Gene-Environment Interactions, Gut Microbiome

Iona Cheng

March 2, 2021

“Genetics loads the gun, but the environment pulls the trigger”

-Francis Collins,
Director of National Institute of Health,
National Human Genome Research Institute

Complex disease result from the interplay of genetic and environmental factors

Gene-Environment Interactions

- Difference in the magnitude or direction of effect of an environmental exposure on disease risk/trait in people with different genotypes (or *vice-versa*)

- Effect modification

- Important because it may:
  - Identify sub-populations with environmental exposures/genotypes at increased risk
  - Clarify biological mechanisms of disease risk
  - Increase power in association analysis
  - Explain some of the missing heritability

Gene-Environment Interaction

- Examples: Individuals with different genotypes could differ in…
  - susceptibility to the health effects of exposures such as smoking, drinking, physical inactivity, etc.
  - responses to life events such as trauma
  - responses to medications (pharmacogenomics)

Statistical Terms of Interaction

	Interpretation
No interaction	The **same effect** of the exposure in people with different genotypes
Synergistic interaction	**Greater effect** of the exposure in people with a genotype of interest than in people with other genotypes
Antagonistic interaction	**Smaller effect** of the exposure in people with a genotype of interest than in people with other genotypes

Study designs to study G-E interaction

- Adoption studies
- Twin studies
- Molecular epidemiologic studies
  - Case-control
  - Nested case-control
  - Case-only

GxE Example: Factor V Leiden Mutations, Oral Contraceptive Use, and Venous Thrombosis

- Question: Could the occurrence of venous thrombosis in young women, who use OC, be explained by the factor V mutations?
- Study Design: Population-based case-control study (155 ca/169 co), women ages 15-49, conducted in Netherland
- OC information assessed my interview
- Factor V (thrombosis susceptibility gene) tested by cleavage products in ethidium-bromide stained agarose gels. Missense variant Arg>Glu (R506Q). Frequency 3-5% in European ancestry populations.

Vandenbroucke et al., The Lancet 1994

GxE: Stratifying by genotype Factor V Leiden Mutations, Oral Contraceptive Use, and Venous Thrombosis

Strata	Cases	Controls
G+E+	25	2
G+E-	10	4
G-E+	84	63
G-E-	36	100
Total	155	169

Odds Ratio

Factor V Leiden mutation:
(25x4)/(10x2) = 5.0

No Factor V Leiden mutation:
(84x100)/(36x63) = 3.7

Vandenbroucke et al., The Lancet 1994

GxE: Alternative Method
Factor V Leiden Mutations, Oral Contraceptive Use, and Venous Thrombosis

Strata	Cases	Controls
G+E+	25	2
G+E-	10	4
G-E+	84	63
G-E-	36	100
Total	155	169

Odds Ratio

G+E+: 34.7

G+E-: 6.9

G-E+: 3.7

G-E-: Reference

Vandenbroucke et al., The Lancet 1994

GxE: Case-Only Study

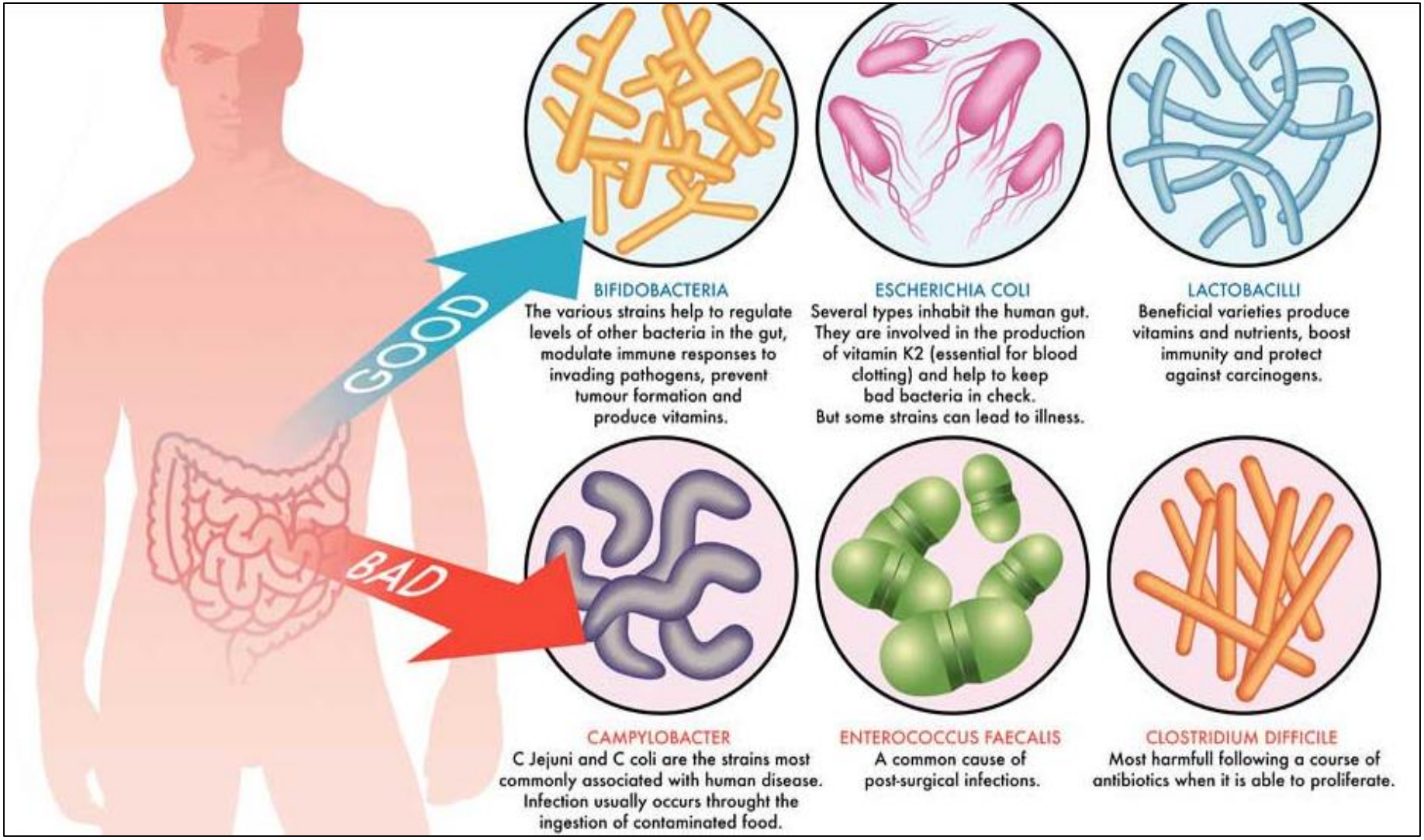
- Tests the association between the genotype and exposure in cases only
- More efficient study design than case-control study
- Greater statistical power than standard case-control studies with the same number of cases
- Key assumption: G and E must be uncorrelated in the population from which the cases arose (the “source population”)
- Higher false positive rate if assumption is violated

Possible bias in GxE Studies

Bias	Issues
Selection Bias	• Inappropriate control population • Incomplete case ascertainment • Who are the non-responders? Those that refuse to participate or unable to provide data or blood/saliva sample?
Information Bias	• Questionnaire errors • Specimen handling • Laboratory assays: Need for quality control measures
Confounding	• Population stratification • Epidemiologic confounders

