Determinants of *S. aureus* carriage in the developing infant nasal microbiome

Emma Accorsi[1,2], Eric A. Franzosa[1,2,3], Tiffany Hsu[1,3], Regina J. Cordy[4], Ayala Maayan-Metzger[5,6], Hanaa Jaber[6], Aylana Reiss-Mandel[6], Madeleine Kline[2,7], Casey DuLong[1], Marc Lipsitch[1], Gili Regev-Yochay[5,6], Curtis Huttenhower[1,2,3]

[1]Harvard T. H. Chan School of Public Health; [2]Harvard Chan Microbiome in Public Health Center; [3]Broad Institute; [4]Wake Forest University; [5]Sackler School of Medicine; [6]Sheba Medical Center; [7]Havard Medical School

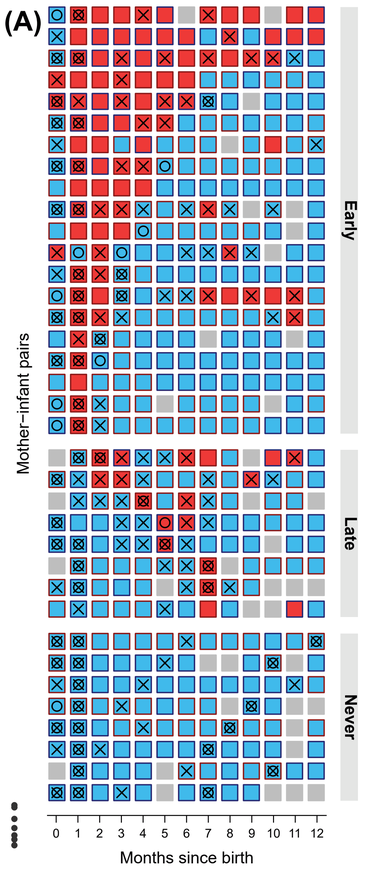
Staphylococcus aureus is a leading cause of healthcare- and community-associated infections and can be difficult to treat due to antimicrobial resistance. About 30% of individuals carry *S. aureus* asymptomatically in their nares, a risk factor for later infection, and interactions with other species in the nasal microbiome likely modulate its carriage. It is thus important to identify ecological or functional genetic elements within the maternal or infant nasal microbiomes that influence *S. aureus* acquisition and retention in early life. We recruited 36 mother-infant pairs and profiled a subset (n=208) of monthly longitudinal nasal samples from the first year after birth using shotgun metagenomic sequencing. The infant nasal microbiome is highly variable, weakly influenced by maternal nasal microbiome composition, and primarily shaped by developmental and external factors, such as daycare. Infants display distinctive patterns of *S. aureus* carriage, positively associated with *Acinetobacter* species, *Streptococcus parasanguinis*, *Streptococcus salivarius*, and *Veillonella* species and inversely associated with maternal *Dolosigranulum pigrum*. In gene-content based strain profiling, infant *S. aureus* strains are more similar to maternal strains. Mothers may represent a sporadic early source for *S. aureus* transmission to the naïve infant microbiome, but microbiome determinants become more important later in the first year. Furthermore, we identified a specific protein family that is highly predictive of infant *S. aureus* status, significantly anticorrelated with *S. aureus* positivity in both infants and mothers, sufficiently prevalent to drive widespread patterns of *S. aureus* carriage, and which ecologically interacts with the commensal species *D. pigrum*. In subsequent companion work, we determined that this (misannotated) protein family was a non-protein-coding sequence acting as a phylogenetic marker of a likely novel bacterial clade. Our study provides an improved understanding of how the infant nasal microbiome develops in early life, and how it can act to promote or exclude *S. aureus* colonization.

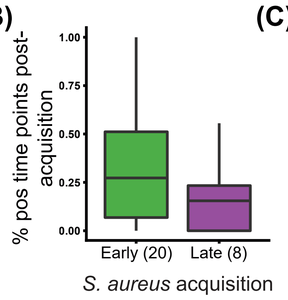
Infants display striking patterns of *S. aureus* carriage

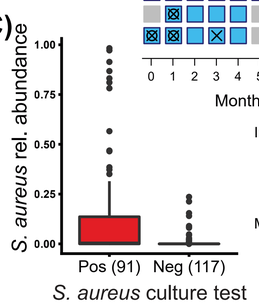
(A) 36 mother-infant pairs gave nasal swabs monthly over the first year after birth. Culture testing for *S. aureus* was performed on all samples and a subset (*n*=208 after QC) were profiled with shotgun metagenomic sequencing.

