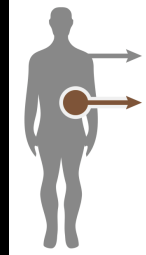

Andy Chan


Eric Rimm

Multi'omics for microbiome epidemiology


Eric Franzosa


Himel Mallick



A Sampling and Multi'omic Profiling Systems in Microbiome Epidemiology

Metagenomics (Microbial species, strains, and gene products)

Metatranscriptomics (Microbial gene expression profiles)

Metaproteomics (Microbial protein profiles)

Metametabolomics (Microbial metabolite concentrations)

Metadata

- Host genetics
- Host epigenetics
- Host gene expression
- Host immune profile
- Host clinical information
- Host demographics

(Multivariate covariates)

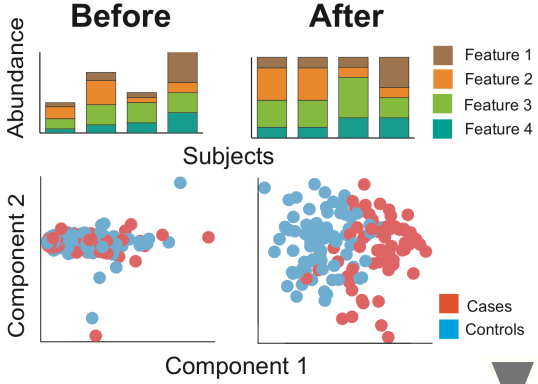
B Multi'omic Data Integration

Cases | Controls

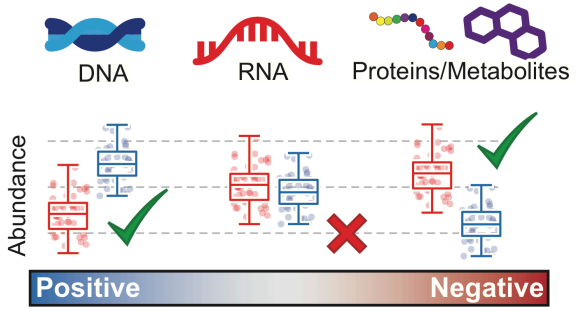
- Microbial species abundance
- Microbial strain profiles
- Microbial gene products
- Microbial gene expression profiles
- Microbial protein profiles
- Microbial metabolite concentrations
- Multivariate covariates

Integrated multi'omic profiles

C Normalization

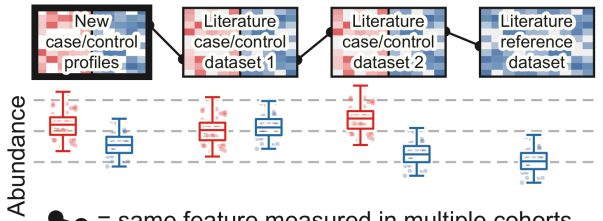


D Quantitative Analysis

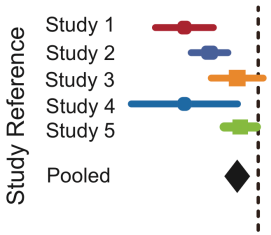


Disease Association

E Meta Analysis of Multi'omic Datasets

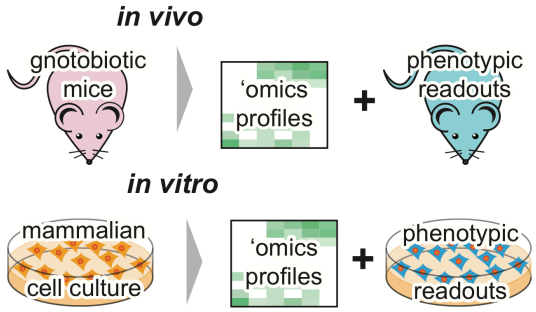


Synthesizing Evidence from Multiple Studies



Prioritized microbial features

F Experimental Validation

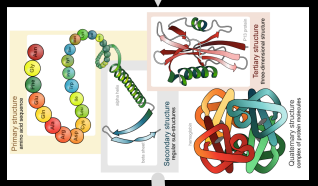


Typical microbial community analysis workflows

Sample

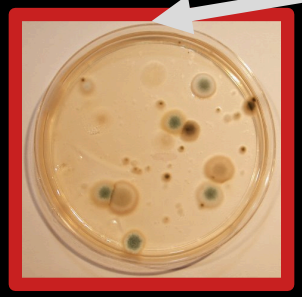
Nucleotides

Biomolecules



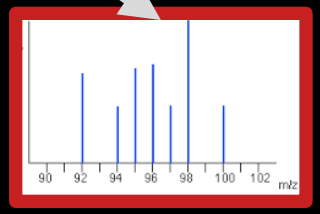
MS

Cells



Culture

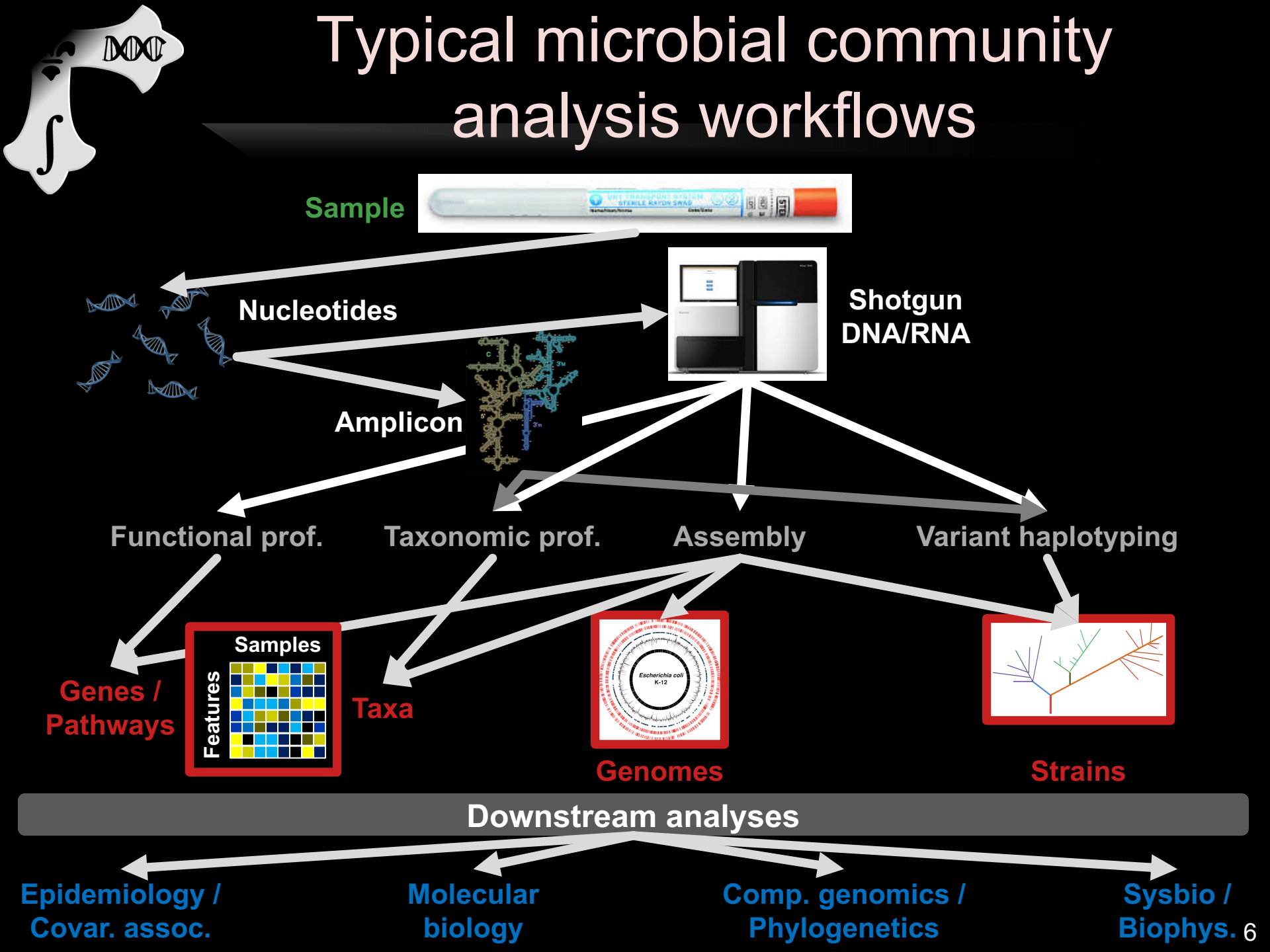
Microscopy



Metabolite / peptide quantification

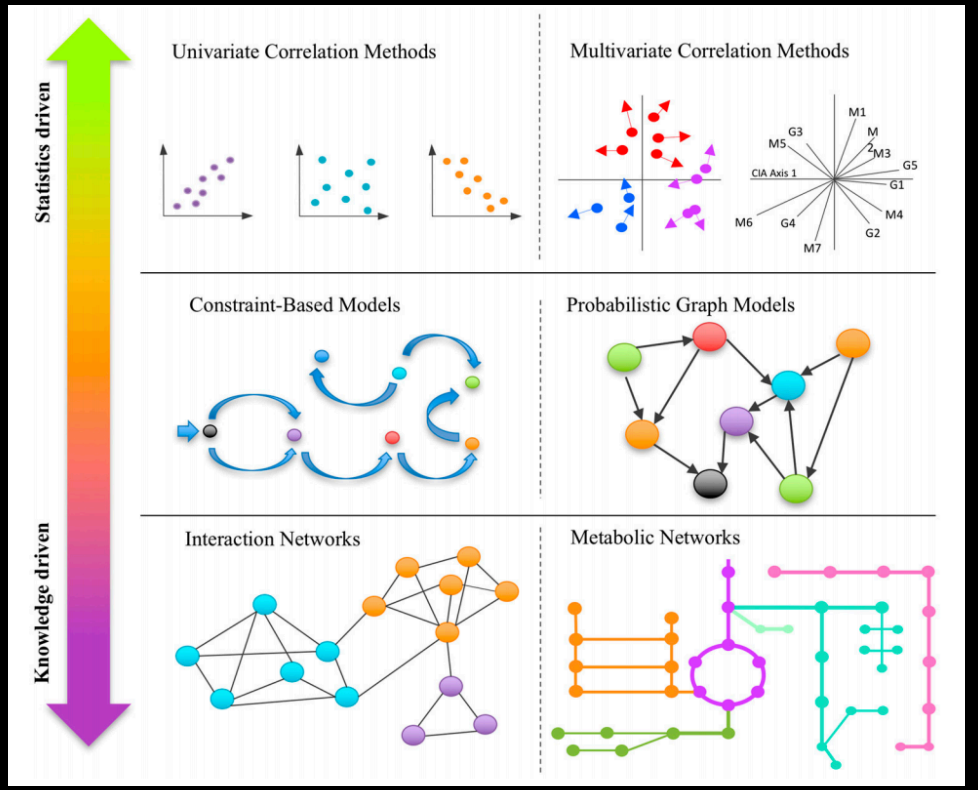
Sorting

Typical microbial community analysis workflows



Multi'omic analysis methods for microbial communities

- Four broad classes of approach:
  - Simple correlations (surprisingly effective!)
  - Factor analyses
    - Joint ordinations, biplots, canonical correlations.
  - Biological prior knowledge
    - Integrate on individual features.
  - Network models
    - Bugs-to-bugs, constraint-based.

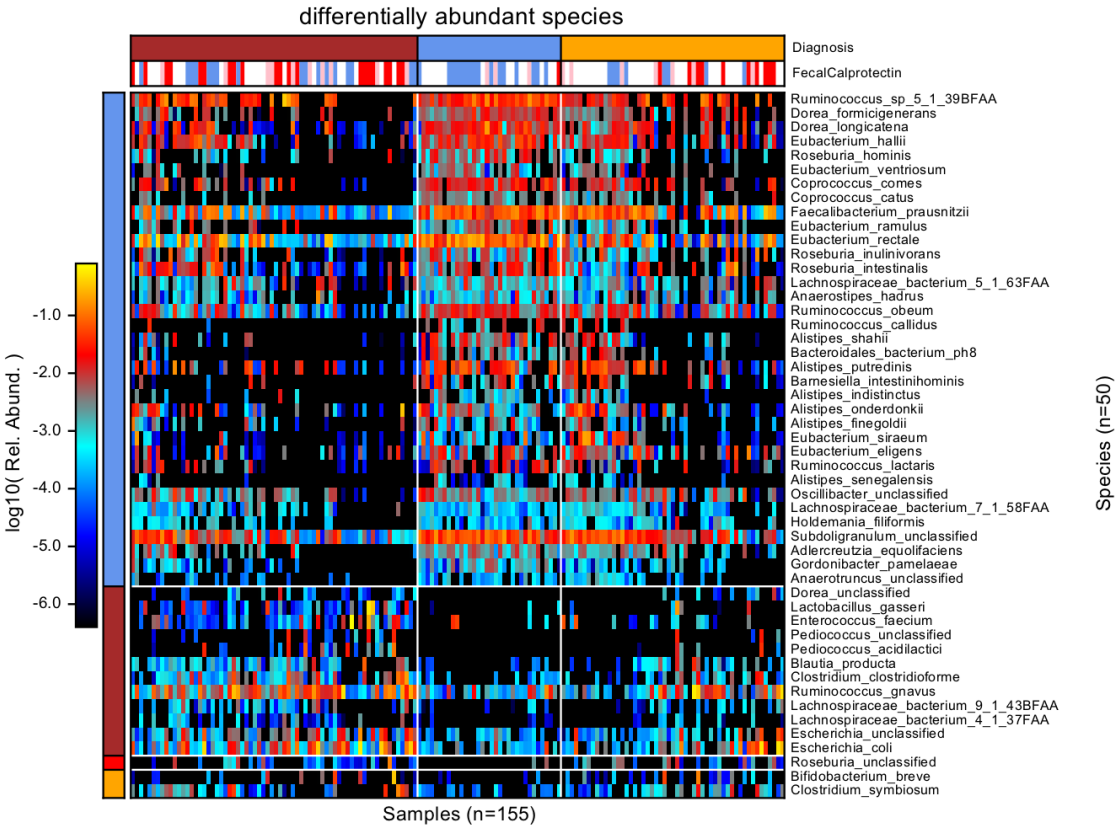


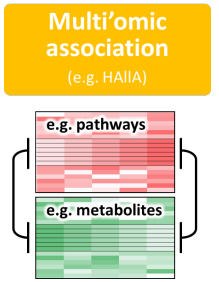
Chong and Xia Metabolites 2017

Integration by pairwise association

- Correlate all-against-all pairs of disparate ‘omic features between two datasets.
  - Usually something nonparametric (e.g. Spearman).
  - Sometimes after regressing out covariates (e.g. medication, repeated measures).
  - Matches on samples.
- Pros:
  - Simple!
  - Interpretable.
- Cons:
  - Difficult to power (lots of hypothesis tests).
  - Not particularly mechanistic.