Garcia-Closs et al., 2004

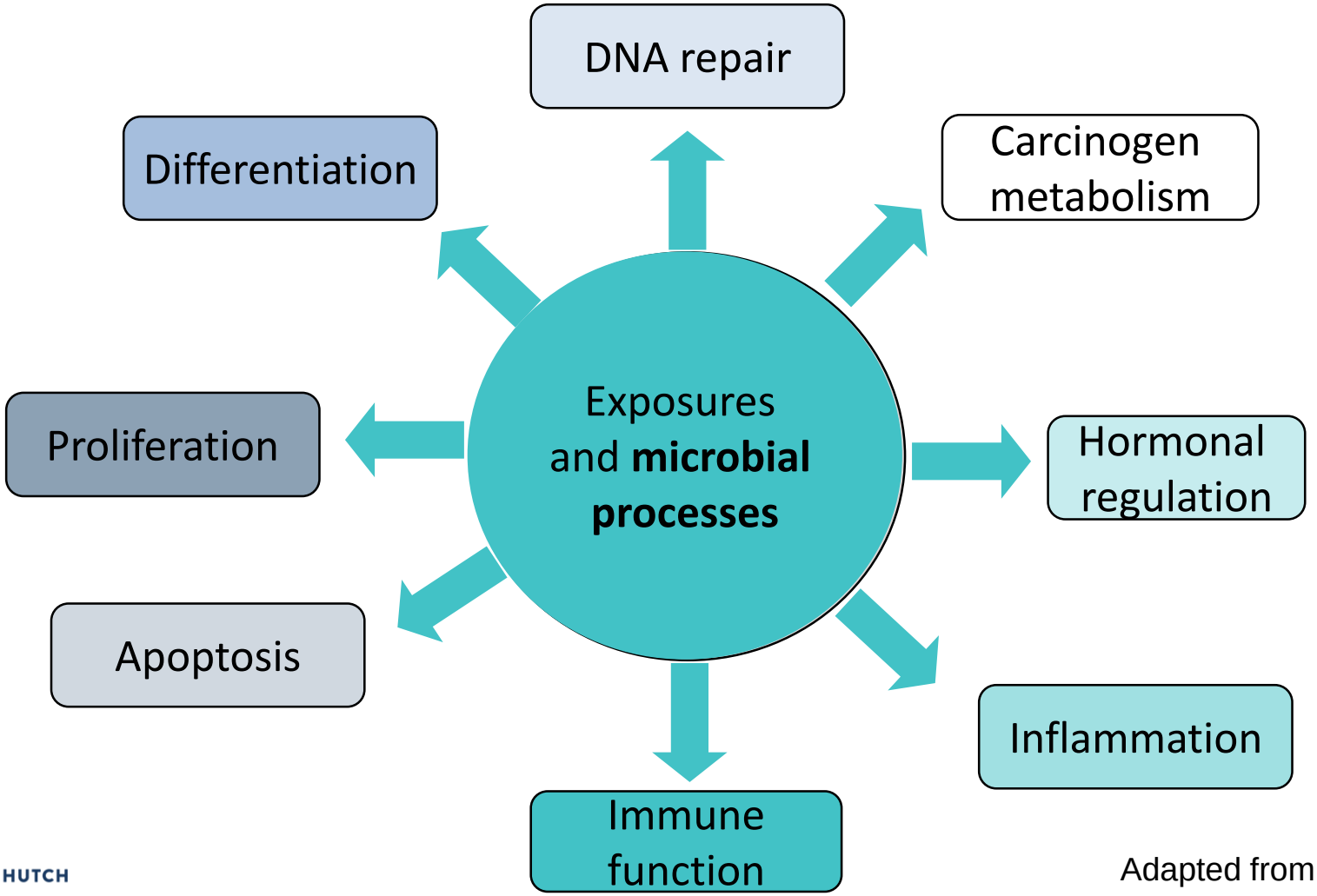
III. Human Gut Microbiome



~3.3. million genes in the human gut microbiome

Exposures and Cellular Processes Linked to Health Outcomes

Involvement of Microbial Processes



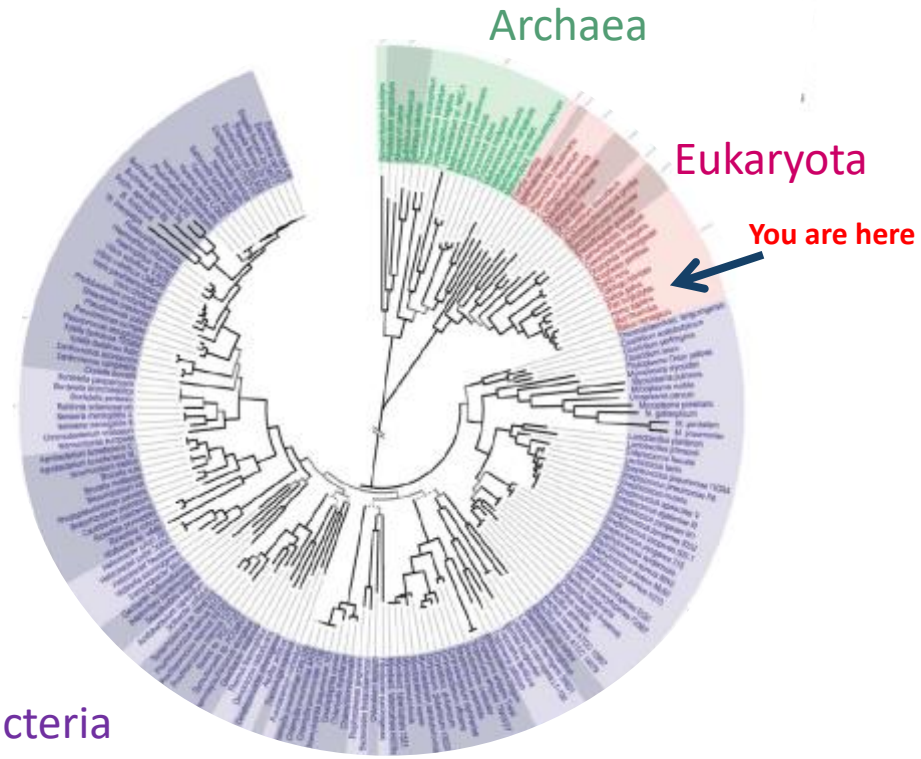
FRED HUTCH

Adapted from
WCRF/AICR 2007 Expert Report

Microbiome in Epidemiology

- **Marker of exposure**
  - **Direct effects:** pathogen
  - **Indirect effects:** microbial metabolites
- **A prognostic factor**
  - Eradication of a pathogen
  - Dysbiosis: microbiome out of balance that leads to a disease state
- **A factor in disease etiology**
  - Disease risk

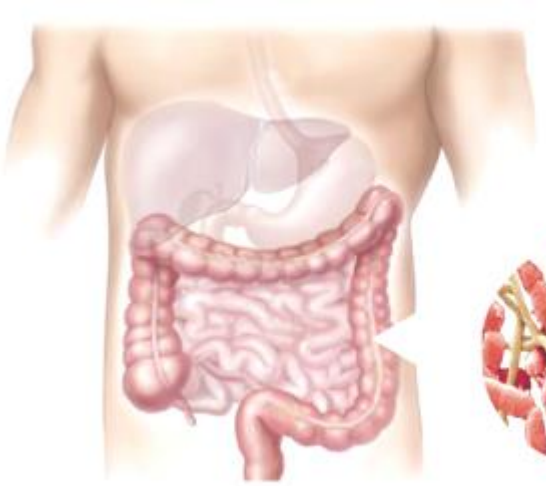
Phylogenetic tree of Life



- Archaea, Bacteria, Eukaryotes, Viruses
- Genome size varies
- Gene content varies
- Cell types vary
- Consortia of bacteria
- Many are not in pure culture –

Ciccarelli et al, 2006,Science,311:1283

Molecular Approaches to Measure the Gut Microbiome



Which microbes are there?

DNA → 16S rRNA genes

What is the functional potential?

DNA → Metagenomics

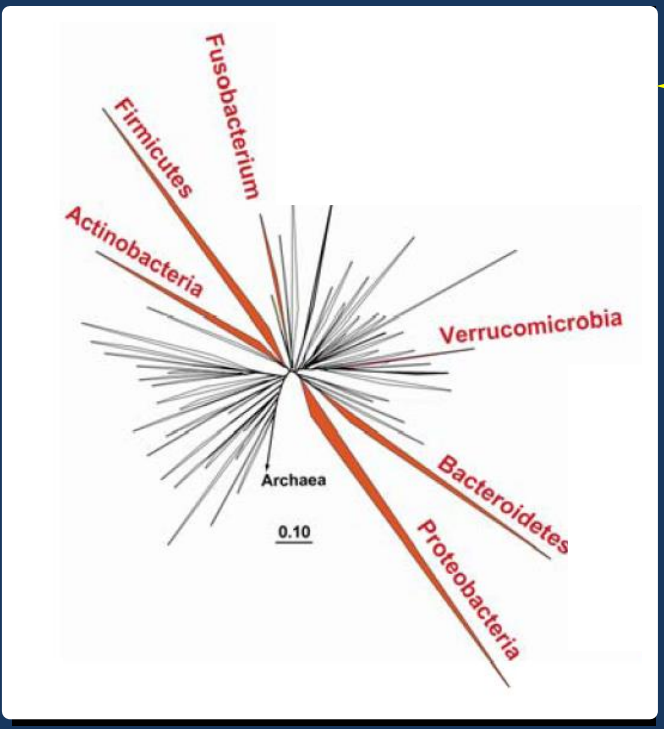
What are the microbes doing?

RNA → Metatranscriptomics

Proteins → Metaproteomics

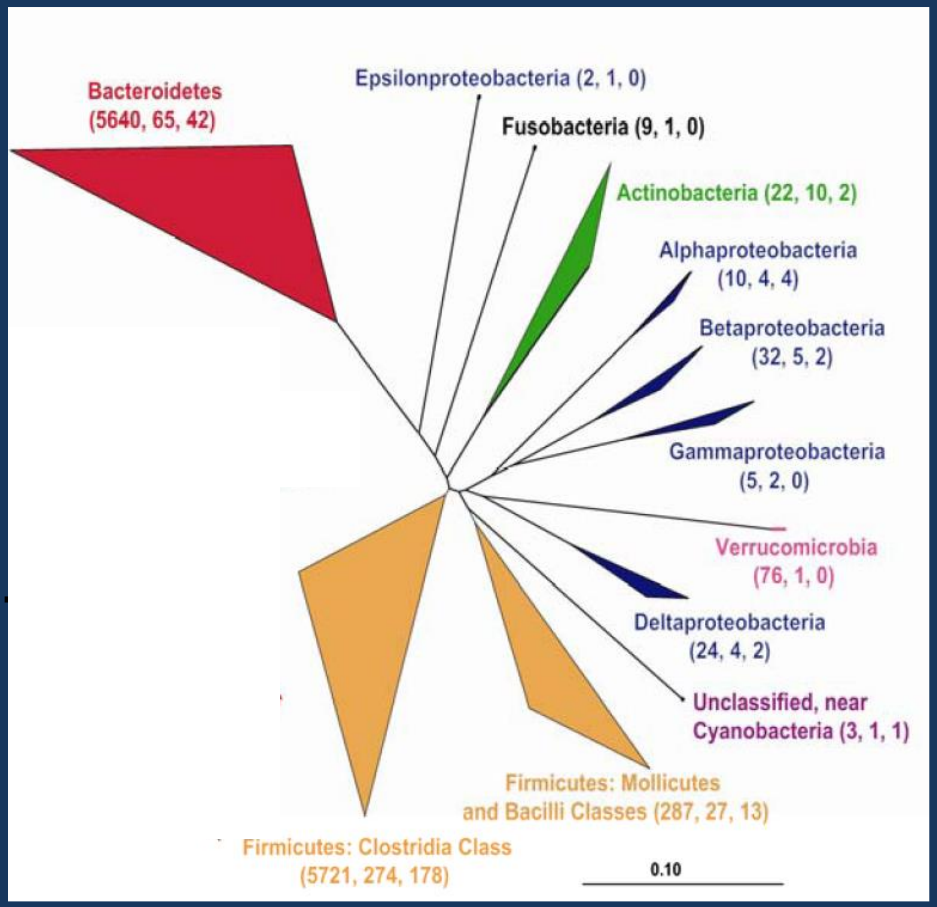
Metabolites → Metabolomics

Bacterial Diversity in the Human Gut



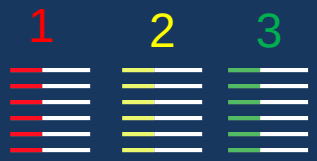
Only 6 major phyla

>90% are Firmicutes & Bacteroidetes



(Eckberg et al., 2005, *Science*)