(B) The percent of positive time points after *S. aureus* acquisition was not significantly different between early and late acquirers, likely due to the small sample size (*n=28*).

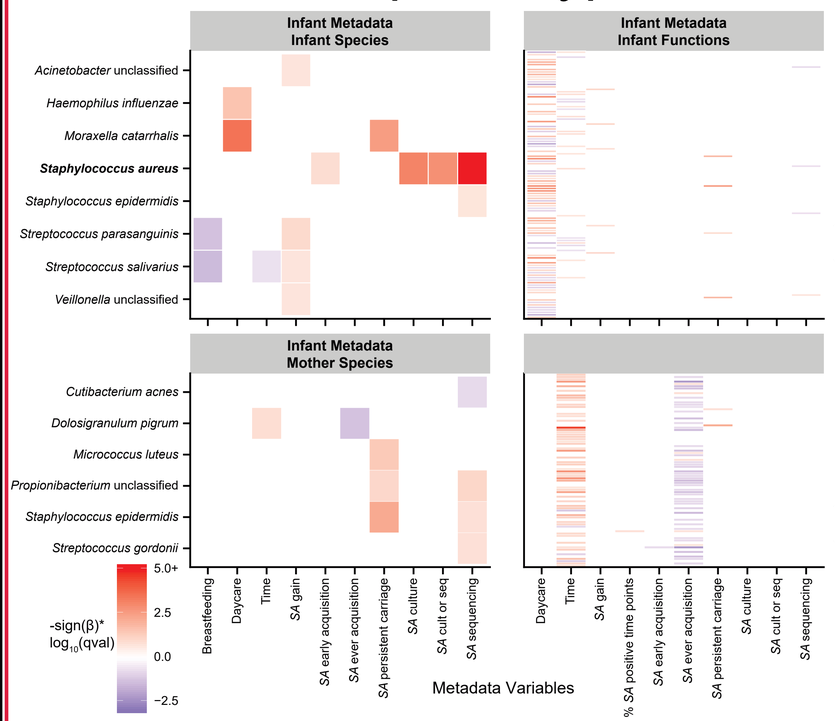
(C) Identification of *S. aureus* by culture and sequencing showed strong, although not complete, concordance.