Multi'omic association (IBD example)



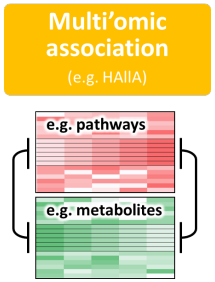


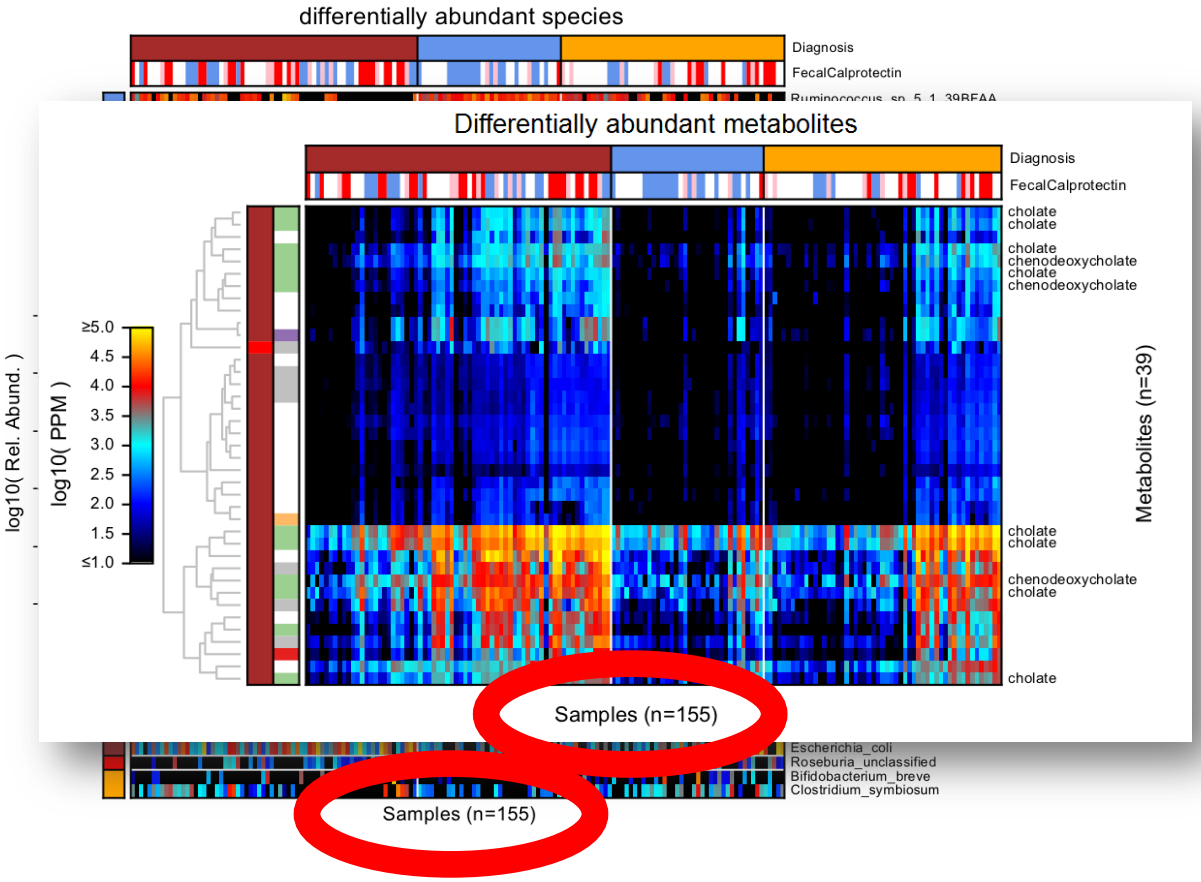
Metagenomic profiles

(here, species)

Diagnosis
- CD
- Control
- UC

Multi'omic association (IBD example)





Metagenomic profiles

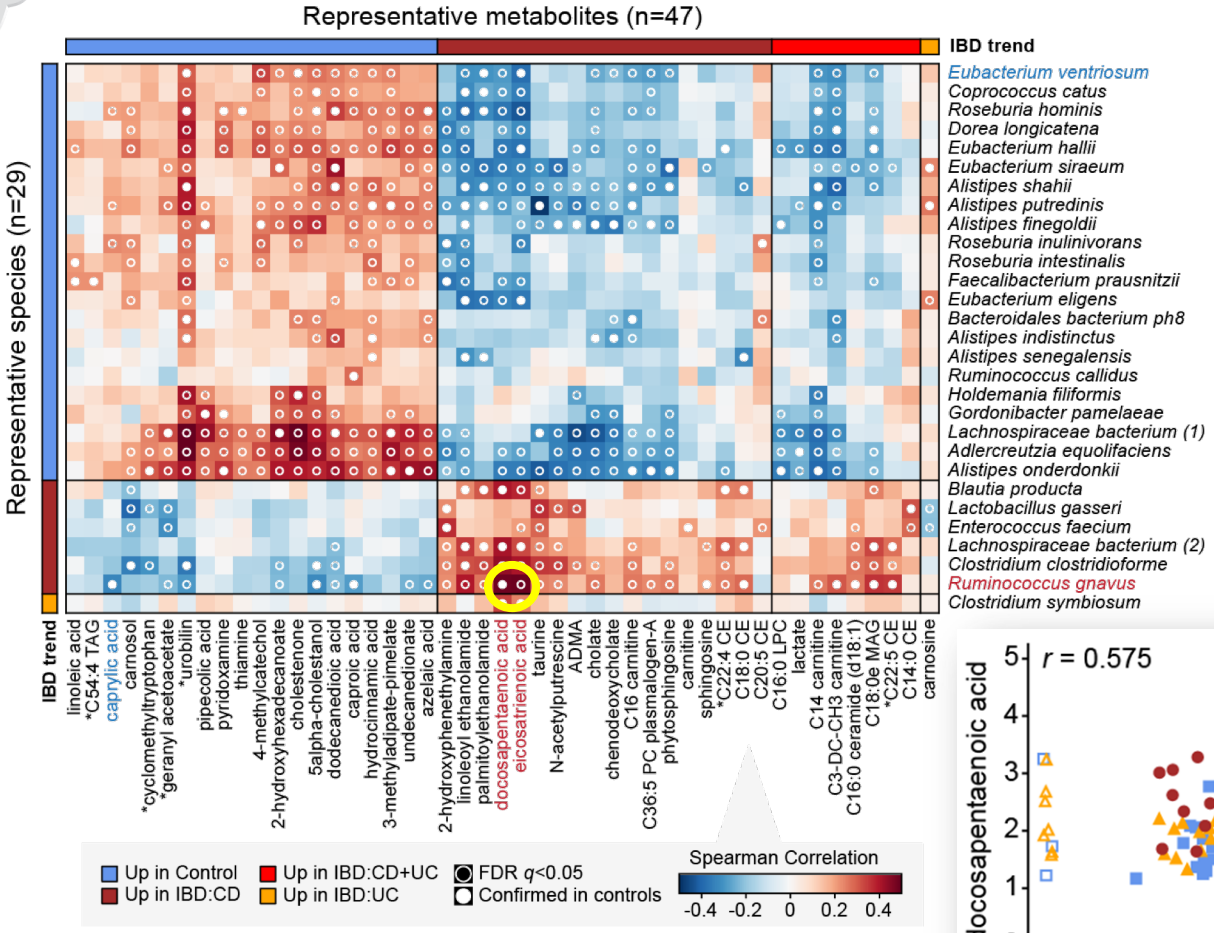
(here, species)

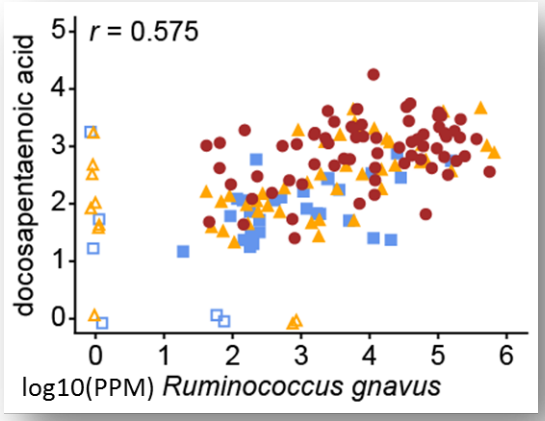
Other high-throughput profiles of the <u>same biosamples</u>

(here, untargeted metabolomics)

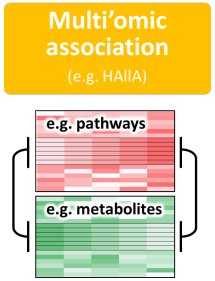
Diagnosis
- CD
- Control
- UC

Multi'omic association (IBD example)





Diagnosis
CD
Control
UC



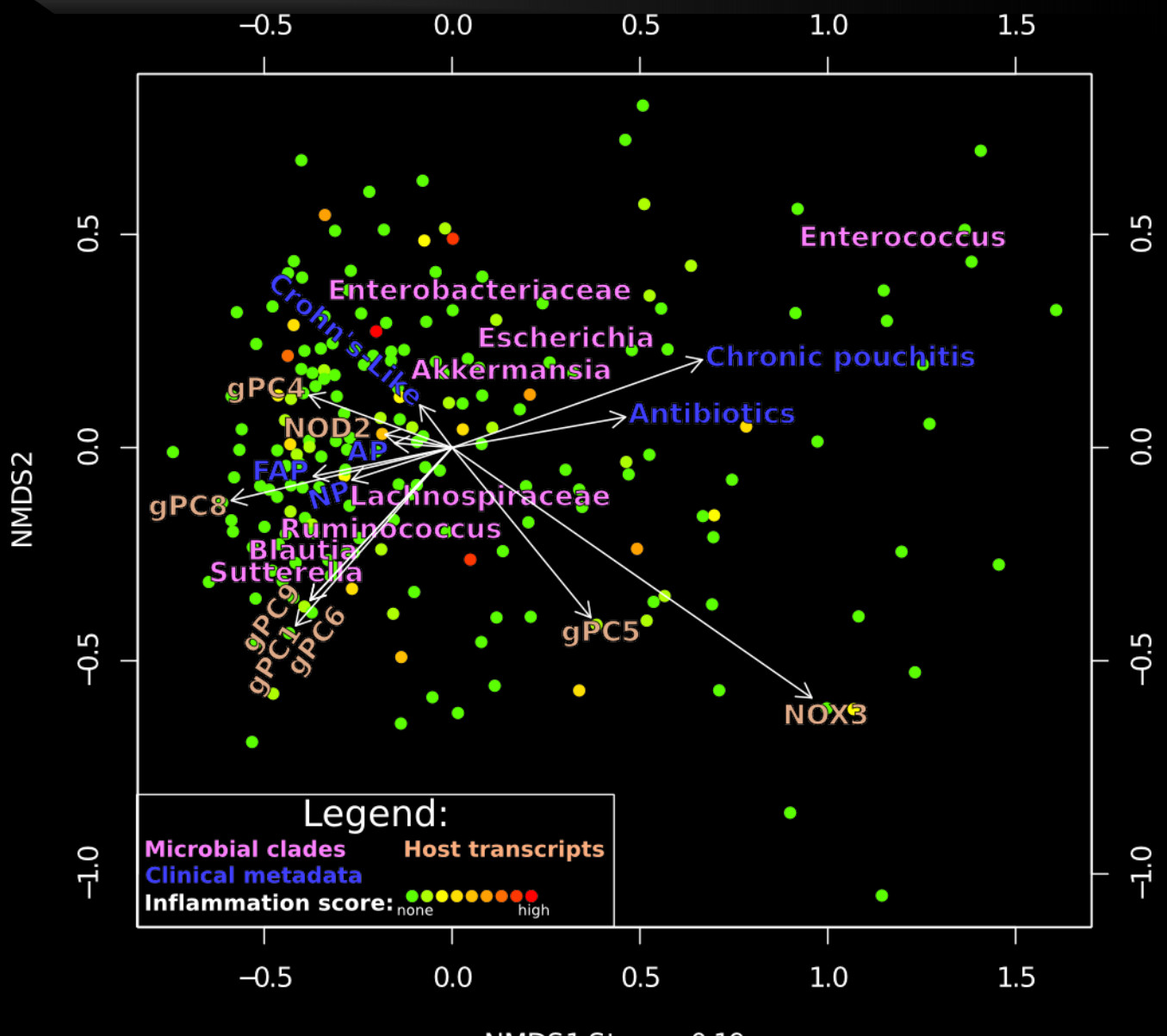
Simple correlation analysis: Identify associations between feature types

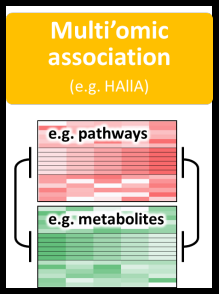
Many potential associations → multiple hypothesis testing

Integration by factor analysis

- Find shared patterns of overall variation between two or more datasets.
  - Some methods anchor on one dataset and project others into the same space (e.g. biplots).
  - Others weight two or more whole datasets and project them together (e.g. Mantel tests, canonical correlation analysis [CCA]).
  - Some can be supervised (e.g. LDA, PLS-DA); dangerous…
  - Matches on samples.
- Pros:
  - Uses “all the data.”
  - Effectively low dimensional and thus well-powered.
- Cons:
  - Finicky and numerically unstable.
  - Difficult to interpret unless you’re really lucky.