Assaying the Gut Microbiome: 16S rRNA Gene Pyrosequencing



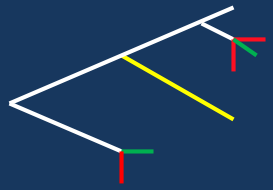
Tag 16S genes with DNA barcode

Mix samples and sequence

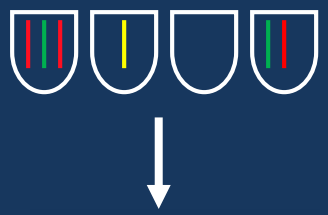
Barcode ID's the sample

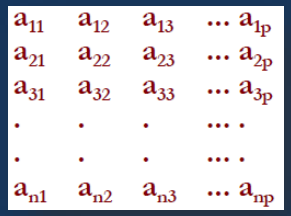
Clip barcode and Align sequences

Phylogenetic analysis



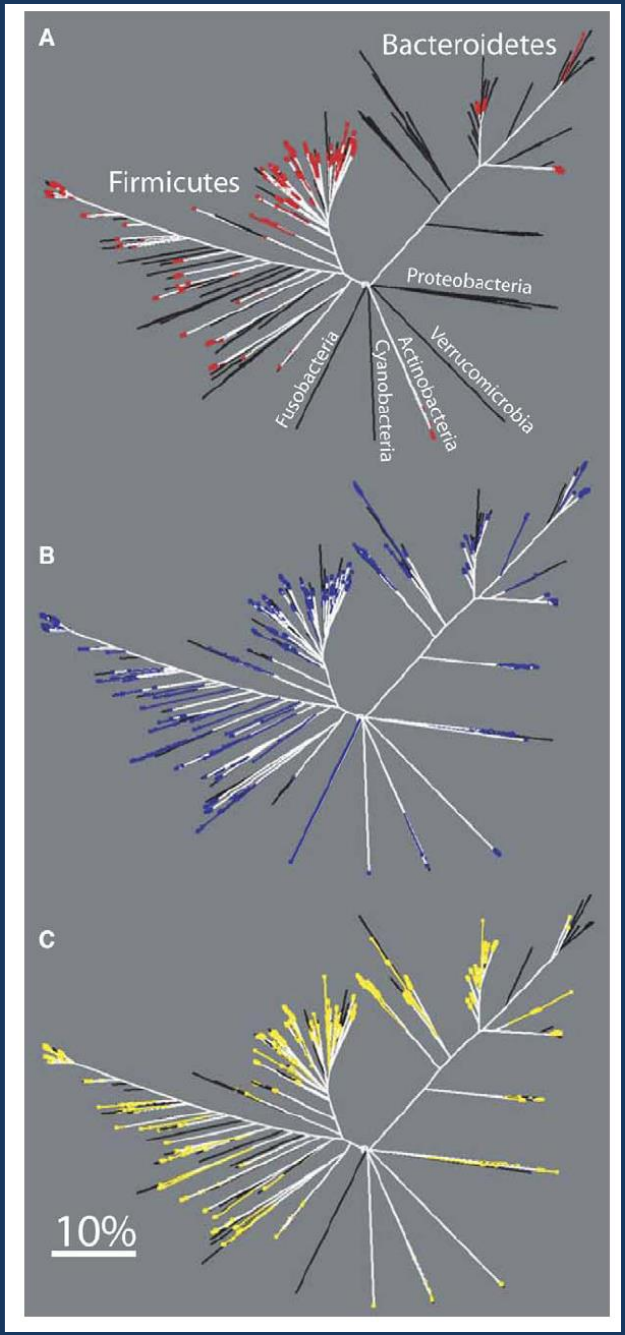
Binning





Multivariate analysis

(After Hamady, et al.,2009, Genome Res.)



16S rRNA Gene Analysis

Intra-individual variation (Alpha diversity)

* the distribution of the different kinds of bacteria within an individual
* Shannon Index/Phylogenetic Diversity

Inter-individual variation (Beta Diversity)

* shared (white)
* unique to an individual (red, blue, or yellow)
* absent (black)
* Unifrac (distance metric)

Objective:

To conduct a genome-wide association study of the gut microbiome in the Multiethnic Cohort

First questions:

1. Quality control assessment of the gut microbiome

2. What will be our base model for the gut microbiome for GWAS analysis?

3. How does the gut microbiome vary by sex, race/ethnicity, lifestyle factors?

Multiethnic Cohort Gut Microbiome Study Population

- Target 2,000 Core P01 MEC subjects (genotype)
- Target 4,200 MEC control subjects with existing GWAS data
- Screening questionnaire:
  - Exclusions:
    - Chemotherapy, insulin, hormones, weight-loss drugs within 6 mos
    - Antibiotics within past 2 mos
    - Bowel cleaning, colonoscopy within 2 mos
    - Substantial weight gain or loss past 6 mos
- Participation rate similar in men and women: Whites (60.9%), Native Hawaiians (53.6%), Japanese Americans (48.7%), African Americans (47.9%), Latinos (45.0%)

Biospecimens

overnight fasting (>8hrs) blood

5 x 10 mL

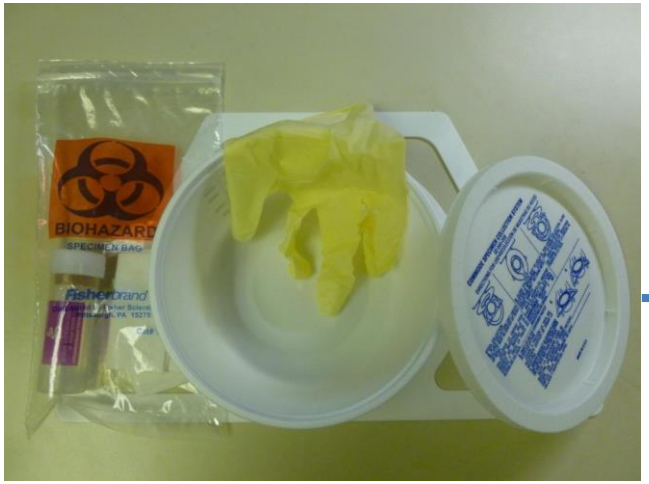
on ice (<4hrs)

(-80 °C; UH, USC)
Buffy coat
Clot
Lymphocytes
Plasma
RBC
Serum

shipped to labs before analysis

mouthwash

stool kit for collection at home



brought in person to clinic or via mail

UH
USC

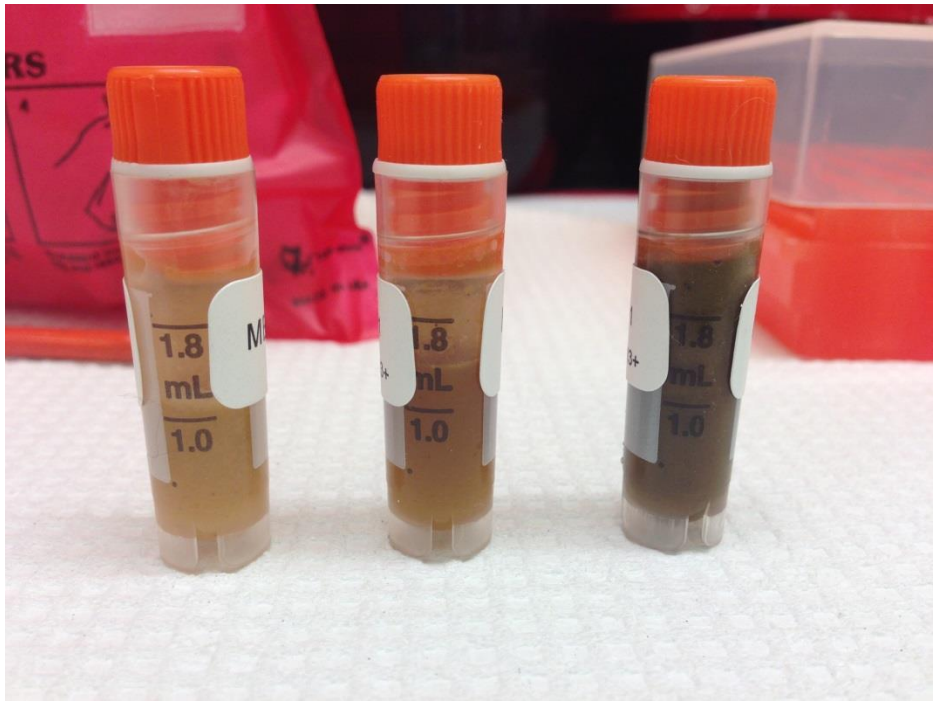
periodically shipped in bulk to FHCRC

DNA Extraction Flow Chart

Samples thawed and noted
Homgenized, aliquoted
Frozen at -80C
Data entered in SQL database

Microbial DNA extraction
QC:
- DNA yield
- DNA Quality

If they pass QC:
- Sent for sequencing



Sample Processing and DNA Extraction

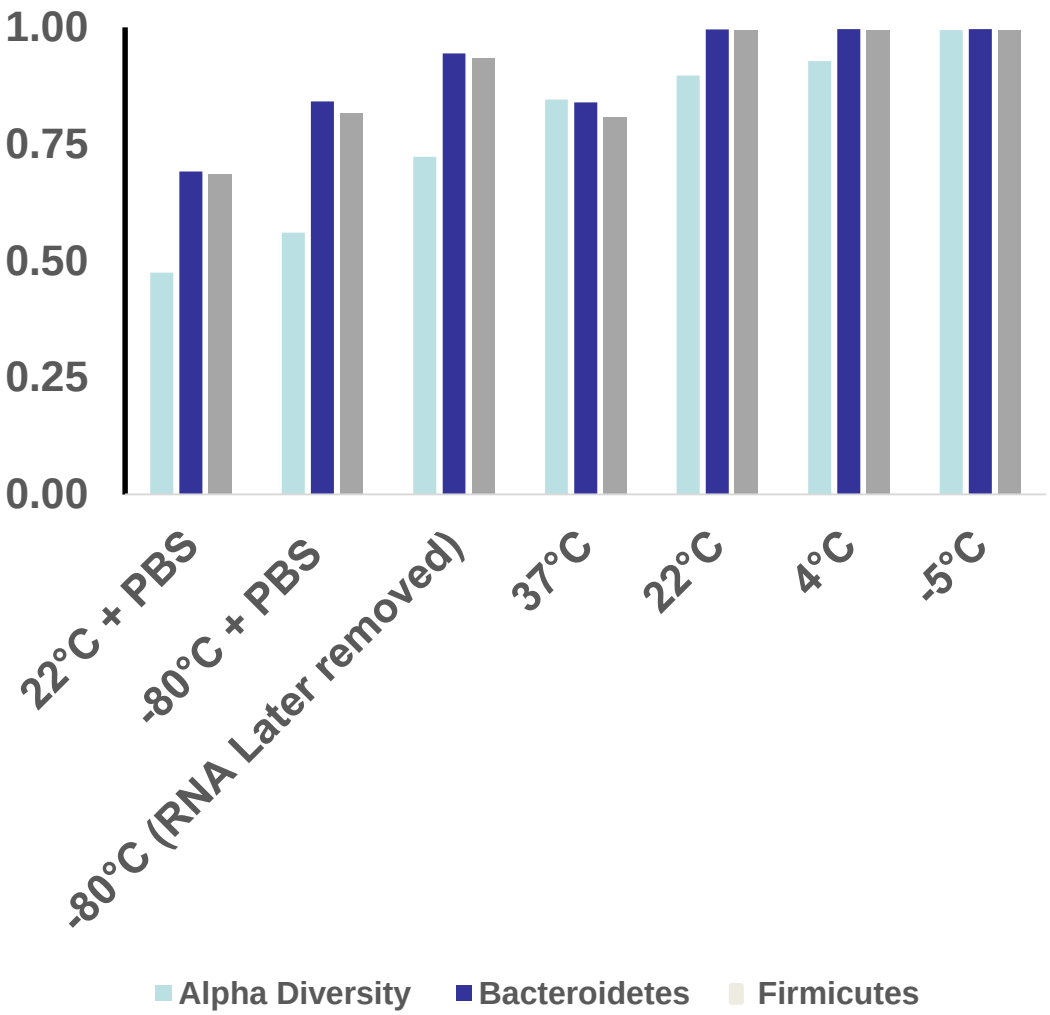
Received: 6719 in RNAlater
Extracted: 6708 samples
Sent for sequencing: 6641
Absolute fail: 7.0%
Passed sequencing: 6176

DNA extraction
(n=3 subjects, duplicate extractions)
ICC ≥0.93 for alpha diversity

ICC ≥0.99 for beta diversity

ICC ≥0.97 for the most abundant phyla.