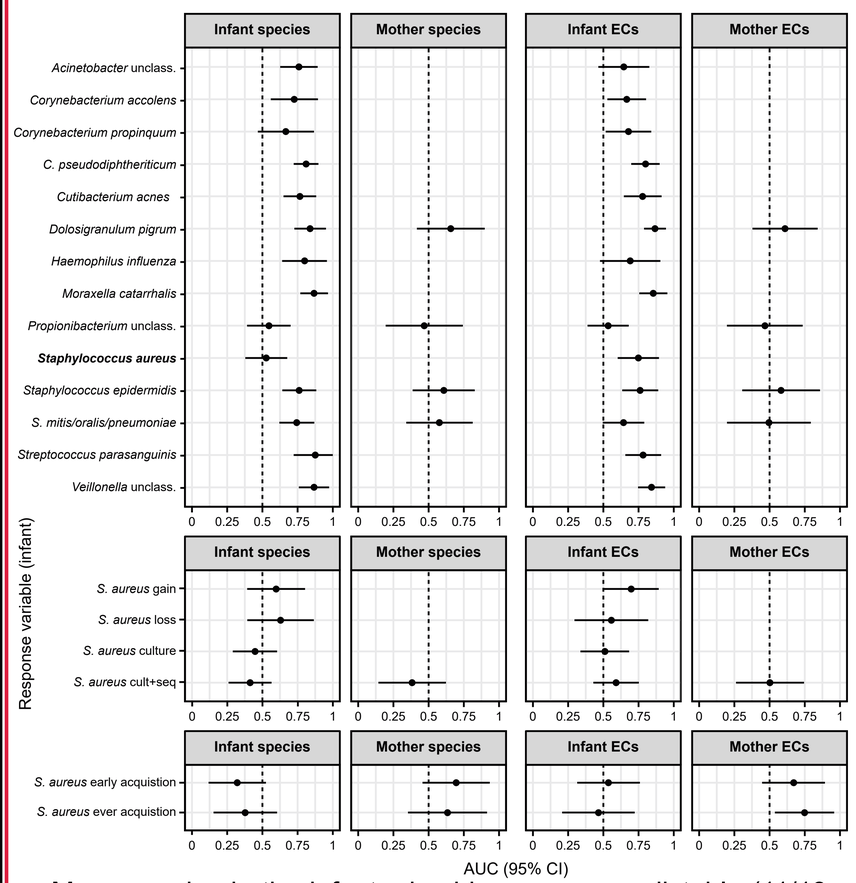




Microbiome drivers of infant *S. aureus* phenotypes



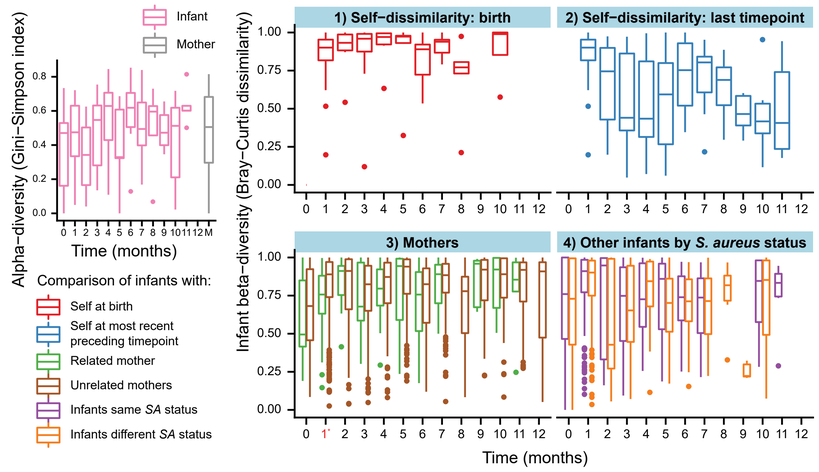
Infant *Acinetobacter* unclas., *S. parasanguinis*, *S. salivarius*, and *Veillonella* unclas. and maternal *D. pigrum* were significantly associated with infant *S. aureus* carriage.



Many species in the infant microbiome were predictable (11/13 using species, 9/13 using ECs), but *S. aureus* was consistently difficult to predict, although performance improved using ECs.

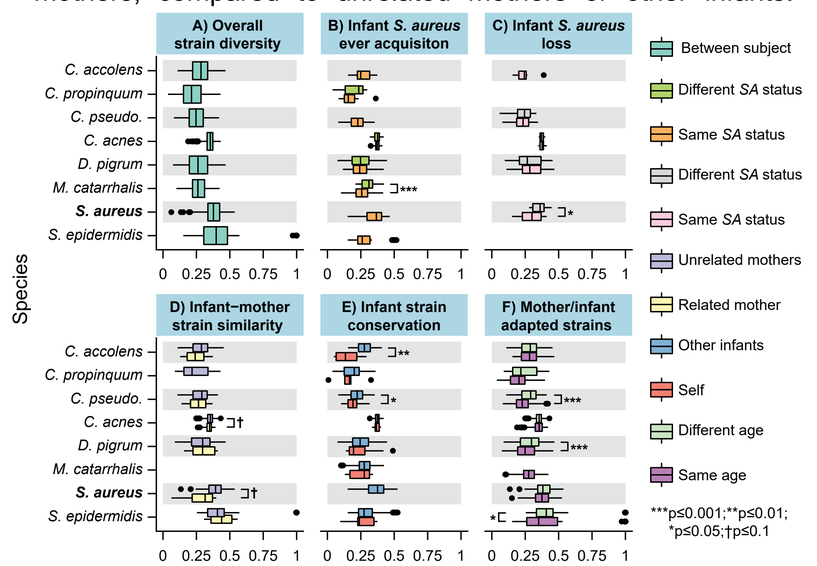
Infant nasal microbiomes mature over the first year, but remain distinct from mothers

Infants rapidly diverged from their species composition at birth, but the rate of change slowed over time indicating stabilization toward a more mature microbiome. Infants were more similar to their own mother than to unrelated mothers at month 1 (PERMANOVA, p=0.005), although infant composition was distinct from maternal composition at all months except 8 (p<0.05).



Strain genotypes show similarity in mother & infant *S. aureus*

Infant *S. aureus* strains were more similar to those of their own mothers, compared to unrelated mothers or other infants.