Host transcripts and the pouchitis microbiome





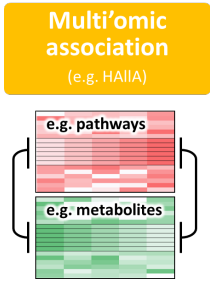
Multi'omics in the IBD gut microbiome

Between individuals

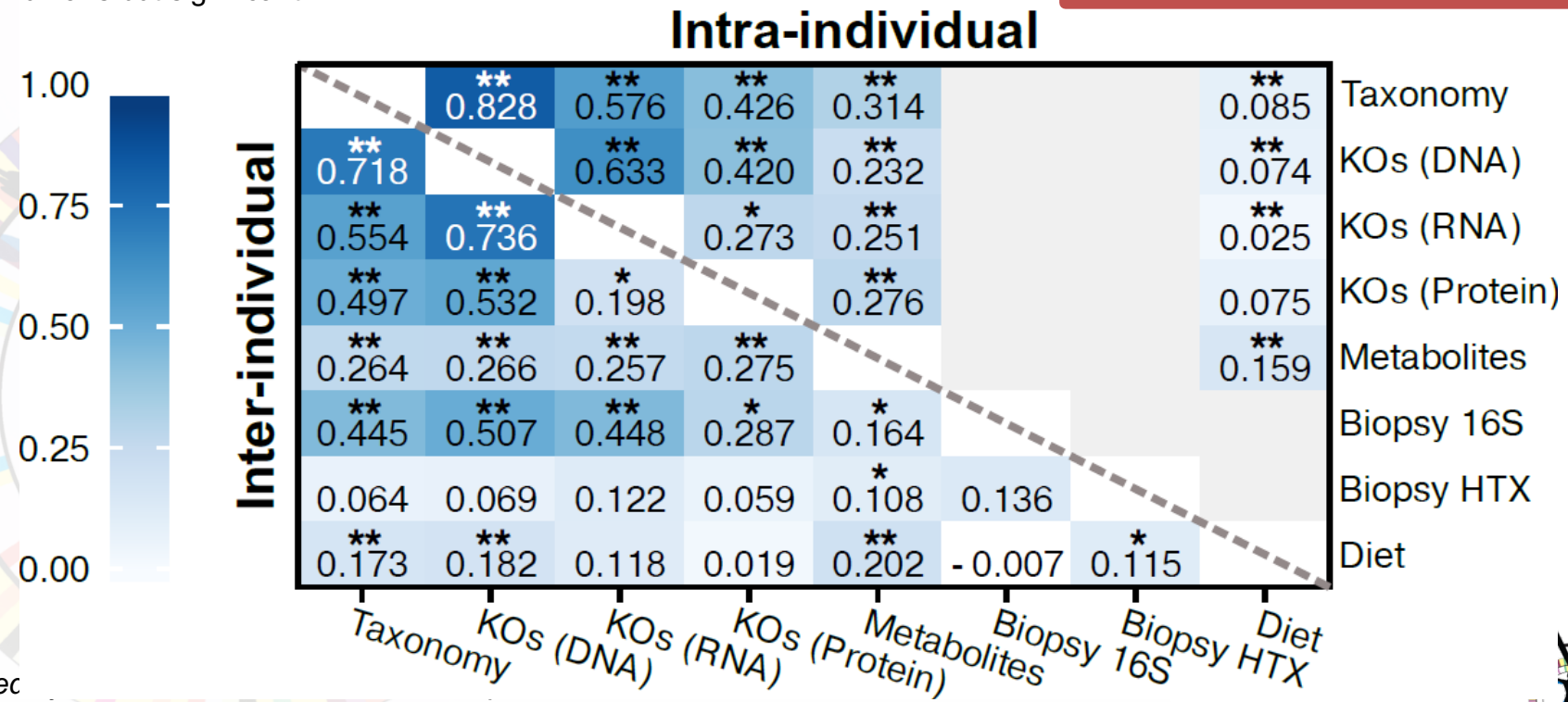
- Function reflects taxonomy, but they're not the same.
- Transcription reflects function, but they're not the same.
- Metabolites integrate information from all, including diet.
- Diet effects are modest overall in humans but significant.

Within individuals

- Similar to between-subjects, but more so.
- Transcripts in particular respond to perturbations beyond mgx function.
- Diet effects very modest.



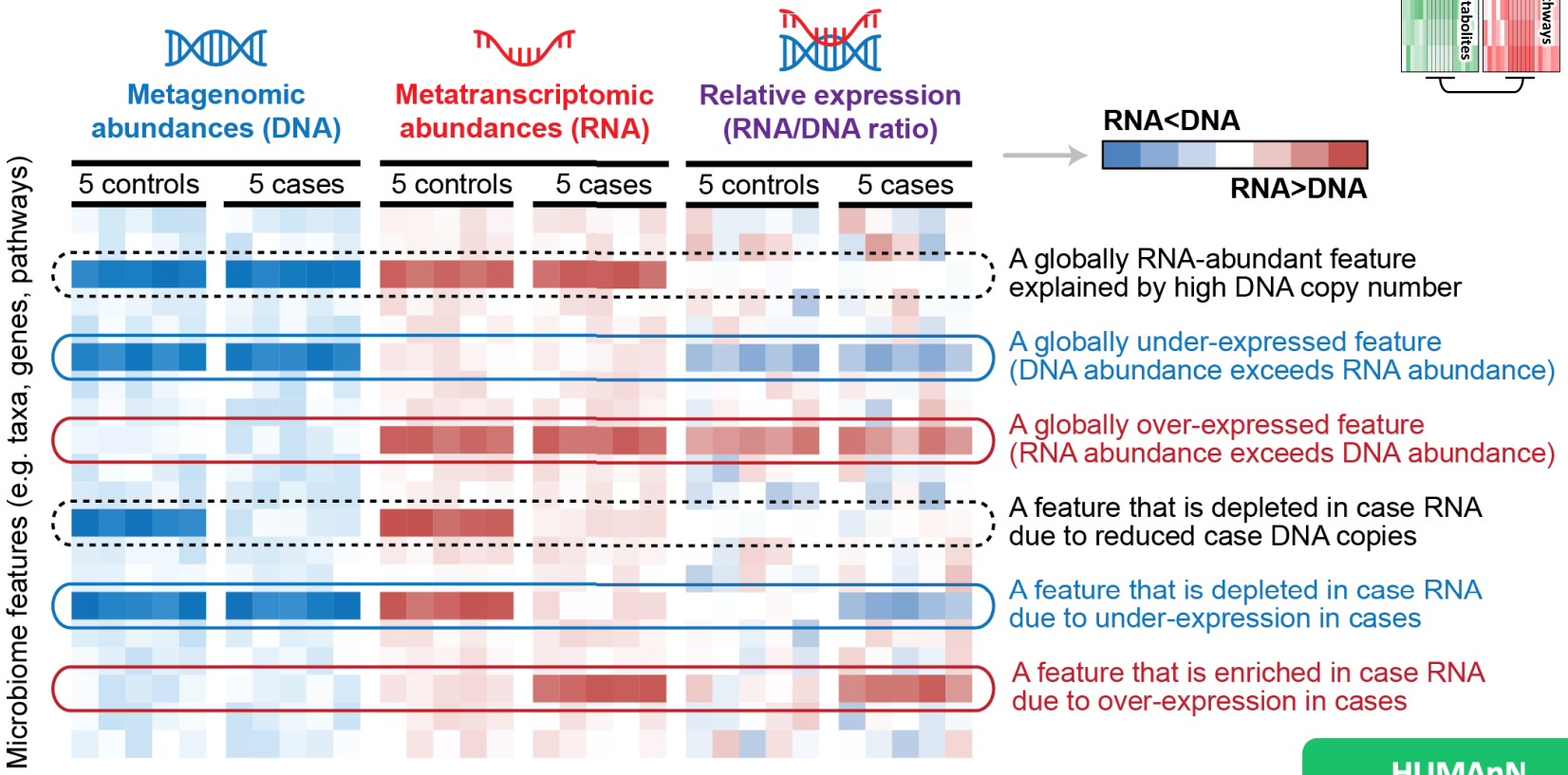
Mantel test: correlation between (dis-)similarity matrices.

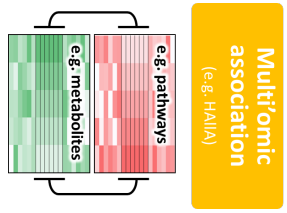


Integration using prior knowledge

- Measure the “same” features multiple ways and compare the results.
  - E.g. genes in DNA, RNA, and proteins.
  - Can be correlated or integrated into quantitative models (e.g. graphical models or ODEs).
  - Matches on features.
- Pros:
  - Very mechanistic and interpretable.
- Cons:
  - Requires the right measurement types.
  - Only works for known / identifiable features.

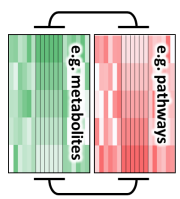
From functional potential to activity

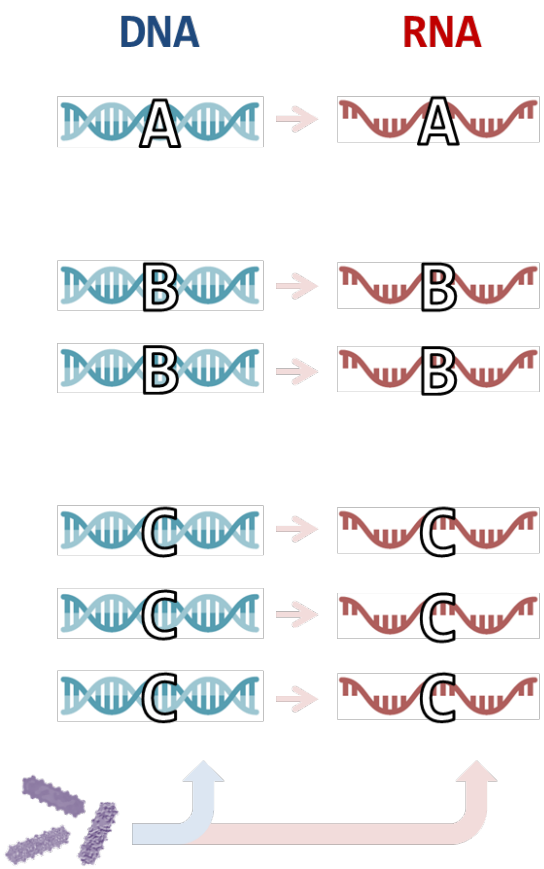




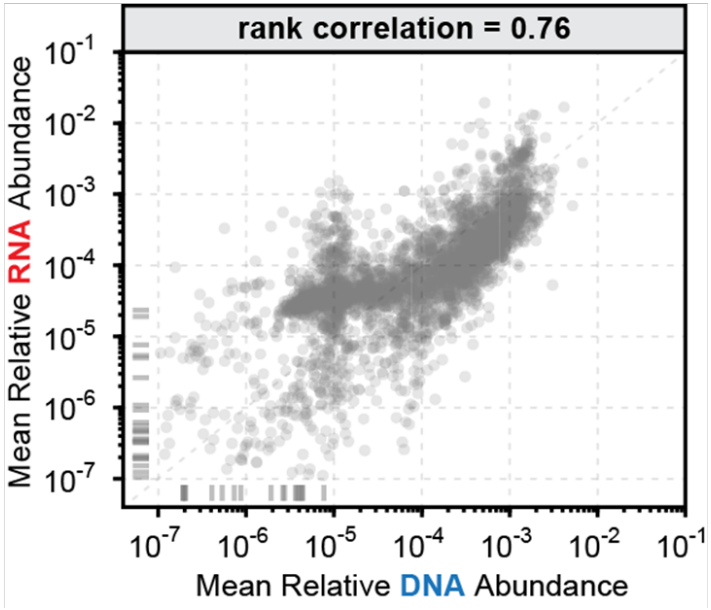
HUMAnN
Profile microbial genes and pathways; stratify by contributing species

From functional potential to activity

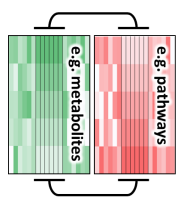


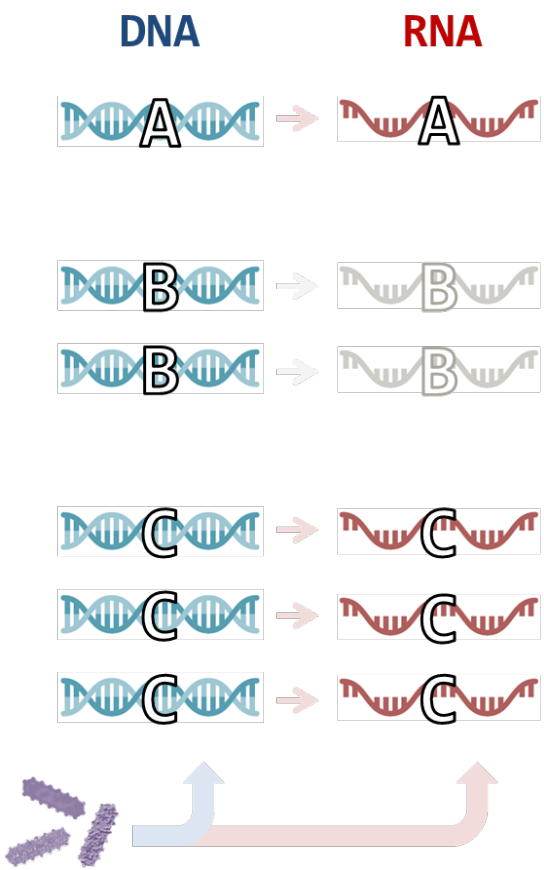


Bug DNA and RNA are actually well correlated in the gut!