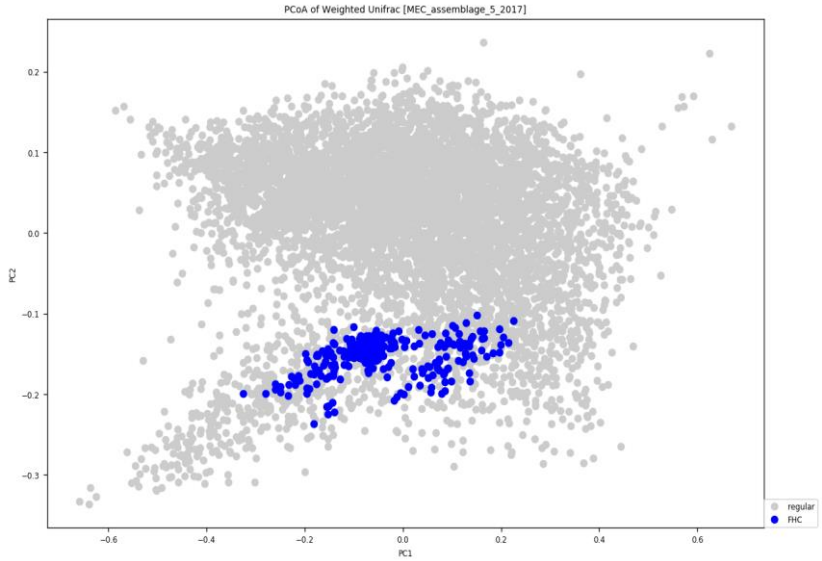
NTC controls: no amplification



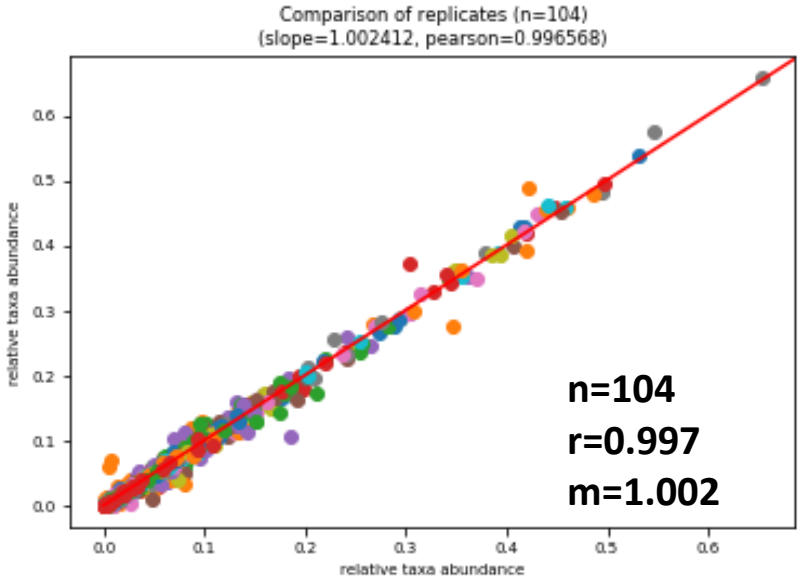
FRED HUTCH

Quality Control Measures

- FHC control

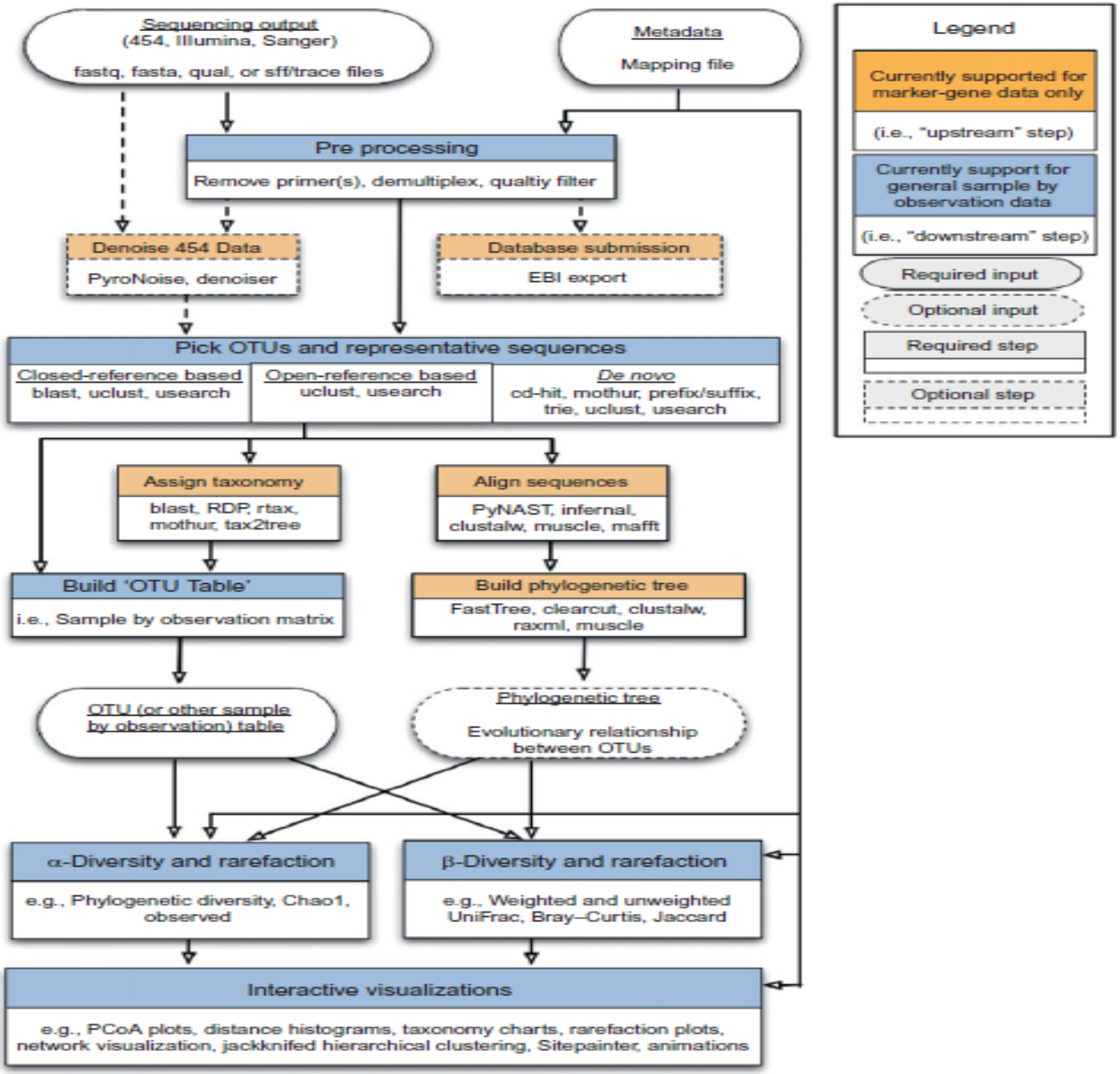


- Duplicate MEC sample



FRED HUTCH

QIIME Pipeline



The Multiethnic Cohort Microbiome Reliability Study

Assess whether one sample from a participant captures the microbiome

N~ 50 MEC participants every 6 mos for 2 yrs
Recruited by race and sex
Stool (Microbiome)(n~250)

Gut Microbiome

16S rRNA gene sequencing $V_{1\text{-}3}$
Illumina paired-end sequencing

Intra Class Correlation (ICC)

$ICC = \sigma b_2 / (\sigma b_2 + \sigma \varepsilon_2)$

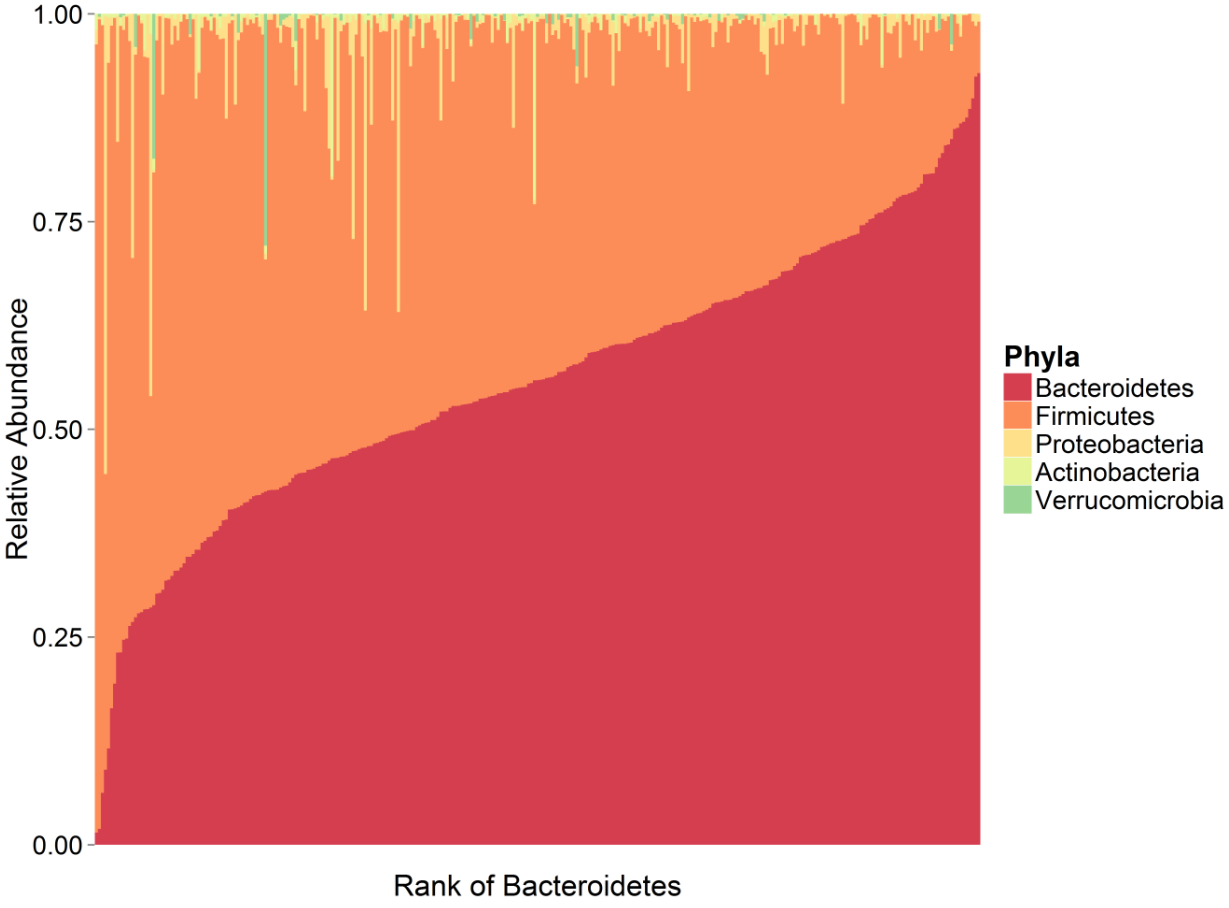
Reliability estimates of diversity, genera abundance
Weighted and unweighted UNIFRAC
Good reliability; ICC >0.6