The authors express their gratitude to all participants, and to the NIH NIAID (grants R21AI112991 to CH and T32AI007535 to EA), the Chief Scientist, Ministry of Health, Israel (grant 3-00000-5622 to GRY), and the Israel Science Foundation (grants 1590/09, and 1658/15 to GRY).

http://huttenhower.sph.harvard.edu

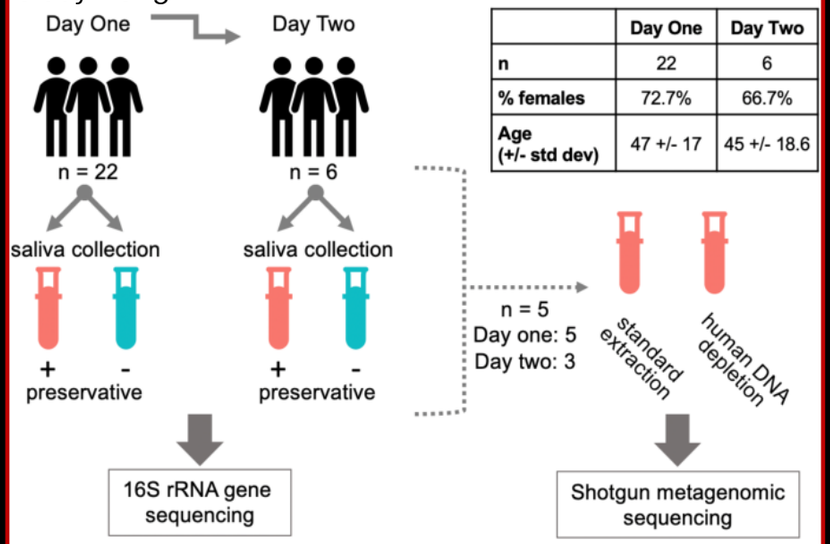
Assessing saliva microbiome collection and processing methods

Abigail JS Armstrong, Veenat Parmar, Martin J Blaser
Center for Advanced Biotechnology and Medicine, Rutgers the State University, Piscataway, NJ

Abstract

The oral microbiome has been connected with lung health and may be of significance in the progression of SARS-CoV-2 infection. Saliva-based SARS-CoV-2 tests provide the opportunity to leverage stored samples for measuring the oral microbiome. However, these collection kits have not been tested for accuracy of measuring the oral microbiome. Saliva is highly enriched with human DNA and reducing it prior to shotgun sequencing may increase the depth of bacterial reads. We examined both the effect of saliva collection method and sequence processing on measurement of microbiome depth and diversity by 16S and shotgun metagenomics. We collected 56 samples from 22 subjects. Each subject provided two saliva samples with and without preservative; 6 subjects provided a second set of samples the following day. 16S rRNA gene (V4) sequencing was performed on all of the samples, and shotgun metagenomics was performed on 8 of the samples collected with preservative with and without human DNA depletion before sequencing. We observed beta diversity distance within subjects over time was smaller than between unrelated subjects, and distances within subjects were smaller in samples collected with preservative. Samples collected with preservative had higher alpha diversity measuring both richness and evenness. Human DNA depletion before extraction and shotgun sequencing yielded higher total and relative reads mapping to bacterial sequencing. We conclude that collecting saliva with preservative may provide more consistent measures of the oral microbiome and that depleting human DNA increases yield.

Study Design



	Day One	Day Two
n	22	6
% females	72.7%	66.7%
Age (+/- std dev)	47 +/- 17	45 +/- 18.6

Figure 1. Saliva collection study design. Subjects provided 2 saliva samples: collected alone and collected with the Spectrum sDNA-1000 kit including preservative. The collection order was randomized. The 6 subjects provided samples the day following initial collection, using the same protocol as previous day. For metagenomic studies, we assessed the effects of a protocol to deplete human DNA, using only samples in which the original preservative was used (n=14 samples).