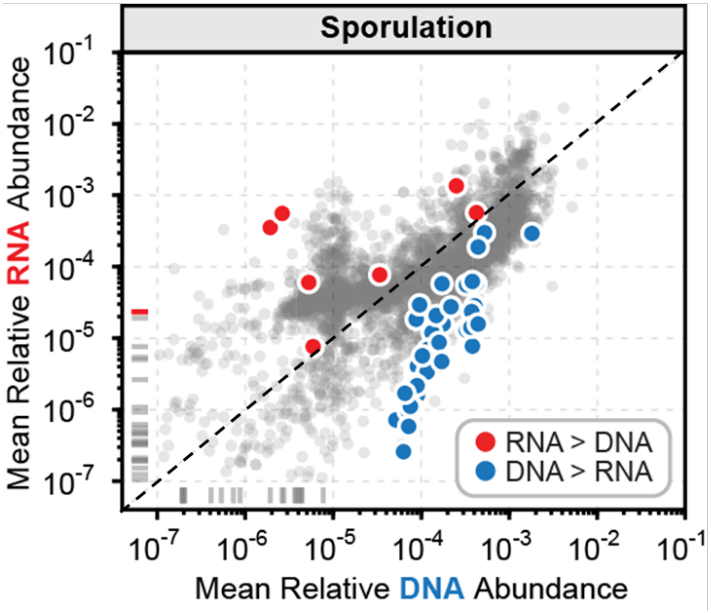
From functional potential to activity



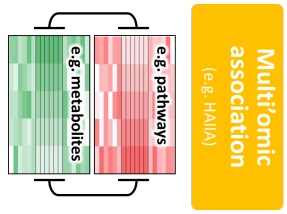


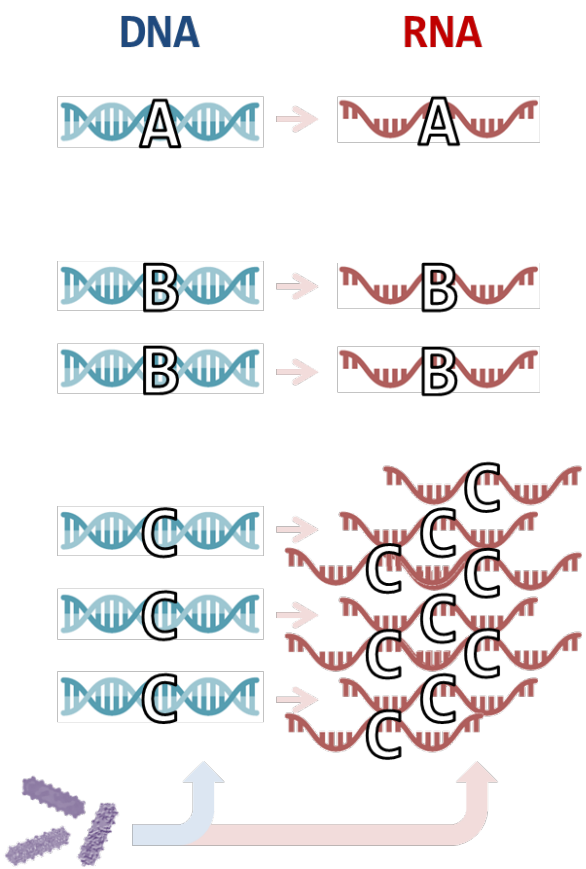
Some functions are **turned off**

e.g. stress responses

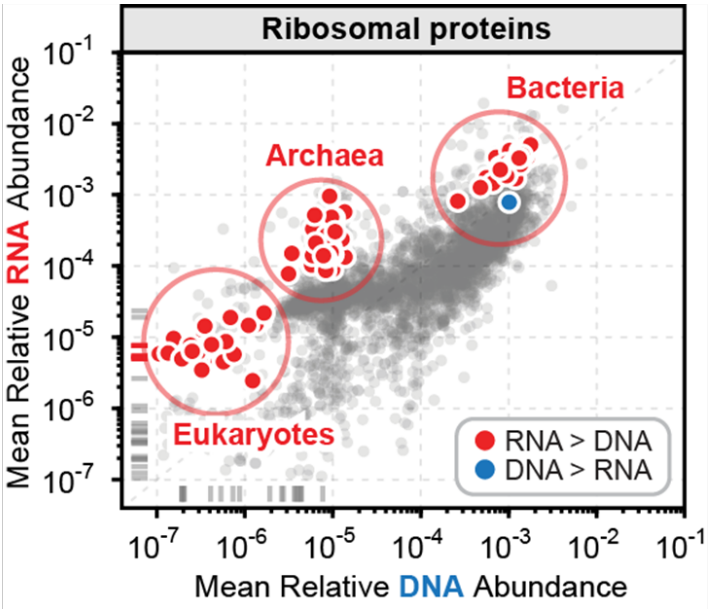


From functional potential to activity





Other functions are
up-regulated
e.g. making ribosomes



Integration using network models

- Construct and compare interaction networks within and between data types.
  - E.g. bugs that should interact based on metabolites.
  - E.g. bugs that co-occur spatially, temporally, or ecologically.
  - Can match on features *or* samples *or* neither.
- Pros:
  - Very detailed, scalable, and generalizable.
- Cons:
  - Requires a lot of detailed, low-noise measurements.
  - And a lot of work!

Human Microbiome Project, Phase One: http://commonfund.nih.gov/hmp

Phase 1 (2007-2012): Survey human microbial variation

Healthy cohort

"Who's there?"

Demonstration projects

Human Microbiome Project, Phase Two: http://ihmpdcc.org

Phase 2 (2013-2018): Integrative HMP "iHMP"

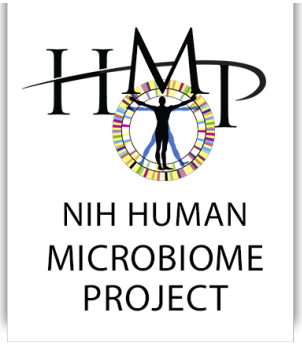
"What are they doing?"

Analyze biological properties of both *microbiome* & *host over time* to understand biomarkers and mechanisms of health and disease.

IBD
Broad Inst.

Pregnancy and Preterm Birth
VCU

Type 2 Diabetes
Stanford/Jax

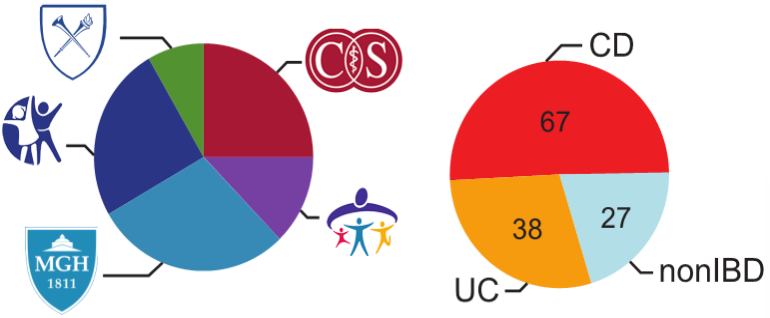


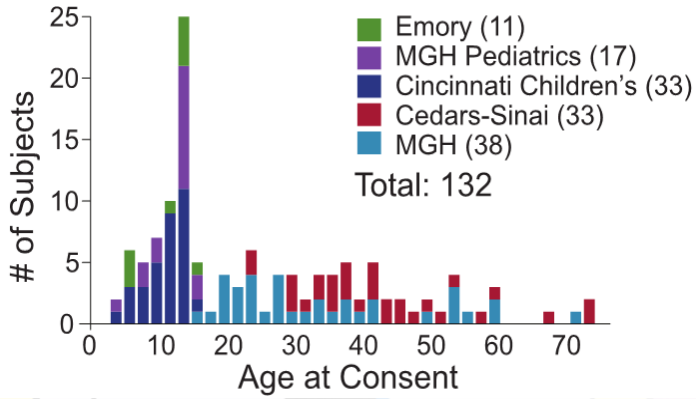
Data Coordination
UMD IGS

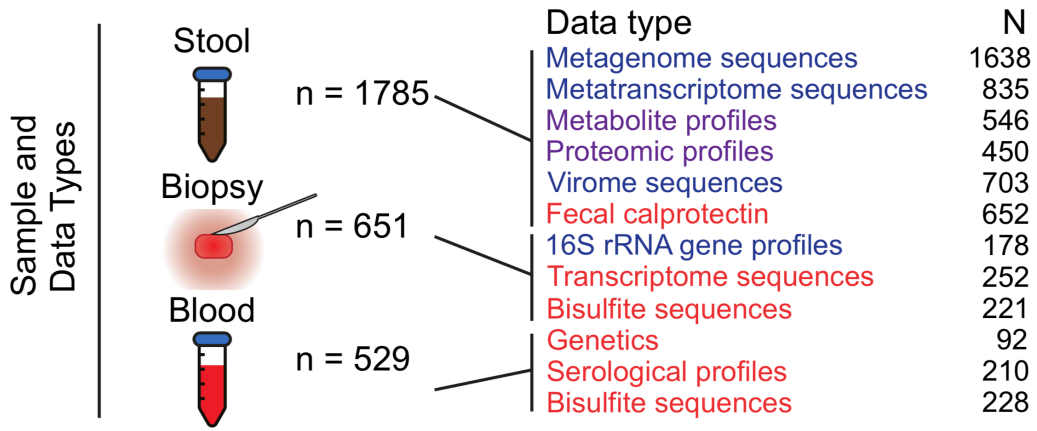
The “HMP2” IBD Multi’omics Data resource

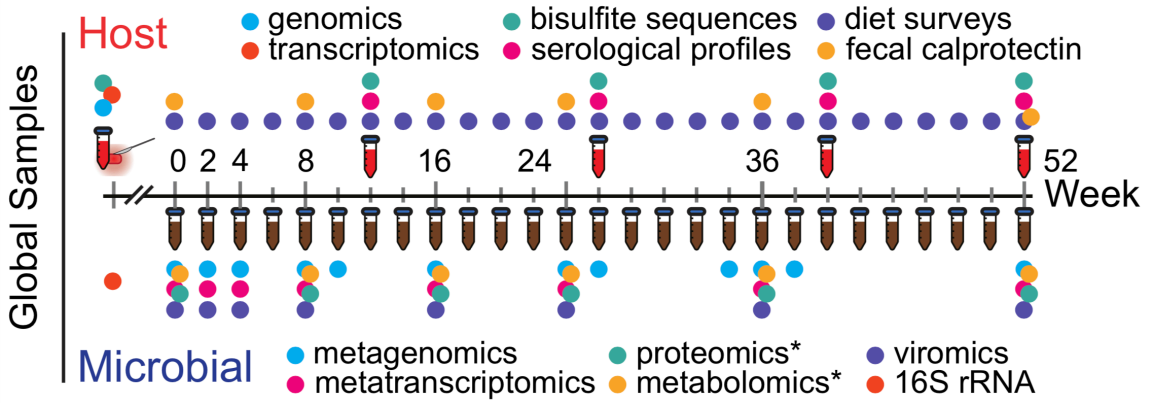
http://ibdmdb.org

With Ramnik Xavier









Jason Lloyd-Price

Funded by National Institutes of Health, Dept. of Health and Human Services

The IBD Multi'omics DataBase

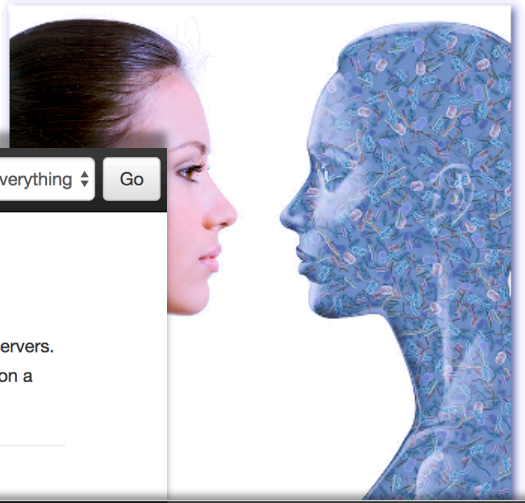
http://ibdmdb.org

Cesar Arze

Home | Download Data | Protocols | Team | Explore Data | Participant Interface | Search | Everything | Go

The Inflammatory Bowel Disease Multi'omics Database

The cells of the human body are outnumbered ten to one by bacteria, but large-scale surveys of the human microbiome were not feasible until the advent of next-generation sequencing. The first stage of the Human Microbiome Project sampled 300 healthy subjects to determine normal microbial composition of healthy Americans (which microbial species were there), their biochemical function (what the microbes were doing), and microbial variation both between individuals and over time. Now that



Home | Download Data | Protocols | Team | Explore Data | Participant Interface | Search | Everything | Go

Results

This page shows the high level results over all of the HMP2 pipelines. Each run is comprised of a set of data that has been uploaded to the HMP2 servers. Once there, it is filtered for quality and error checked for completeness and saved under the `raw files` page. After the QC phase, the data is run on a specific AnADAMA2 pipeline, producing several types of data products. Each data product for a project is saved on the `products`.

Download HMP2 Metadata | Download Provenance Log

Available Studies

Name	Data type
HMP2	Metagenomes
HMP2	Proteomics
HMP2	Viromics
HMP2	Metatranscriptomes

Home | Download Data | Protocols | Team | Participant Interface

(Home) » Clinical Protocols

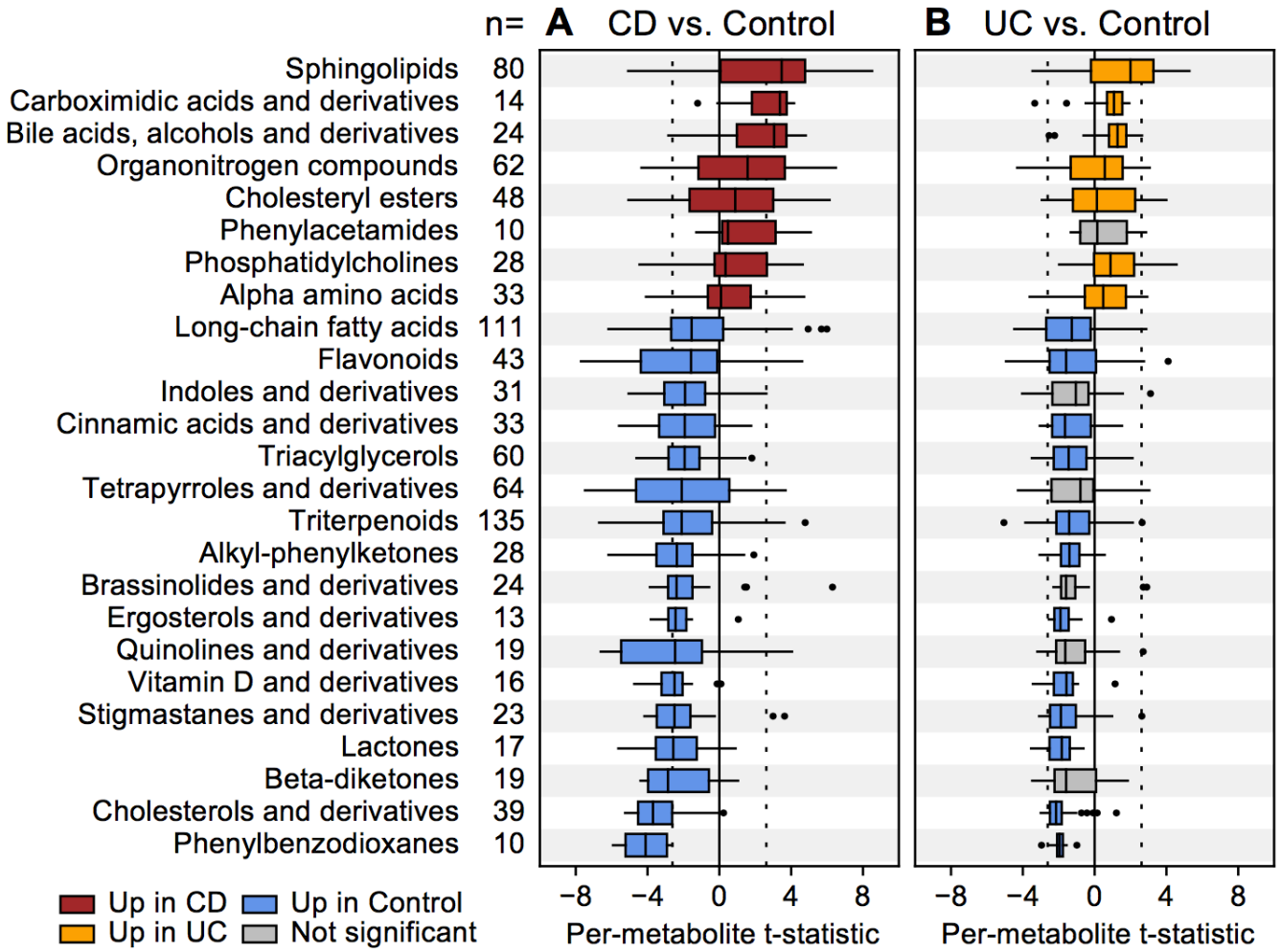
Containers

- *Clinical Protocols*
- Sample Handling Pro...
- Data Generation Pro...