B. Fu, Jlampe Labs

The Multiethnic Cohort Microbiome Reliability Study

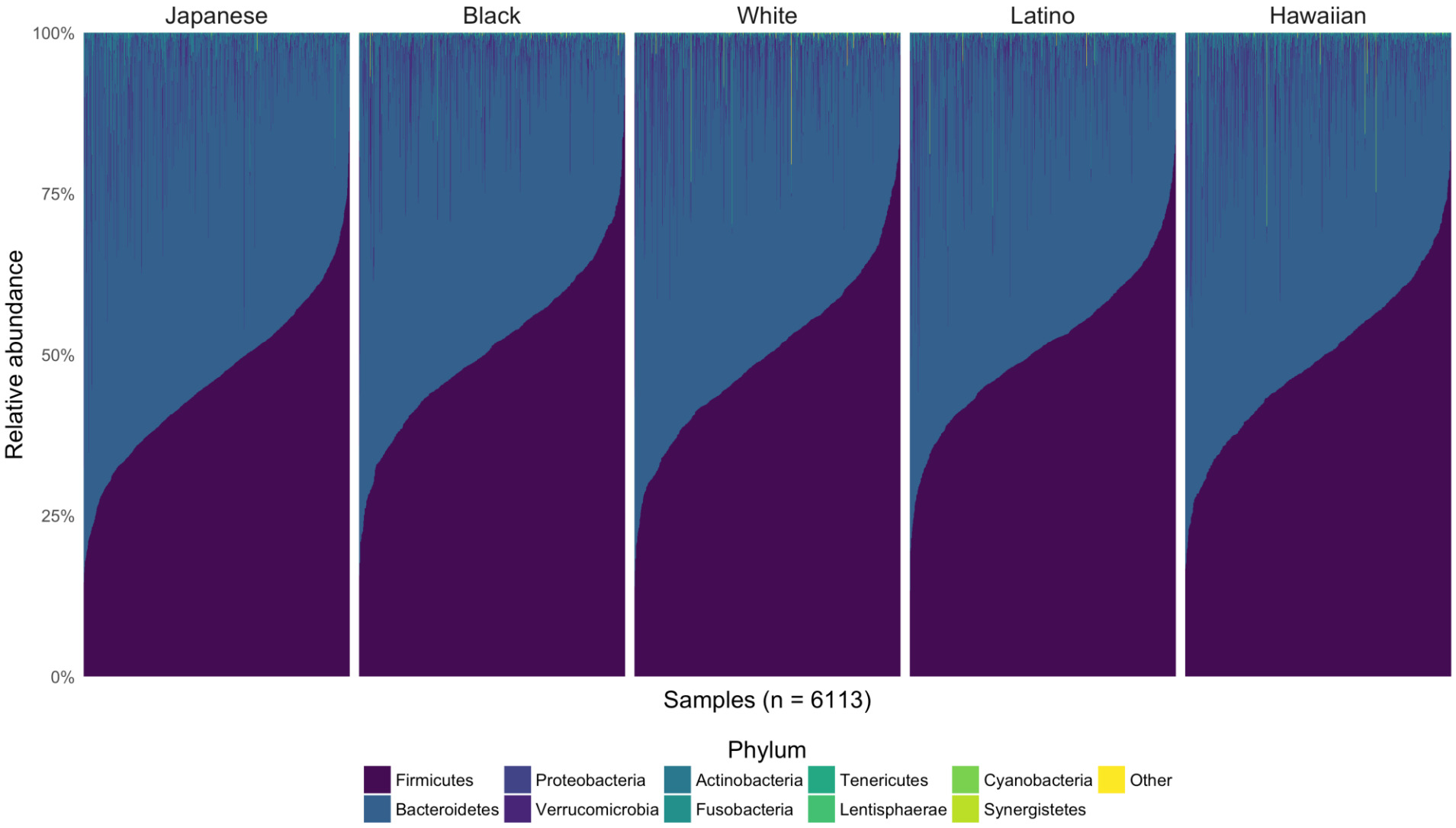
	ICC
Phylum	
Firmicutes	0.64
Bacteroidetes	0.62
Proteobacteria	0.65
Actinobacteria	0.67
Alpha diversity: diversity *within* sample	
Phylogenetic diversity	0.75
Shannon index	0.67
Chao1	0.56
Beta diversity: diversity *between* samples	
Unweighted UniFrac PC1	0.93
Weighted UniFrac PC1	0.65
Bray-Curtis PC1	0.94

Characterizing the Gut Microbiome: 16S rRNA gene

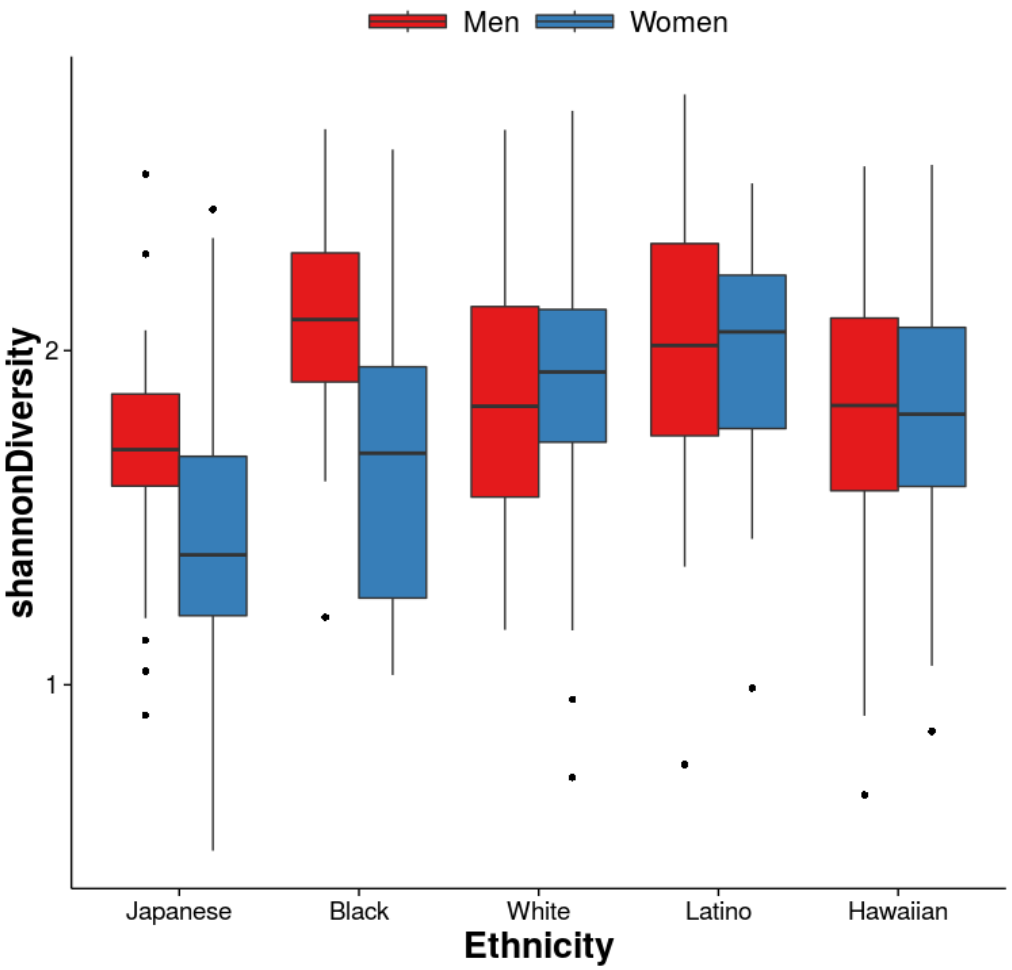


- 16S rRNA gene, V1-V3
- n=274
- High between person variation
- Bacteroidetes and Firmicutes dominate