Results

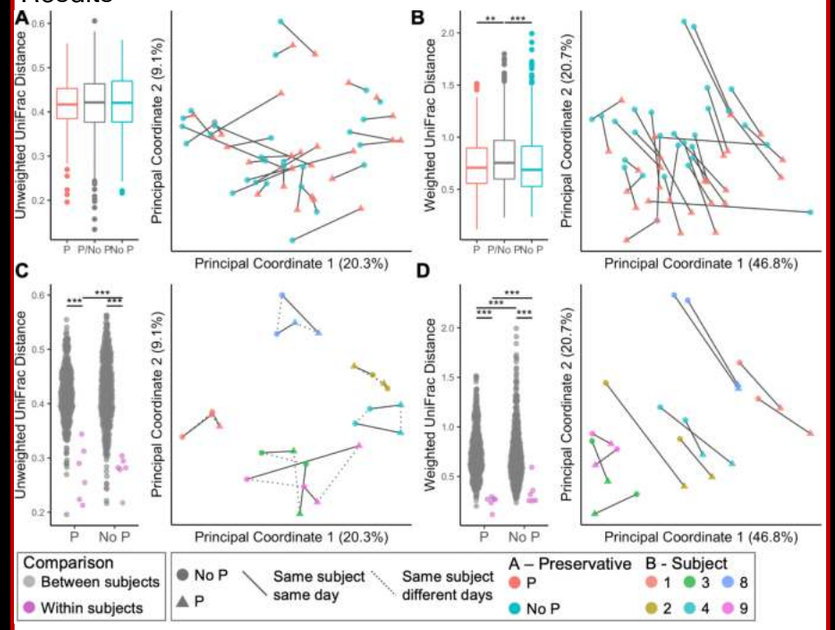


Figure 2. Phylogenetic beta diversity for 16S rRNA analyses of saliva samples by collection method and across time. **Top panels:** Unweighted **(A)** and weighted **(B)** UniFrac distance of all samples according to collection method. **Left panels:** Median (and IQR) distances in within-sample comparisons (P vs. P; No P vs No P), and across samples (P vs. No P). **Right panels:** PCoA plots of all samples by sample collection method. **Bottom panels:** Unweighted **(C)** and weighted **(D)** UniFrac distances of the paired specimens from 6 subjects sampled on two consecutive days and according to collection method. **Left panels:** Distance between unrelated subjects (gray) or within an individual across days (pink). **Right panels:** PCoA plot of all samples with only 6 multi-day subjects visualized. For all left panels, pairwise Wilcox test with FDR correction ** q < 0.01, *** q < 0.001. Lines connect specimens collected from the same subject on the same day (solid) or different days (dotted).

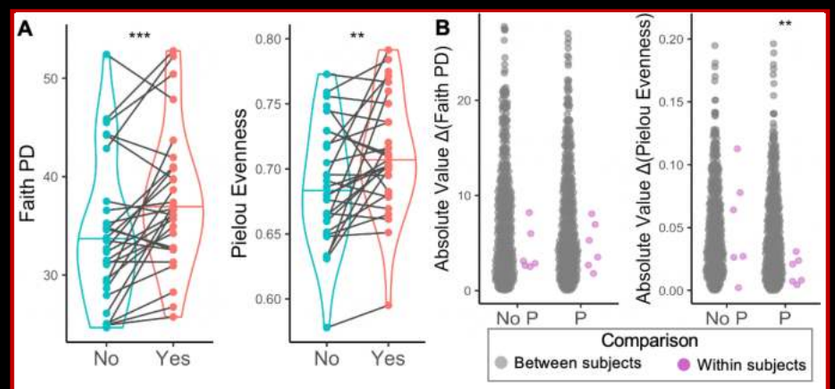


Figure 3. Alpha diversity of samples, by collection method and over time. **A.** alpha diversity measures based on Faith PD and Pielou evenness, by sample collection with preservative (P) or not (No P). Lines connecting points indicate sample pairs. ** p < 0.01, *** p < 0.001; linear mixed effects model. **B.** Absolute value of the differences in alpha diversity between all unrelated subjects (pink circles), and longitudinal samples within the same subject (gray circles). * p <0.05, ** p < 0.01; Kruskal-Wallis test.