Name	Content Type
Coordinator Data Forms	--
Coordinator Protocols	--
Patient Instructions	--

Funded by National Institutes of Health, Dept. of H

Metabolomic associations with metagenomic taxonomy and function

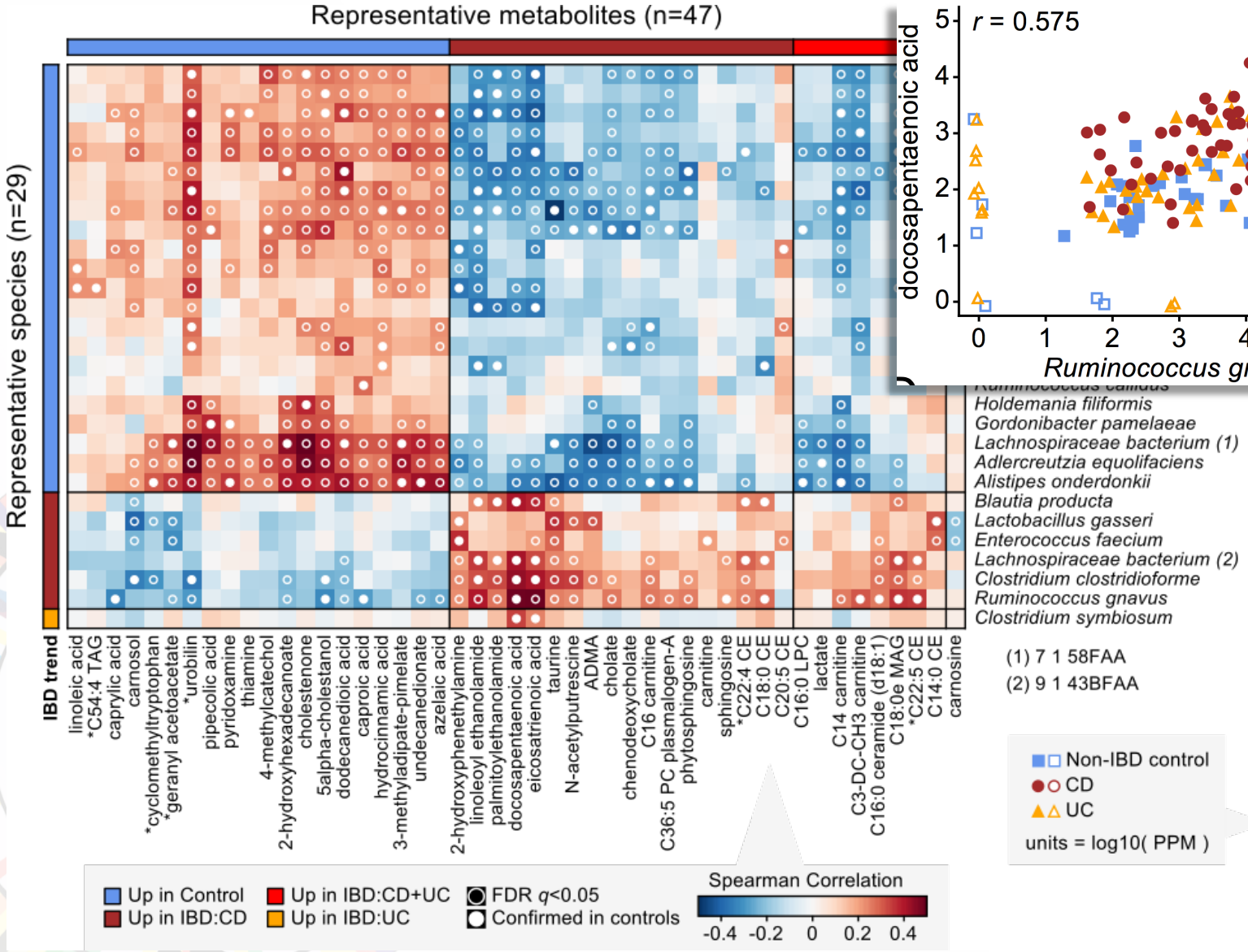


Hera Vlamakis

Eric Franzosa

Funded by National Institutes of Health, Dept. of Health and Human Services

Metabolomic associations with metagenomic taxonomy and function



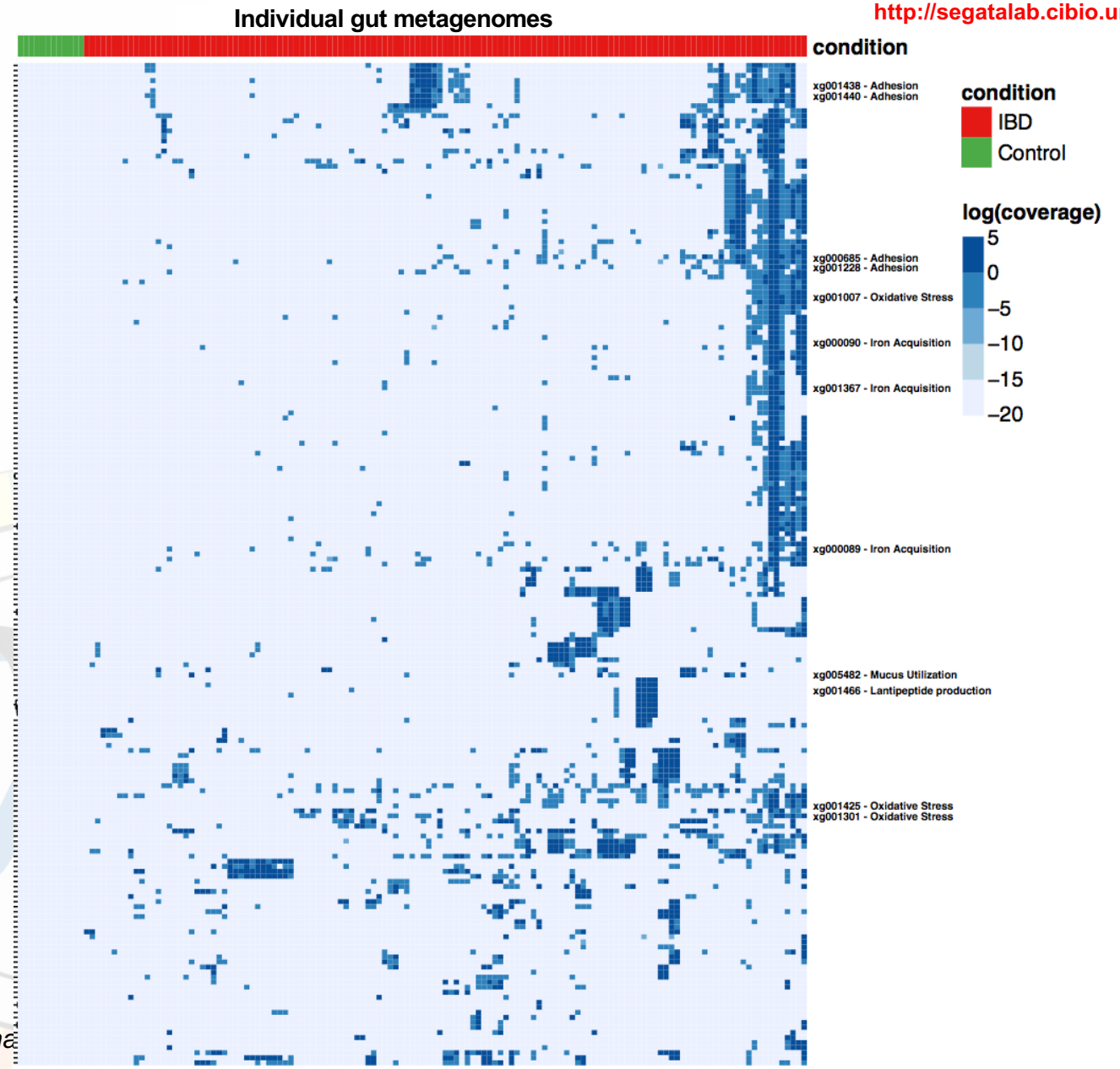
Hera Vlamakis

Eric Franzosa

Funded by National Institutes of Health, Dept. of Health and Human Services

Linking microbial function to strain-specific phenotypes in IBD

http://segatalab.cibio.unitn.it/tools/panphlan

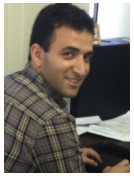


Brantley Hall

Moran Yassour

Funded by National

Linking microbial function to strain-specific phenotypes in IBD



Ali Rahnavard

Well-annotated

Homology-based

Novel / hypothetical

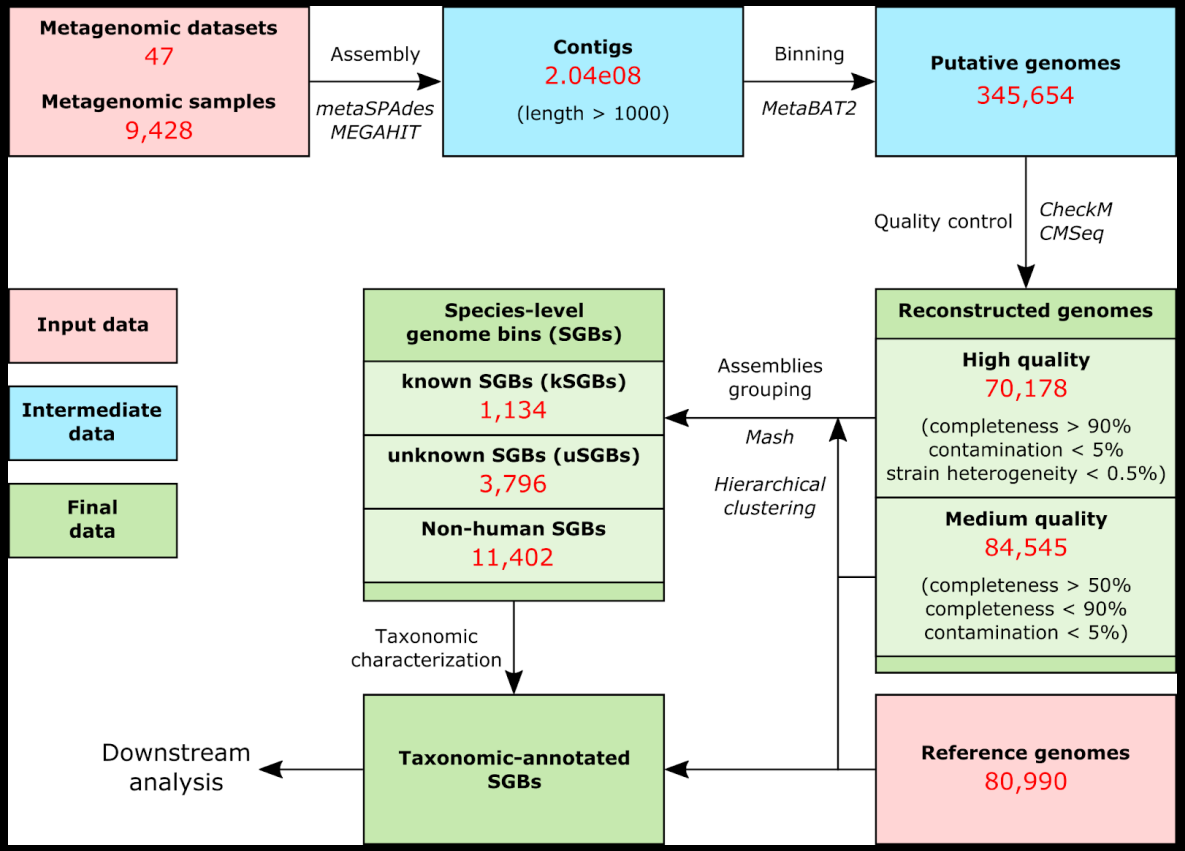
Prevalent, health-relevant functional novelty abounds in the microbiome and is strain-specific.

Brantley Hall

Moran Yassour

Funded by National

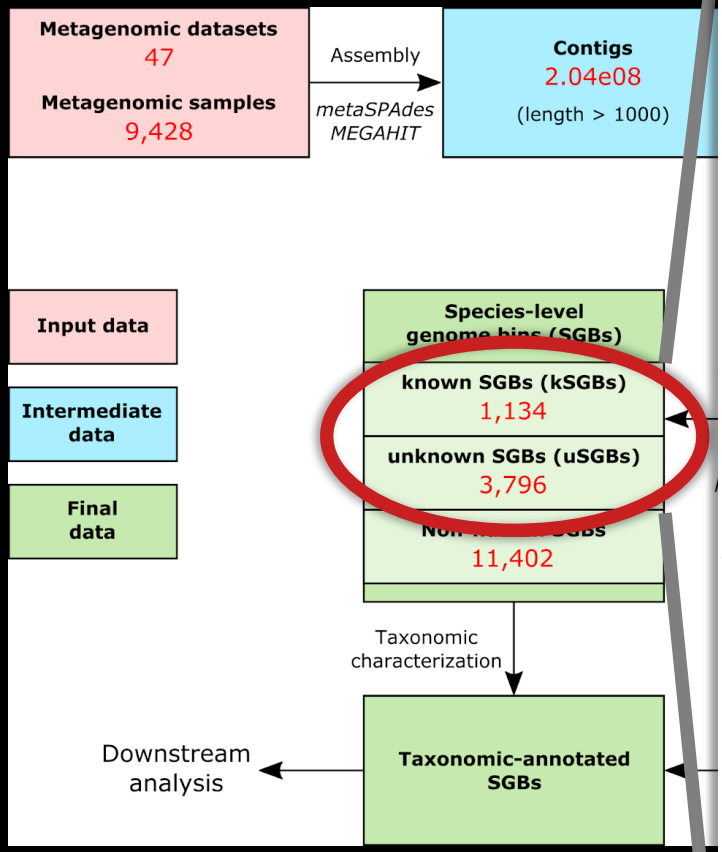
Strain-specific functional novelty abounds body- and world-wide, and is analytically accessible