The distribution of phyla (%) across self-report ethnicity

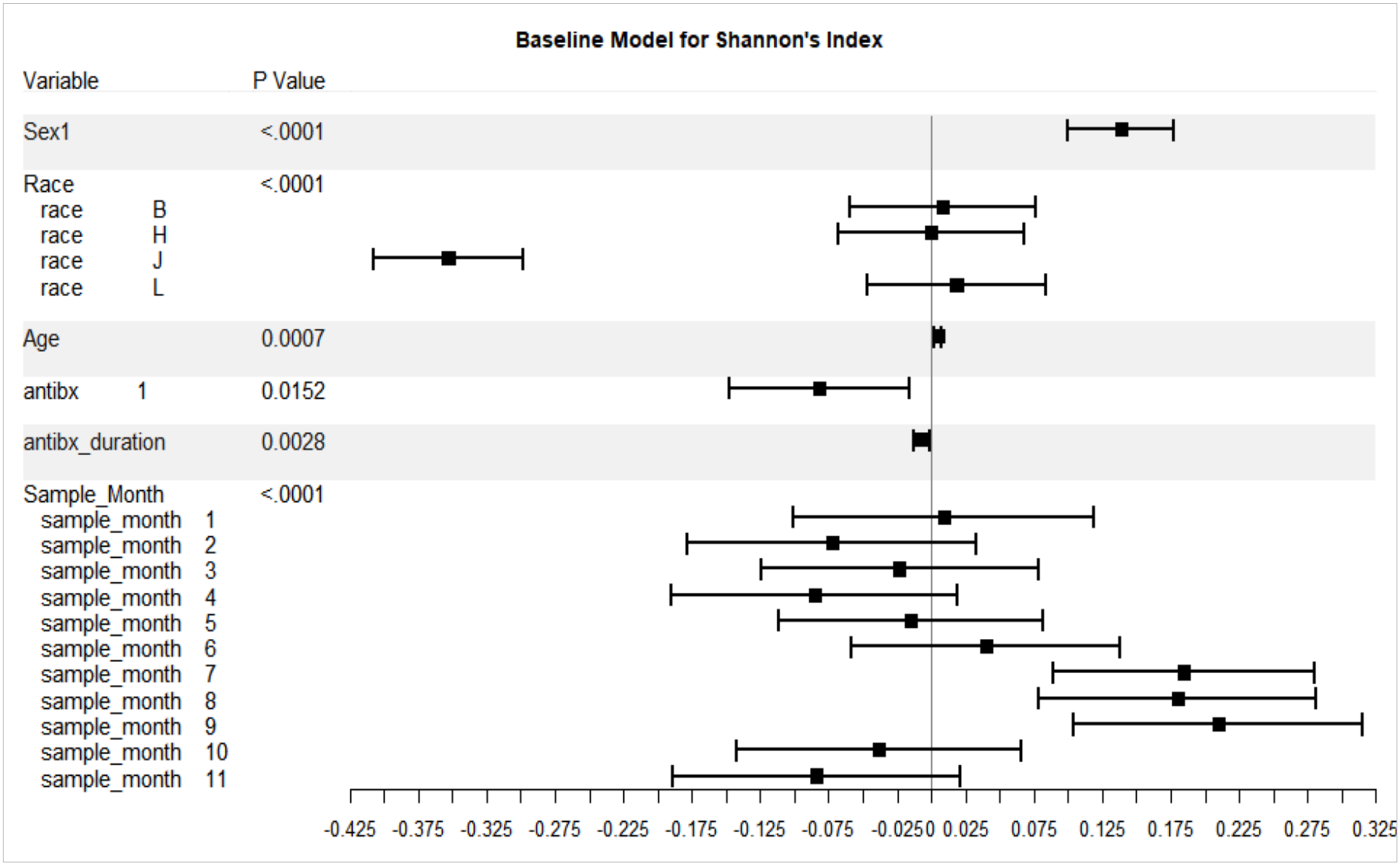


Characterizing the Gut Microbiome Diversity



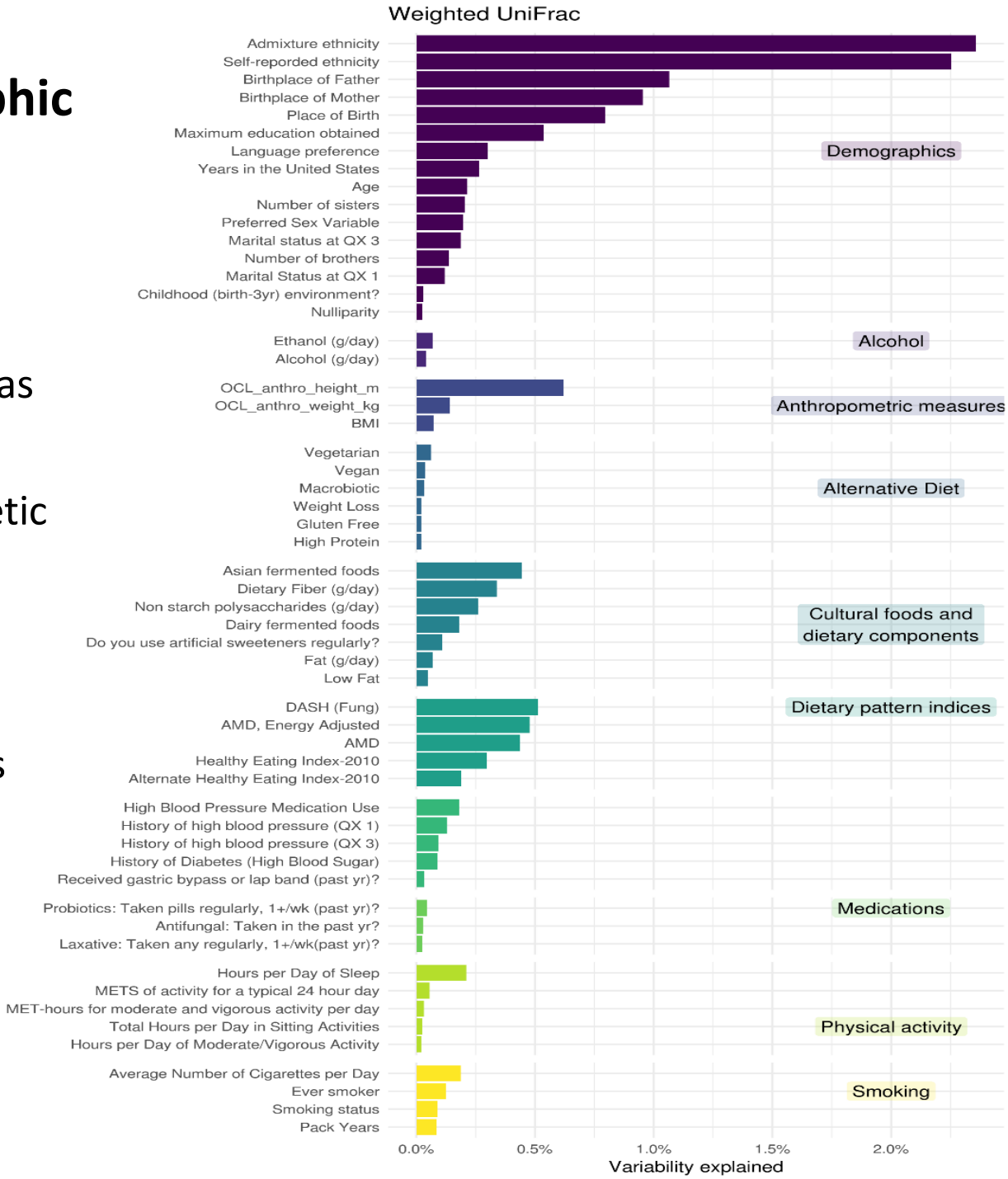
- Microbial diversity varies by sex in Japanese American and Blacks
- Diversity is similar between White, Latino, and Hawaiian men and women

Baseline model for Alpha-diversity

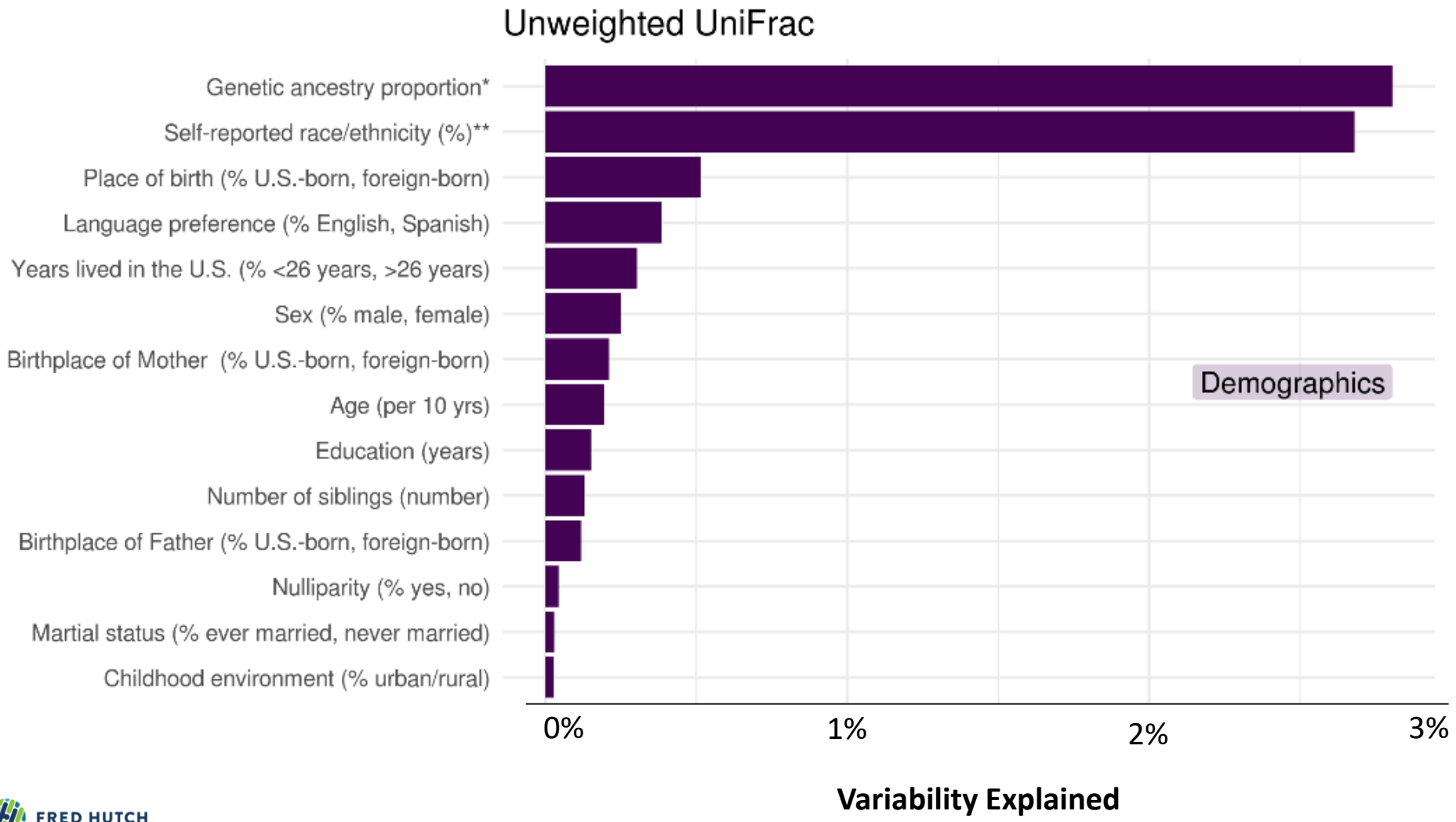


Variation in the GM explained by demographic and lifestyle factors (n=5,112)

- Each variable was modeled as perMANOVA
- Ethnicity, measured by genetic and self report ancestry, explained about 3% of the variation
- Dietary patterns and other dietary intake explained less than 0.5% of the variation.