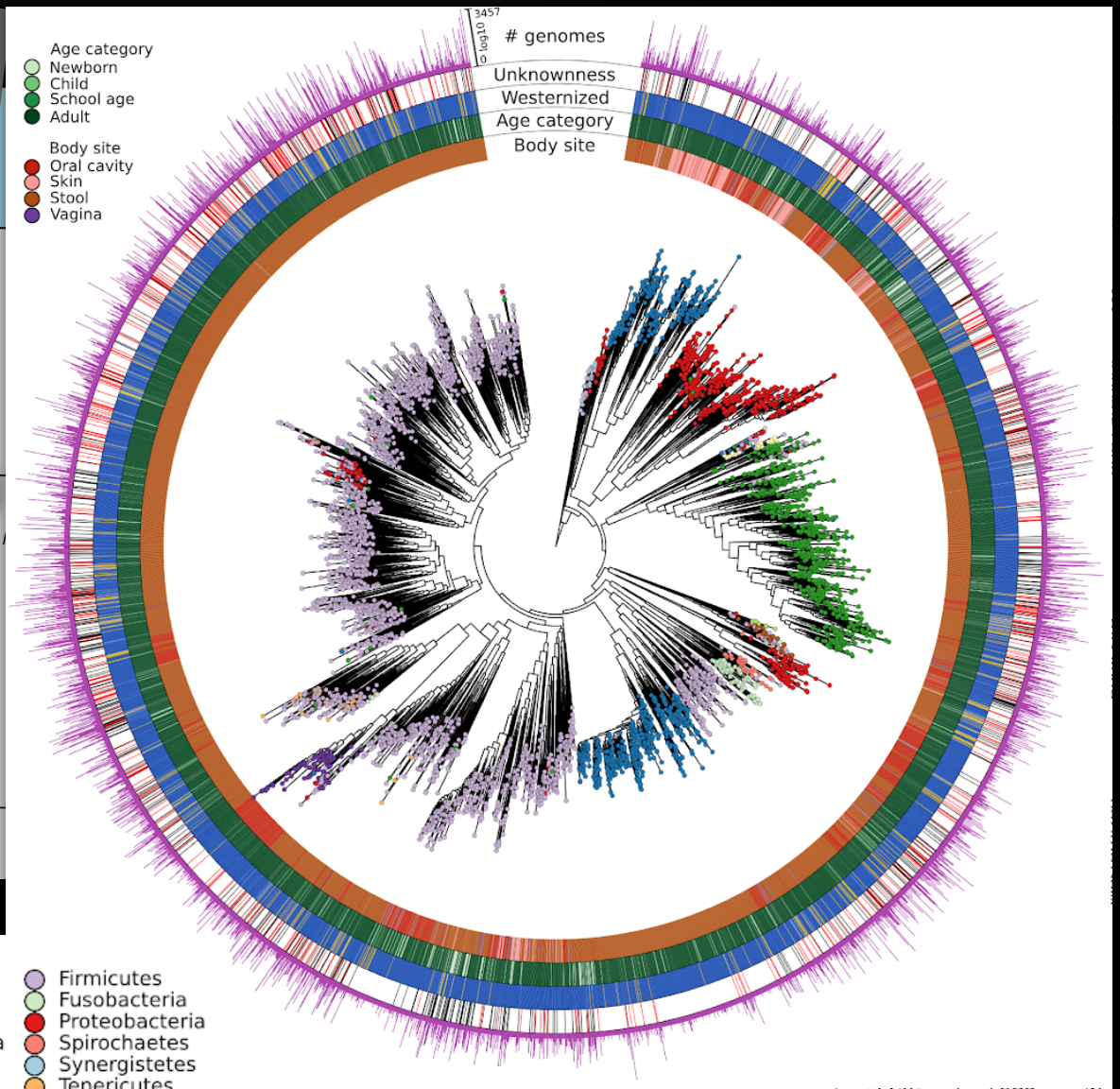


High-quality assemblies for ~150,000 genomes in ~5,000 species-level bins from ~10,000 metagenomes spanning ~50 datasets from around the world and across the human body.

Nicola Segata Edo Pasolli

Strain-specific functional novelty abounds body- and world-wide, and is analytically accessible



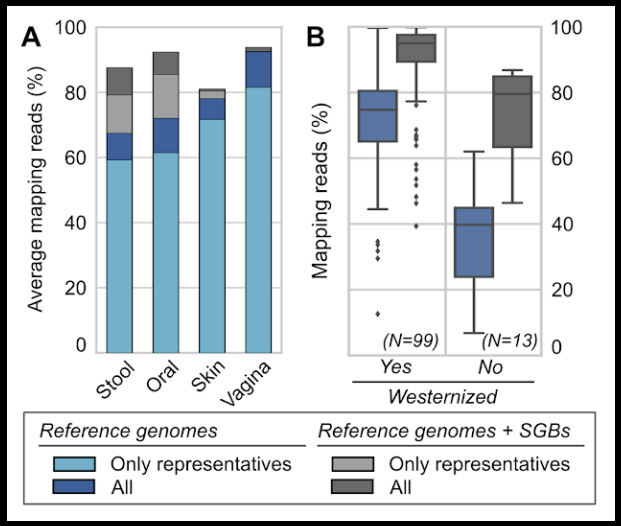


Nicola Segata

Edo Pasolli

Phylum
- Actinobacteria
- Bacteroidetes
- Cand. Melainabacteria
- Cand. Saccharibacteria
- Chlamydiae
- Elusimicrobia
- Euryarchaeota
- Firmicutes
- Fusobacteria
- Proteobacteria
- Spirochaetes
- Synergistetes
- Tenericutes
- Verrucomicrobia
- Others

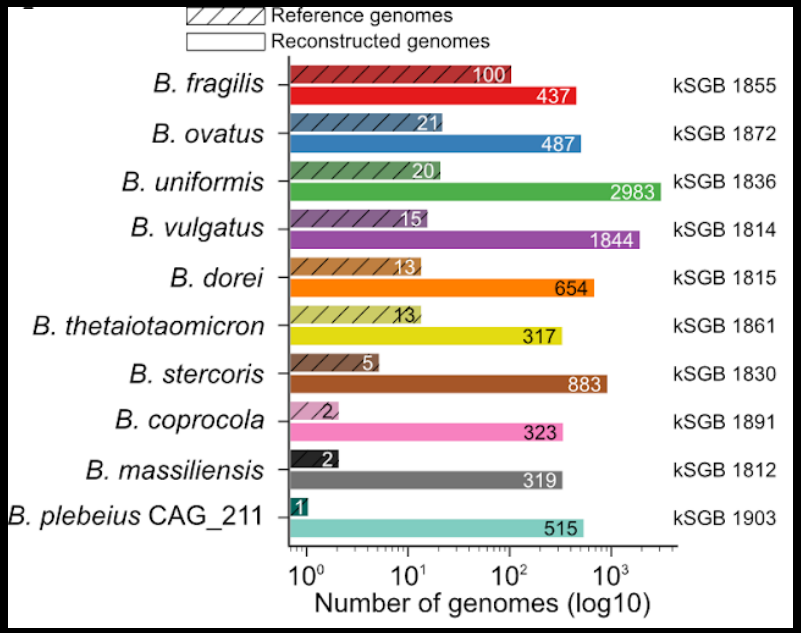
Strain-specific functional novelty abounds body- and world-wide, and is analytically accessible

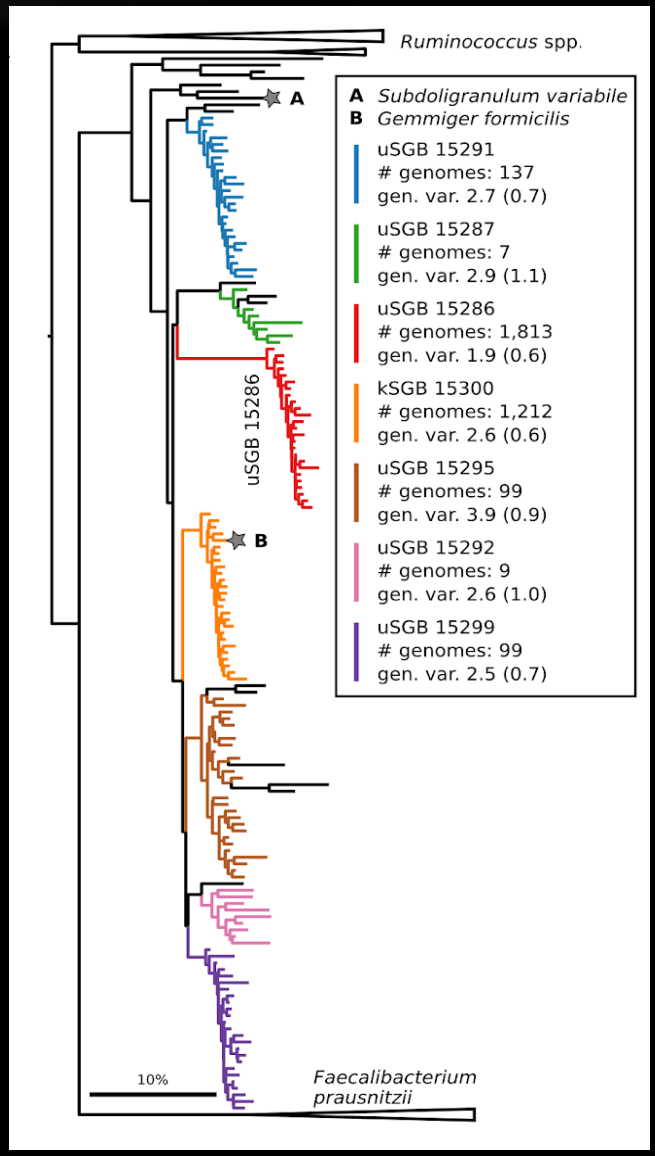


Most novelty is from undercharacterized populations...

...but some is prevalent...

...and most is functional.



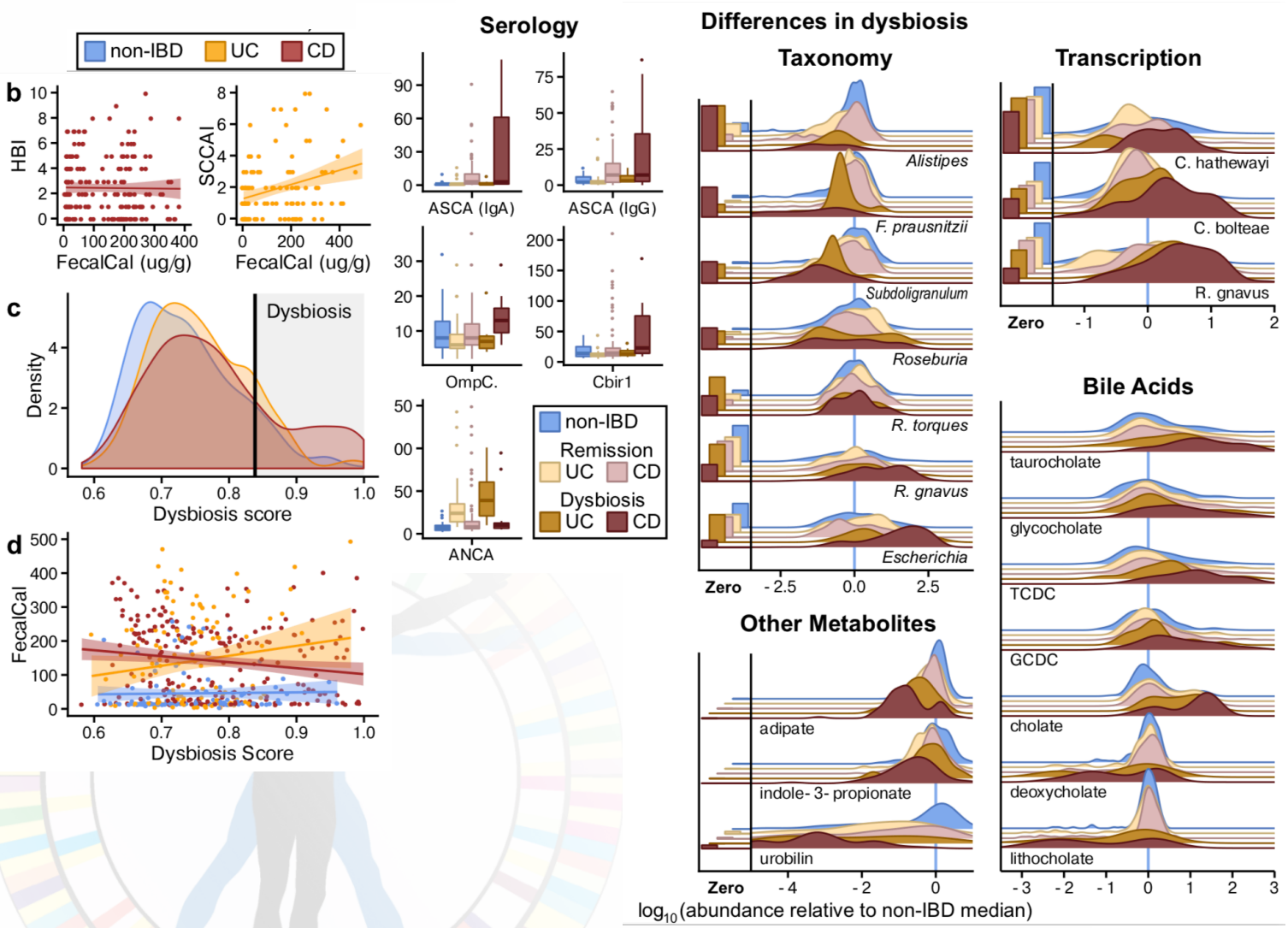


Nicola Segata

Edo Pasolli

Other findings from the HMP2

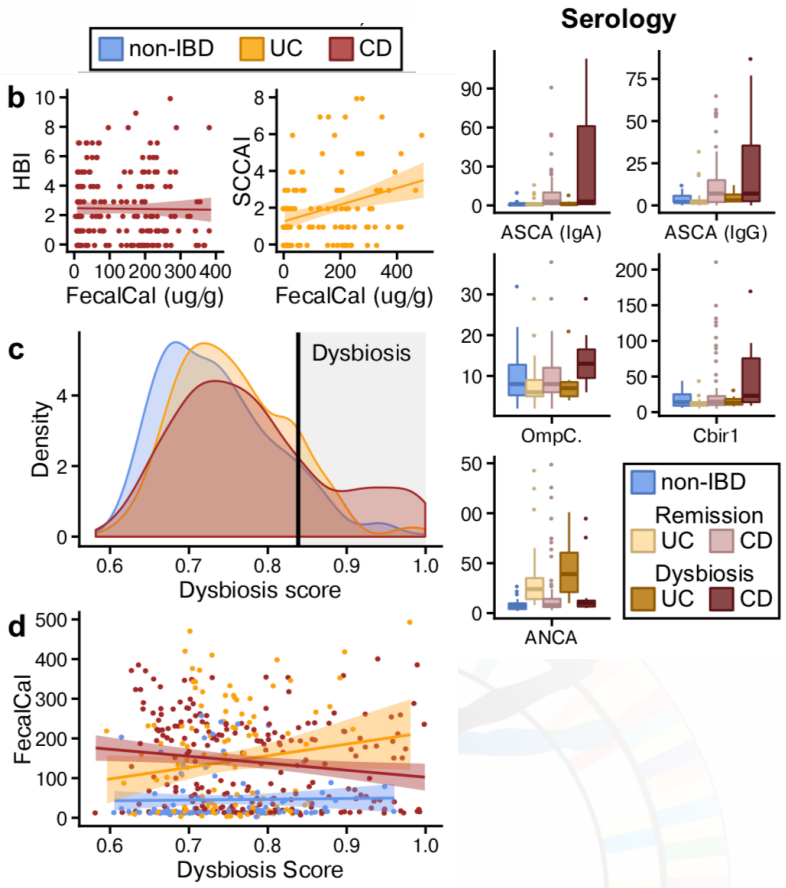
Jason Lloyd-Price

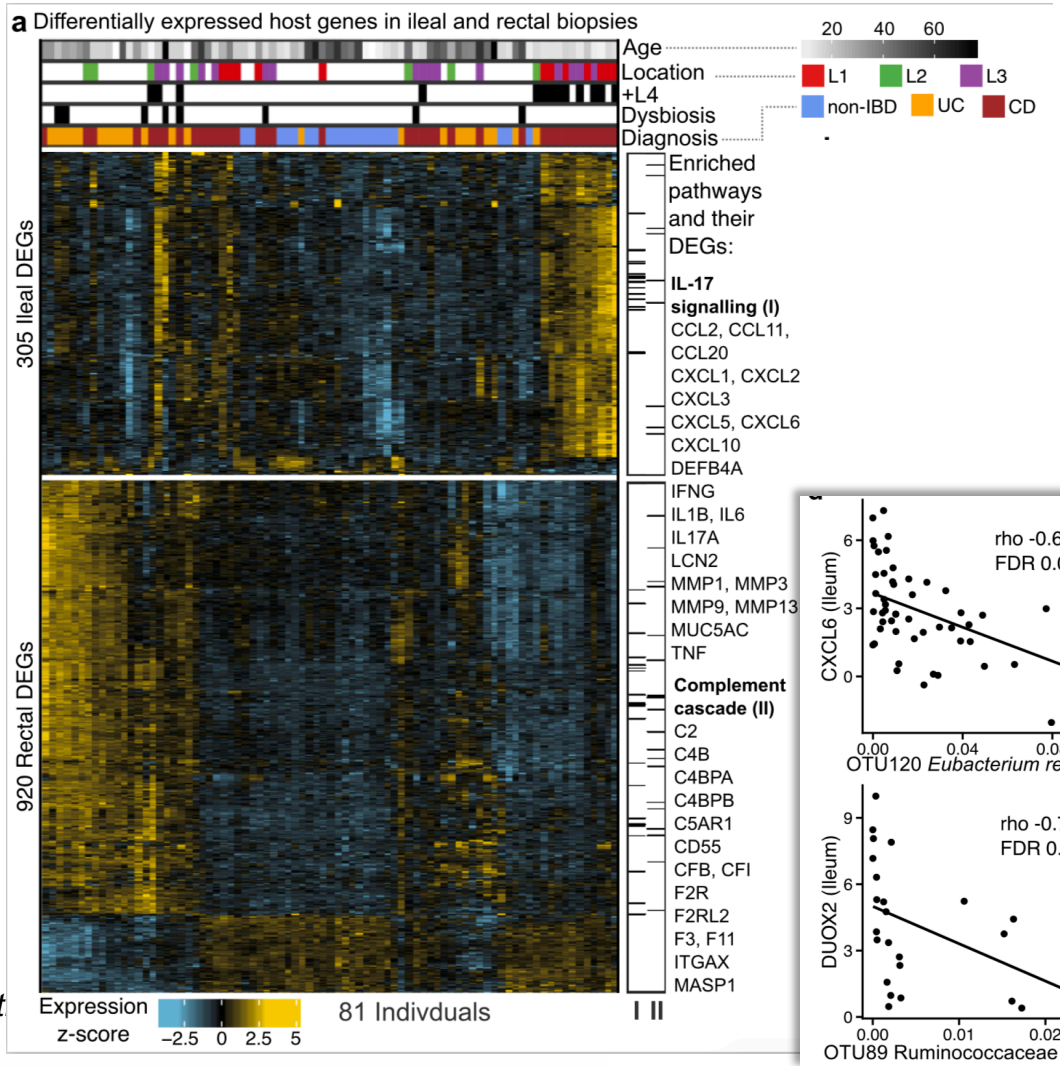


Funded by National Institutes of Health, Dept. of Health and Human Services

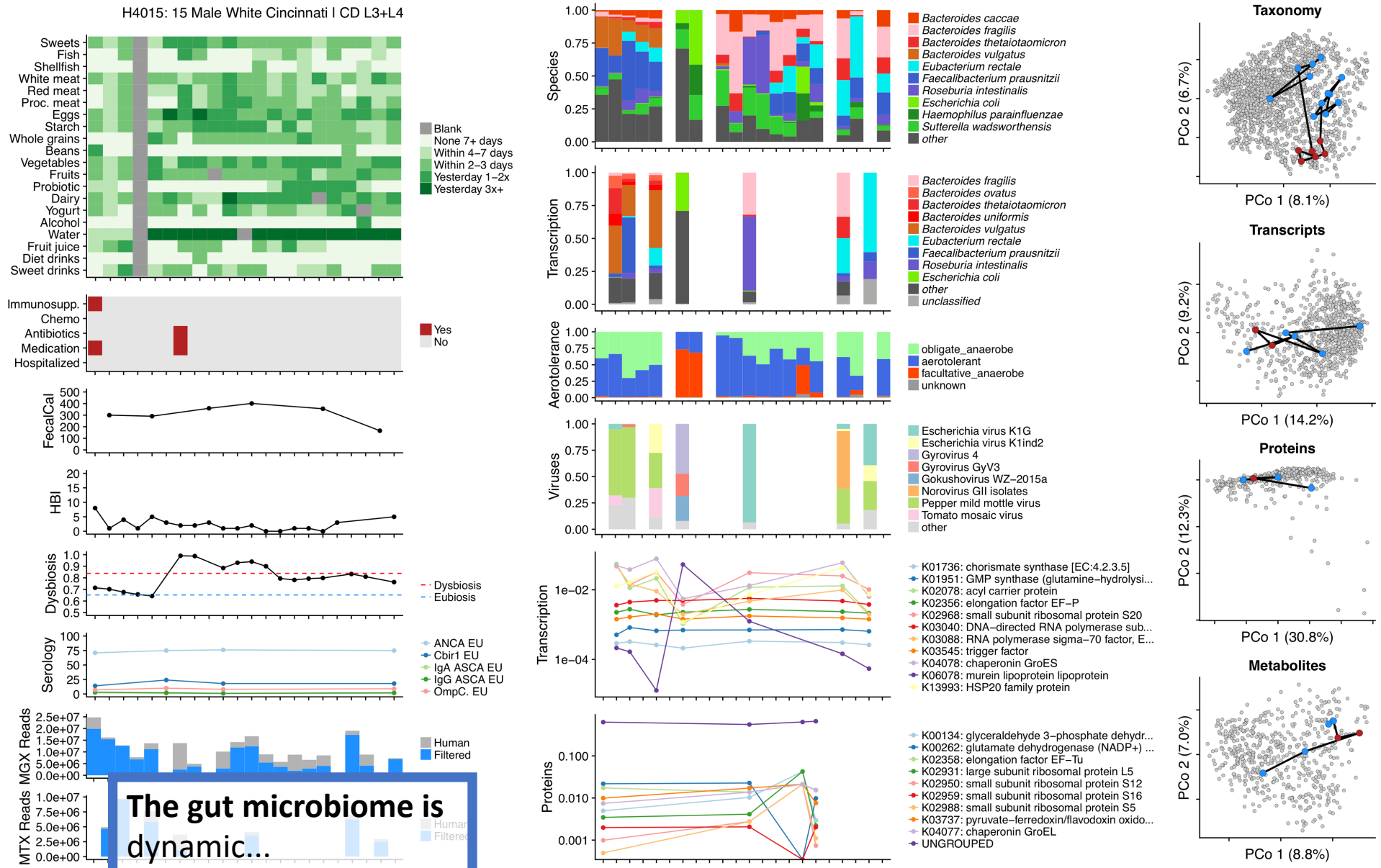
Other findings from the HMP2

Jason Lloyd-Price

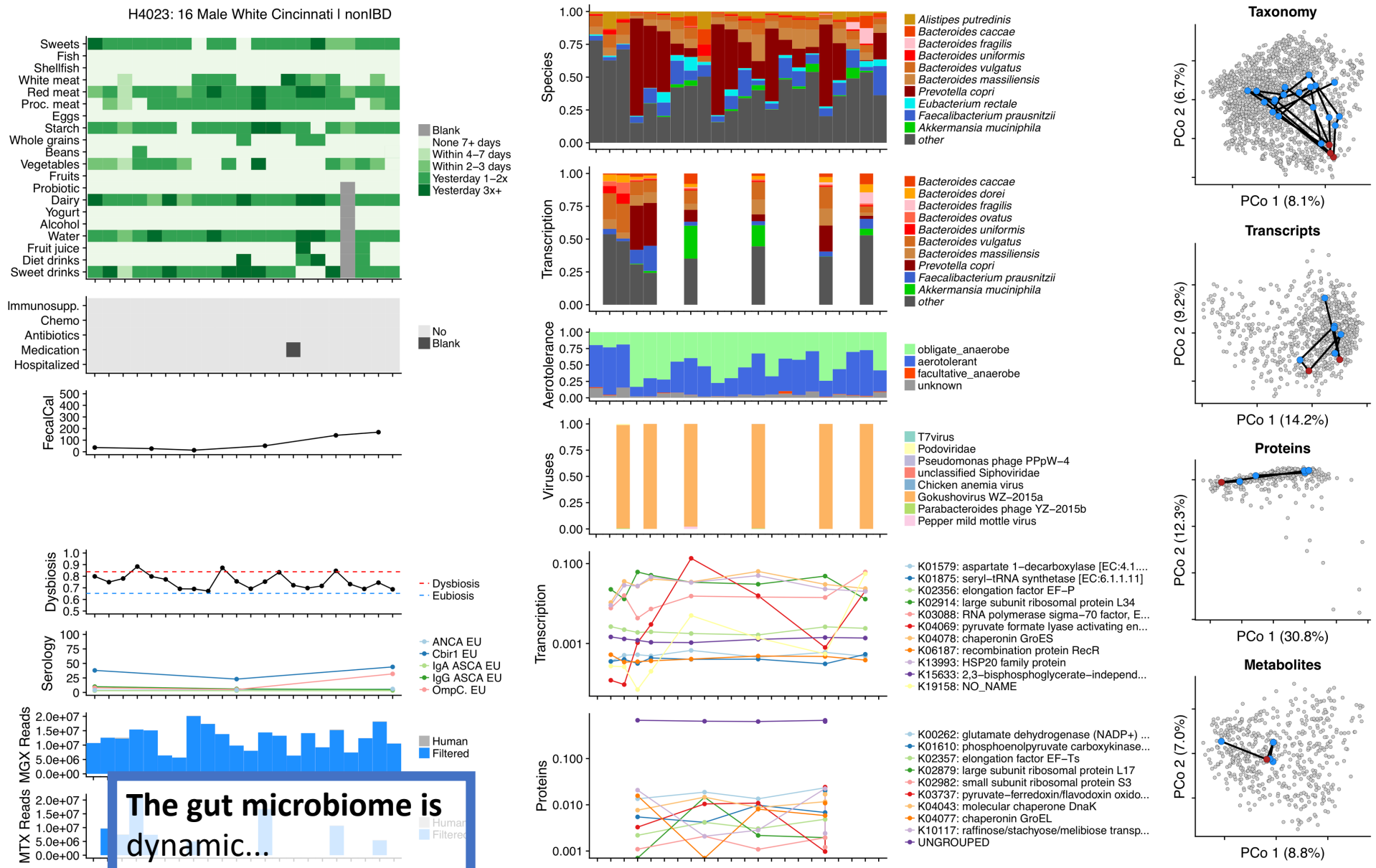




Funded by National Institutes of Health, Dept. of Healt



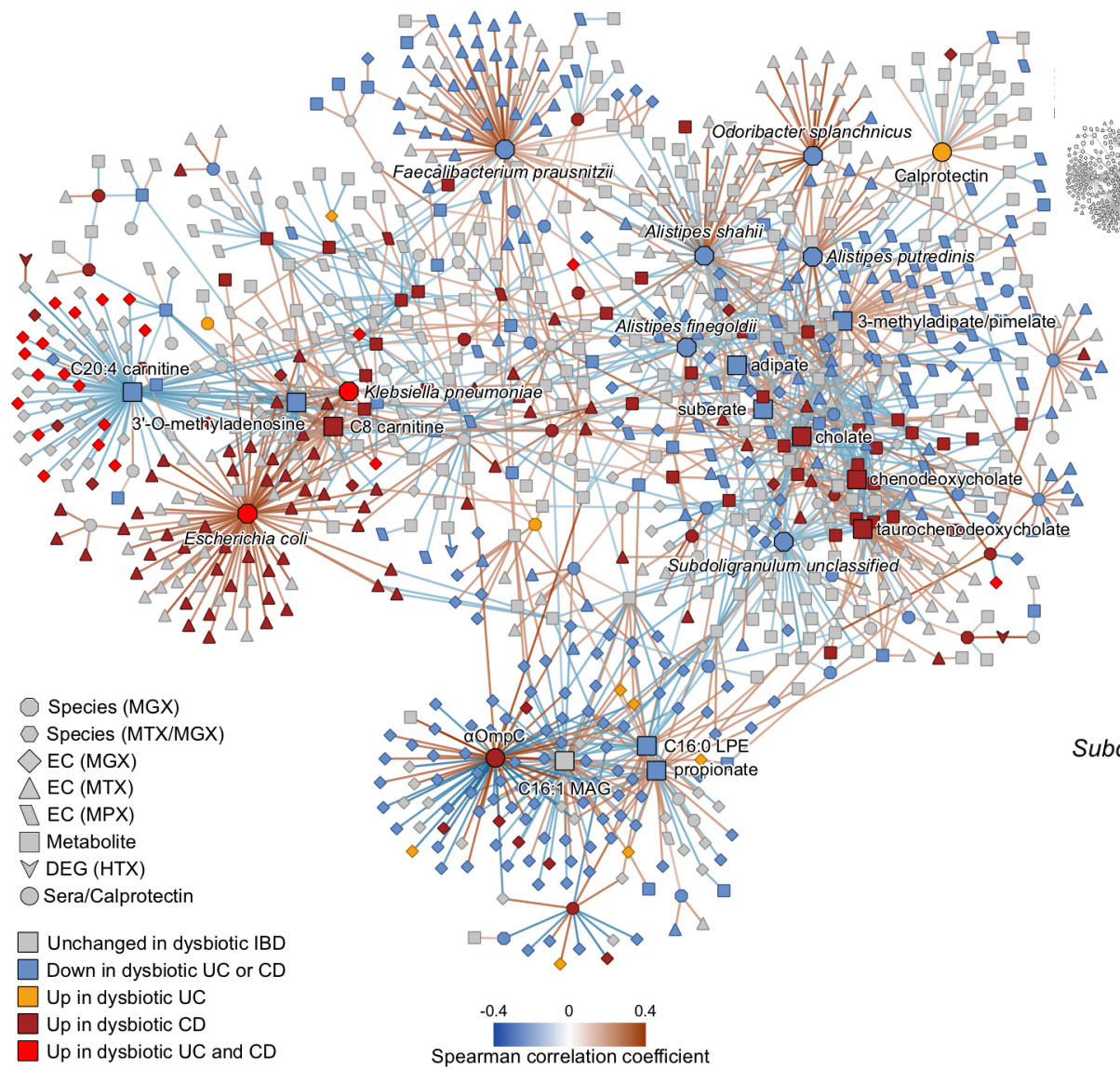
The gut microbiome is dynamic...