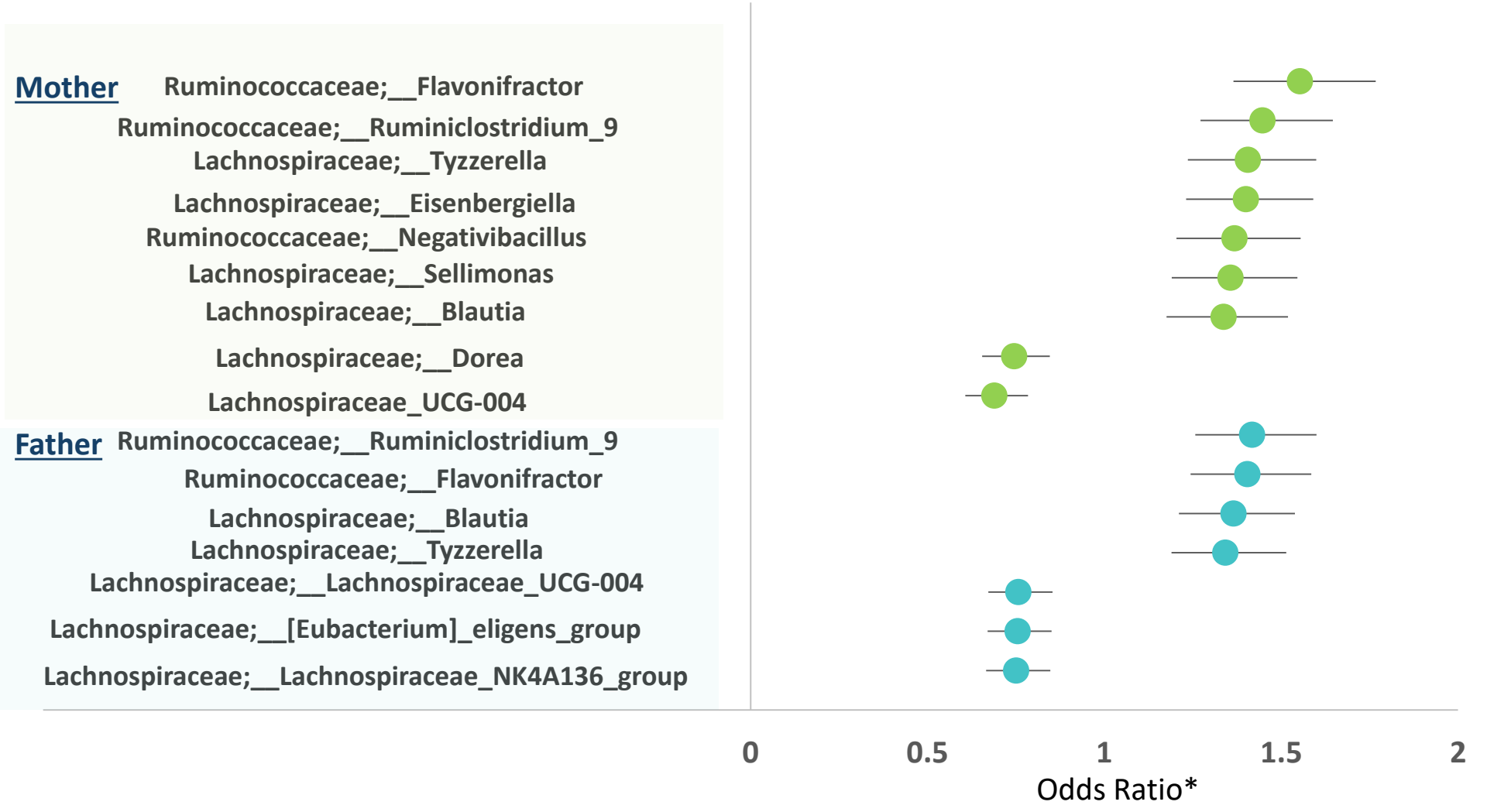


Association between Demographic Variables and the Microbial Community

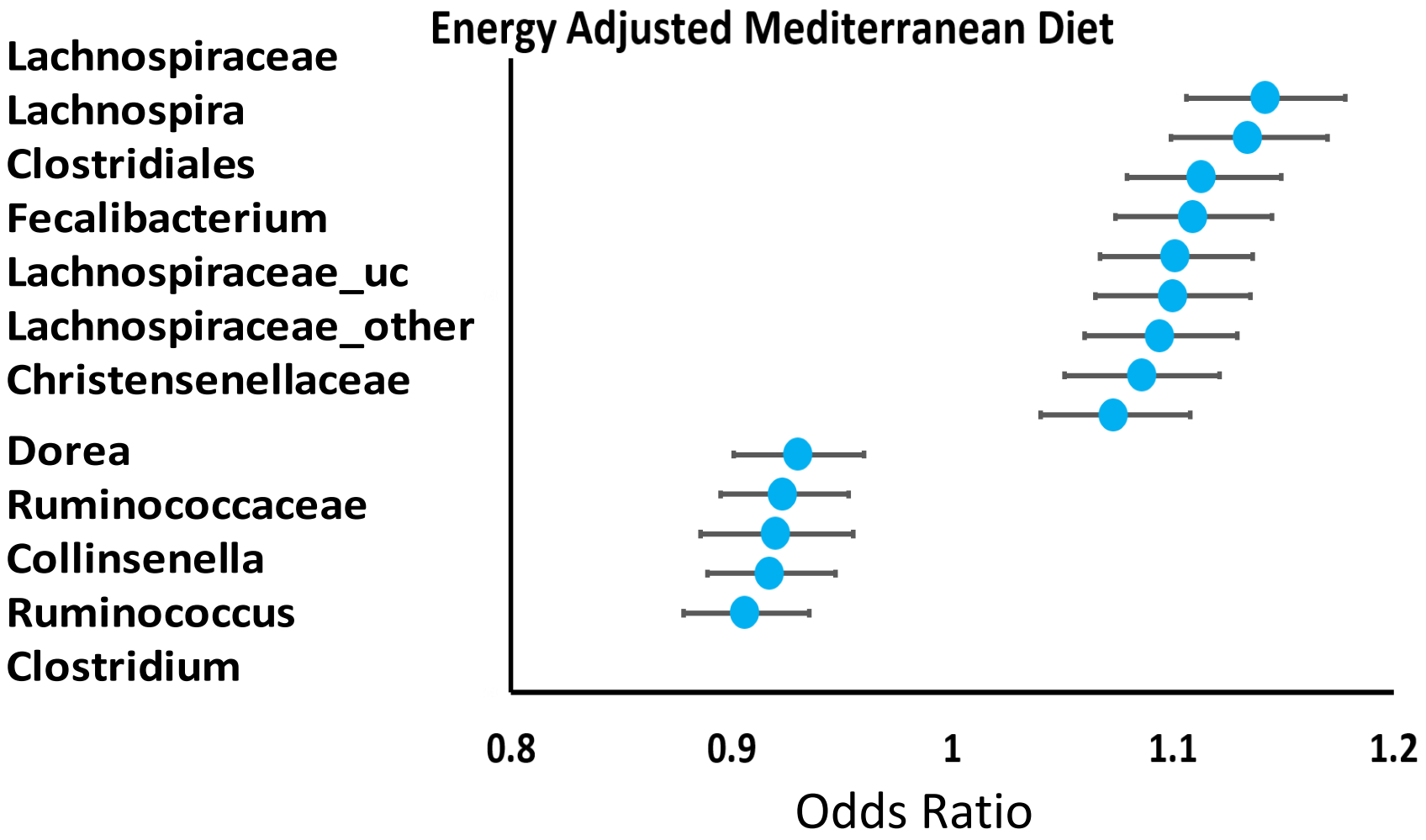


FRED HUTCH

Parental Birthplace May Influence Child's Gut Microbiome Composition

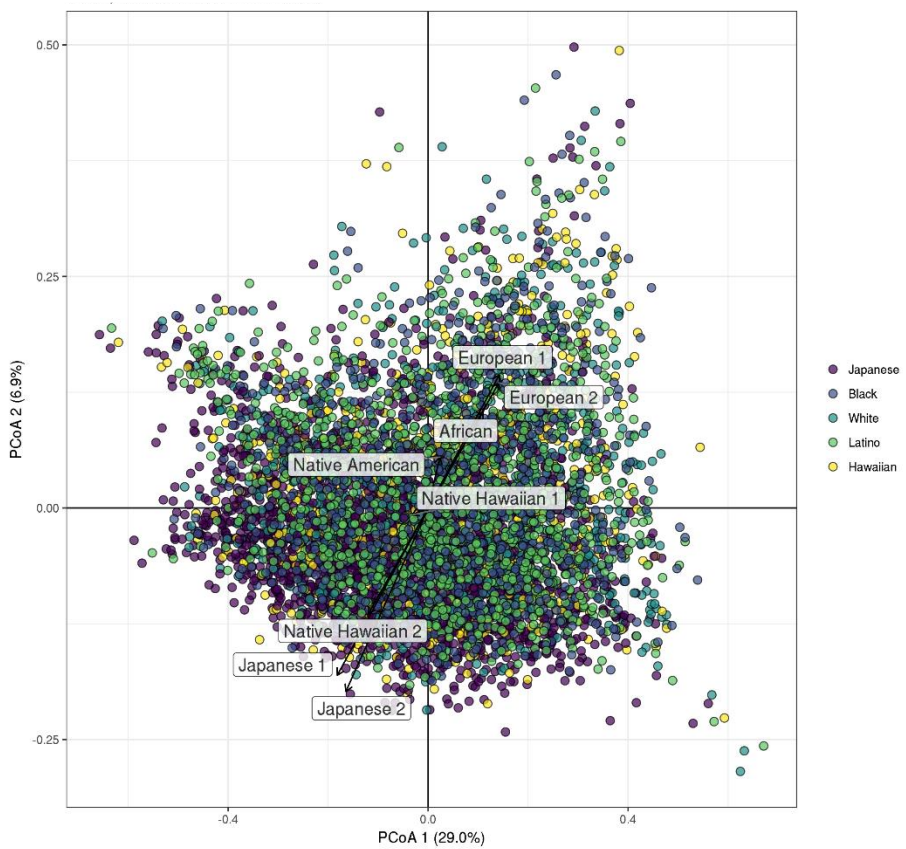


Gut Microbiome and Diet



- Higher dietary pattern scores (AMD) were associated with significant enrichment in bacteria associated with lean phenotypes (members of the family Christensellenaceae) and metabolism of phytochemicals.