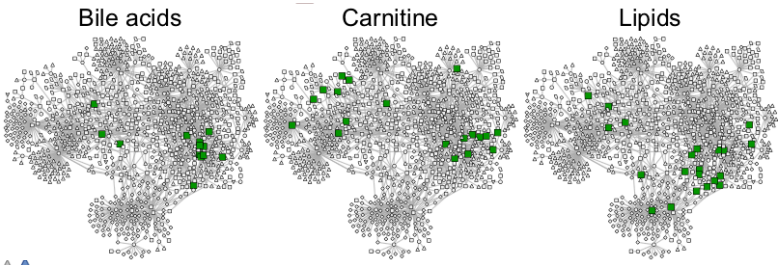


The gut microbiome is

dynamic...

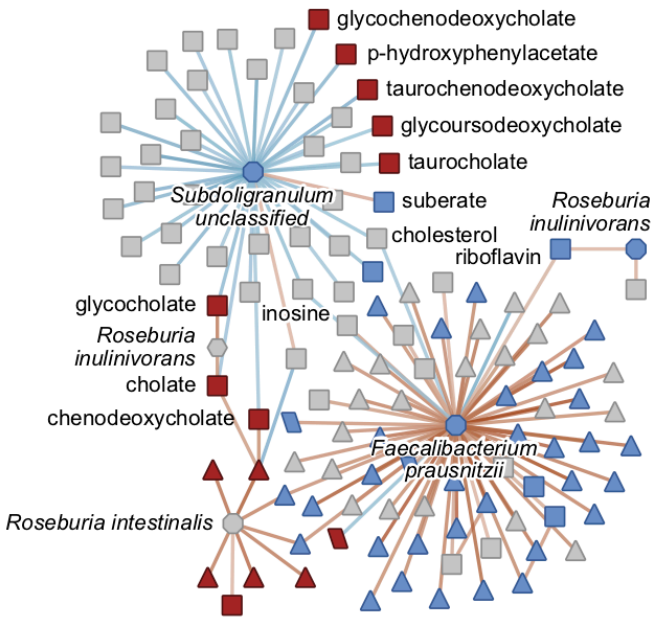
personalized...



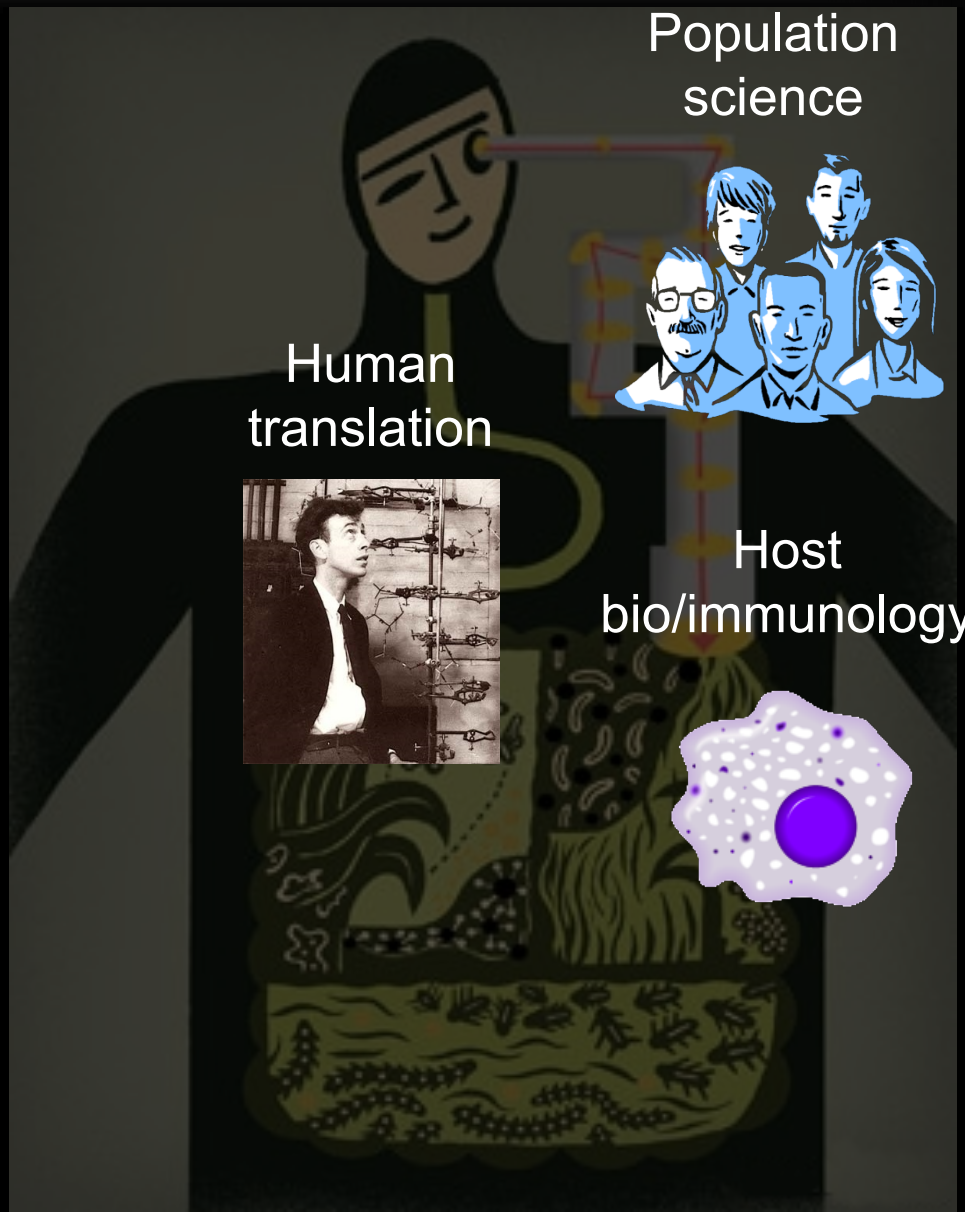


The gut microbiome is
dynamic...
personalized...
and complex!

Subdoligranulum, *Faecalibacterium*, and *Roseburia* sub-networks



Working toward high-impact outcomes from microbial communities and the microbiome



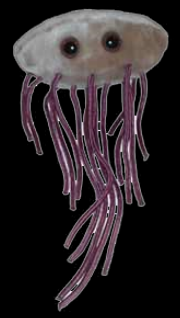
Population science

Environment and ecology

Human translation

Host bio/immunology

Microbiology



Working toward high-impact outcomes from microbial communities and the microbiome

Translation

Phenotype association for diagnostics

- Human disease risk: lifetime, activity, outcome
- Longitudinal analysis and study design
- Dense longitudinal measures, multiple nested outcomes

Systems analysis for intervention

- More and simpler model systems
- Systematic understanding of current models
- Ecological models for ecosystem restoration

Quantitative methods

Biostatistics

- Noisy multimodal data integration
- Dynamical systems models

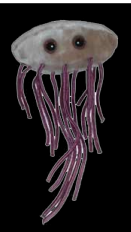
Computational biology

- Massive new data resource
- Machine learning for e.g. personalized medicine

Basic biology and molecular mechanism

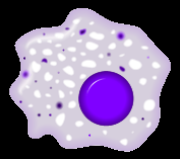
Microbial experiments

- Community quantitation
- Integration/meta-analysis of genomes and metagenomes



Host-microbe-microbiome interactions

- Immunity in specific host tissues
- Non-immune mechanisms (metabolites, peptides)
- Model system perturbations

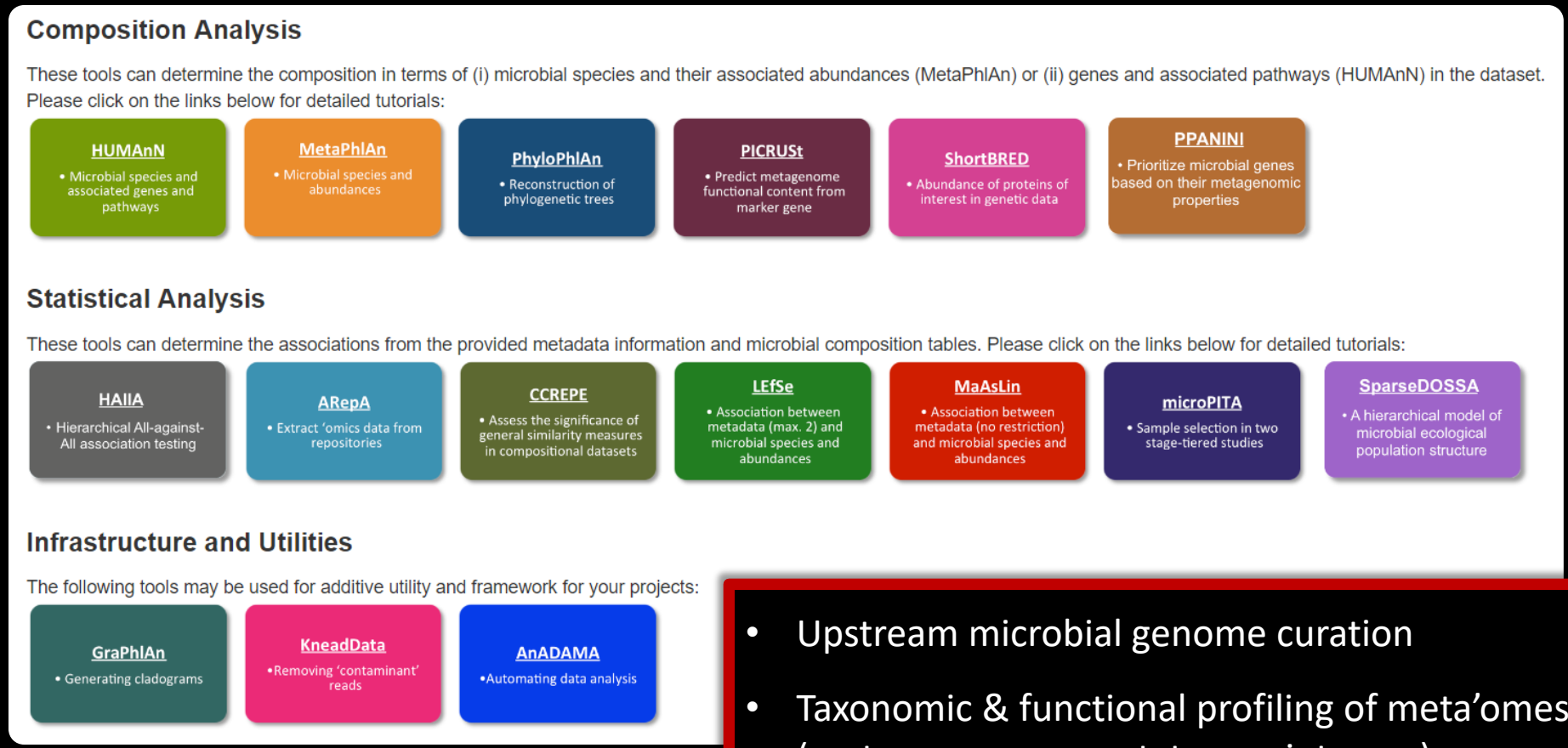


bioBakery

A virtual environment for meta'omic analysis

http://huttenhower.sph.harvard.edu/biobakery

First hit when you google "biobakery"



Source code, documentation, & tutorials

- Upstream microbial genome curation
- Taxonomic & functional profiling of meta'omes (metagenomes + metatranscriptomes)
- Microbial & clinical association discovery
- Reproducible workflows & infrastructure

Thanks!

http://huttenhower.sph.harvard.edu


Sena Bae


Yan Yan


Cesar Arze

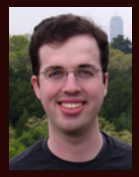

Lea Wang


Kevin Bonham

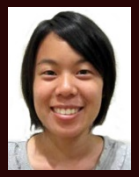

Lauren McIver

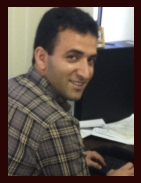

George Weingart


Ana Mailyan


Eric Franzosa


Daniel Wang


Tiffany Hsu


Ali Rahnavard

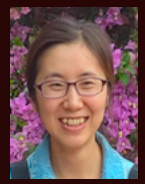

Jason Lloyd-Price

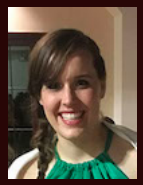

Dmitry Shungin


Tommi Vatanen


Jeremy Wilkinson


Siyuan Ma


Yancong Zhang


Kelsey Thompson


Emma Accorsi


Melanie Schirmer


Himel Mallick

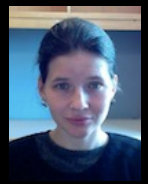

Jingjing Tang


Long Nguyen


Ramnik Xavier


Hera Vlamakis


Wendy Garrett


Xochitl Morgan

Human Microbiome Project 2

Lita Procter
Jon Braun
Dermot McGovern
Subra Kugathasan
Ted Denson
Janet Jansson
Owen White
Bruce Birren
Chad Nusbaum
Clary Clish
Joe Petrosino
Thad Stappenbeck


Nicola Segata


Edo Pasolli
Adrian Tett


Owen White


Anup Mahurkar