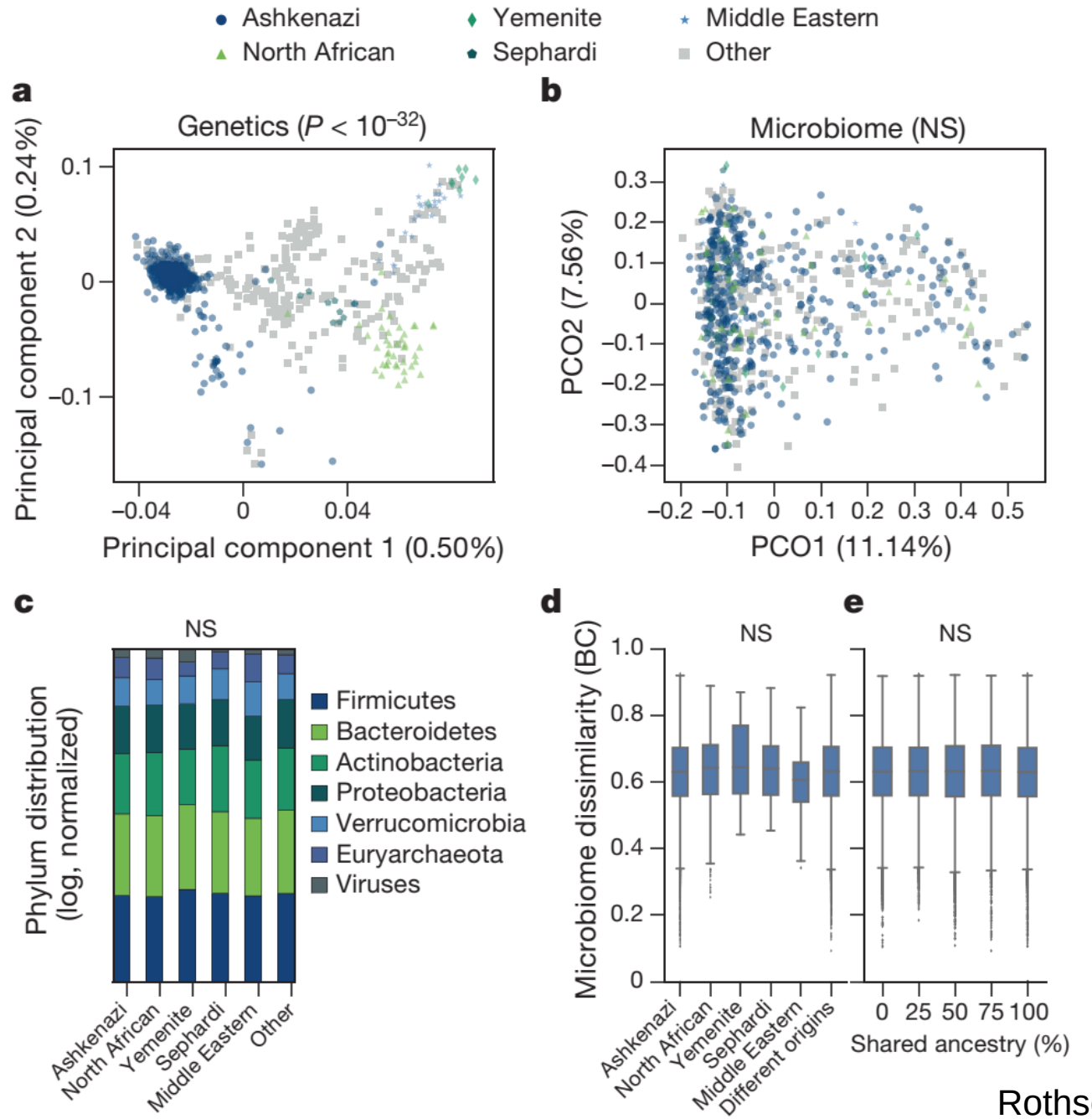
Self-Report Ethnicity, Genetic Ancestry, and the Microbiome



- The microbiome categorized across different 5 self-report ethnicities (n=6113)
- Vectors of ancestry proportions (n=5212)
  - 8 groups: South & North European, Japanese, Okinawa, African, Native American, Native Hawaiian, Native Hawaiian/Chinese/Filipino
- Microbiome composition varied across ethnicity

$R^2 = 0.026$, $p = 0.001$; perMANOVA

FRED HUTCH



Genetic ancestry is not associated with microbiome composition

Ashkenazi (*n*=345), North African (*n*=42), Middle Eastern (*n*=24), Sephardi (*n*=10), Yemenite (*n*=8) and admixed/other (other) (*n*=286) ancestries

Rothschild, Nature Genetics, 2018

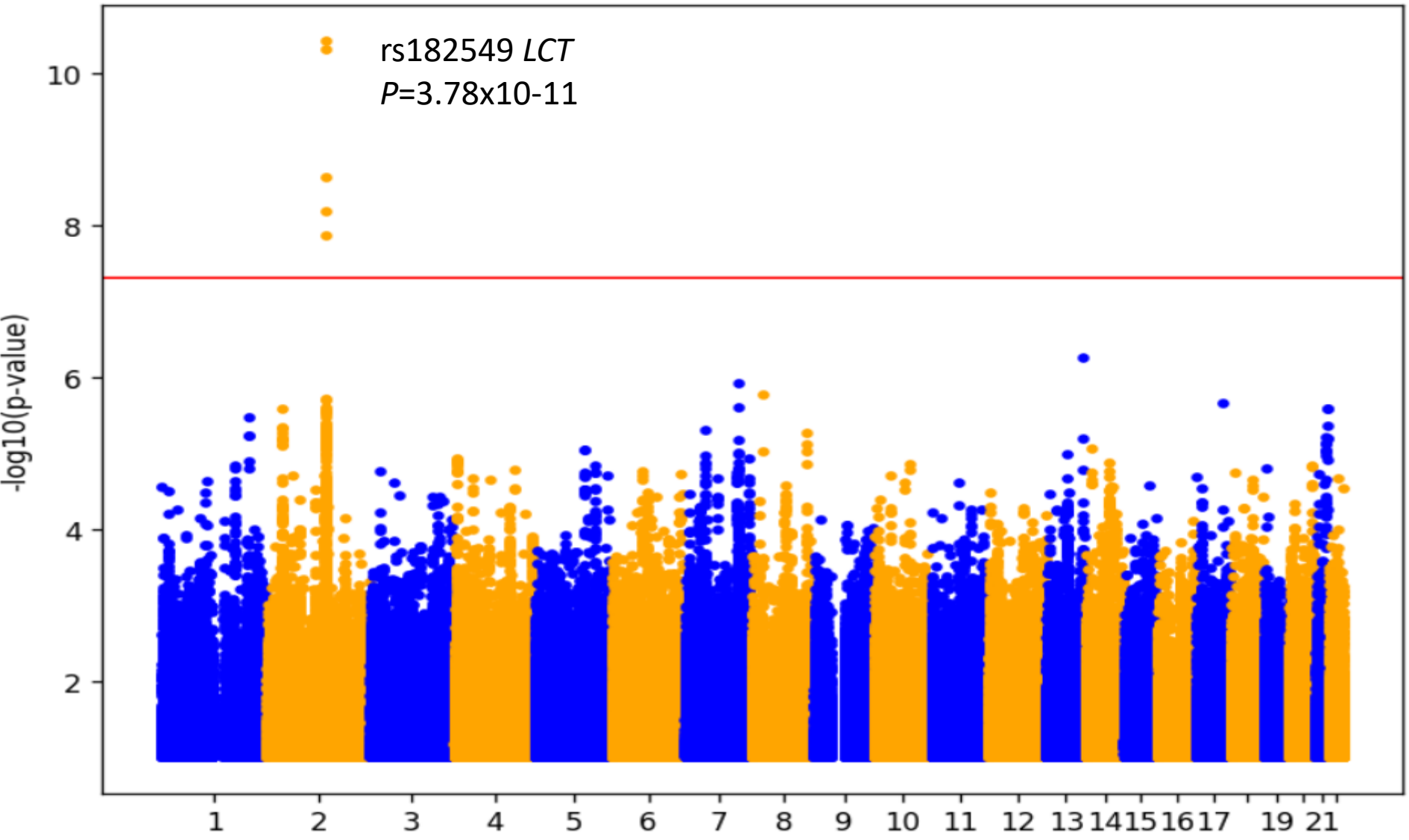
Gut microbiome and Host Genetics

- Twin studies
  - MZ and DZ similar GM compared to non-twin pairs ($p<0.009$)
  - Another study showed similarity decreases when twins start living a part
  - 33 significantly heritable bacteria taxa have been identified. Hereditability estimates 10-30%
- GWAS
  - Israel (n=1056): No association (Rothschild, Nature Genetics, 2018)
  - Germany (n=1812): *VDR* gene and gut microbial characteristics (Wang, Nature Genetics, 2017)
  - Dutch (n=1515); 9 loci with microbial taxonomies & 33 loci with microbial pathways and gene ontology terms (Bonder, Nature Genetics, 2016)
  - Hutterites (n=1415): No GWAS significant findings (Davenport, PLoS One, 2015)
- Meta-analysis GWAS (Kurilshikov, 2021)
  - MiBioGene Consortiu m Initiative
  - 18,340 (24 cohorts)
  - 31 loci meet $P<5 \times 10\text{-}8$ with lactase gene $P=1.28\text{x}10\text{-}20$ with Bifidobacterium abundance

GWAS of Gut Microbiome: Approach

- **Gut microbiome:** 16S rRNA gene sequencing V_{13} by Illumina paired-end sequencing
- **Genotype:** Genotype MEGA^{ex} array (n=2000) & existing GWAS (n=4000)
- **Outcomes**: alpha diversity, common genera
- **Genetic ancestry:** PCA of 42K independent SNPs
- **Genetic ancestry proportions:** ADMIXTURE program (Alexander, 2009)
- **Statistical Approach**: Ordinal logistic regression; adjusting for age, sex, PCs, antibiotic use, batch, season
- **Replication:** MiBioGene

Manhattan Plot Bifidobacterium



Summary

- The large study of the microbiome in a multiethnic population
- General characterization of the microbiome shows
  - Ethnicity/race explains a small but significant amount of variation in the microbiome
  - Bacterial genera are significantly enriched in different racial/ethnic groups
  - Acculturation across generation may be important
- Replicate association with the lactase gene and Bifidobacterium

Looking forward

- Very large and well characterized cohorts may be the key to finding associations between the microbiome and disease
  - Population-based
  - Well-described
  - Rigorous inclusion criteria
  - Rigorous data collection
- Standardization is needed across all aspects of sample collection, lab analysis, and bioinformatics analysis for pooling, as well as, new statistical approaches to account for covariates (Microbiome Quality Control Project, Sinha R et al, Nat Biotechnol 2017)
- Development of large databases and repositories for the increasing volume of complex data
- Allows pooling of information across studies which is especially important in cancer epidemiology and the microbiome

Questions
Thank you