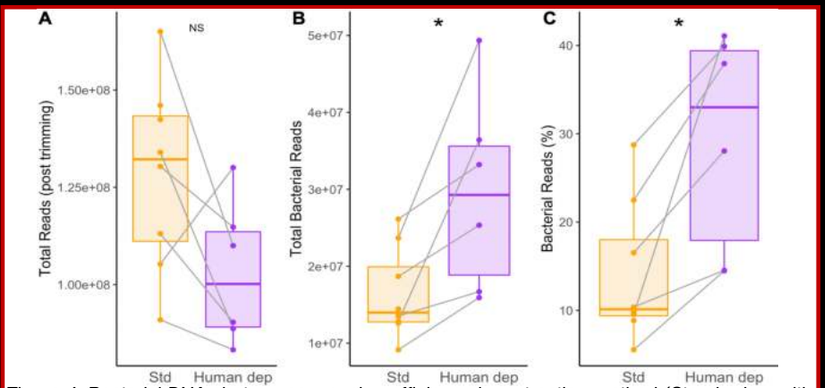


Figure 4. Bacterial DNA shotgun sequencing efficiency by extraction method (Standard or with depletion of human DNA). **A.** Total reads (post-trimming). **B.** Total bacterial reads. **C.** Bacterial reads as a percent of total reads; paired T-tests, * p < 0.05: NS = p > 0.05.

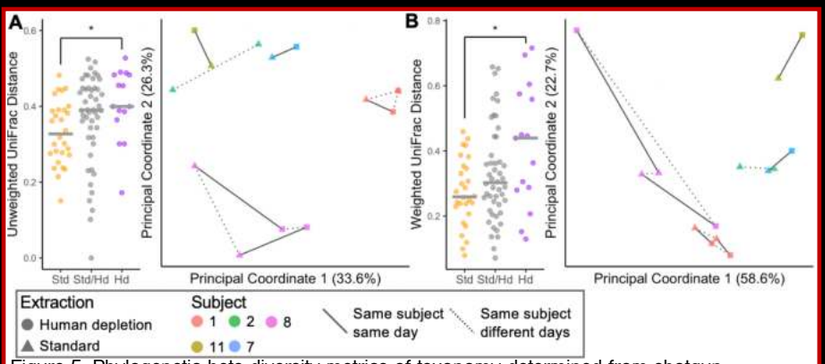


Figure 5. Phylogenetic beta diversity metrics of taxonomy determined from shotgun sequencing. Unweighted **(A)** and weighted **(B)** UniFrac of all samples according to DNA extraction method. **Left panels:** Median distances in within-sample comparisons (standard vs. standard; human depletion vs human depletion), and across samples (standard vs human depletion). **Right panels:** PCoA plots of all samples by extraction method; pairwise Wilcox test with FDR correction * q < 0.05. Lines connect specimens collected from the same subject on the same day (solid) or different days (dotted).

Conclusions

- Kits used to collect saliva for the purpose of SARS-CoV2 testing sufficiently preserve the microbiome DNA and are comparable to saliva collected without preservative.
- Preservative did not hinder human DNA depletion which increased bacteria DNA in shotgun sequencing.
- We found less variation within individuals over time compared to unrelated individuals, suggesting that longitudinal evaluation of subjects may provide valuable insights into oral microbiome changes.
- These results make it practical to use saliva samples obtained for SARS-CoV-2 testing to examine the salivary microbiome

Acknowledgements This work has been funded in part by Danone North America and National Institutes of Health U01AI2285.
Contact Abigail Armstrong, aa2253@cabm.rutgers.edu

Microbiota-Induced Vitamin A Mobilization by Serum Amyloid A and Its Role in Intestinal Immunity

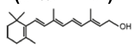
Ye-Ji Bang[1], Zehan Hu[1], Yun Li[1], Sureka Gattu[1], Kelly A. Ruhn[1], Joachim Herz[2,3,4,5], and Lora V. Hooper[1,6]

[1]Department of Immunology; [2]Department of Molecular Genetics; [3]Center for Translational Neurodegeneration Research; [4]Department of Neurology and Neurotherapeutics; [5]Department of Neuroscience, [6]The Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, TX 75390

Vitamin A and its derivative retinol are essential for the development of intestinal adaptive immunity. Retinoic acid (RA)-producing myeloid cells are central to this process, but how myeloid cells acquire retinol for enzymatic conversion to RA is unknown. Here, we show that serum amyloid A (SAA) proteins, retinol binding proteins induced in intestinal epithelial cells by the microbiota, deliver retinol to myeloid cells. We identify LDL receptor-related protein 1 (LRP1) as an SAA receptor that facilitates endocytosis of SAA-retinol complexes and promotes retinol acquisition by RA-producing intestinal myeloid cells. Consequently, SAA and LRP1 are essential for vitamin A-dependent immunity, including T and B cell homing to the intestine and immunoglobulin A production. Our findings identify a key mechanism underpinning vitamin A's effects on the immune system and provide insight into how the microbiota promotes intestinal immunity.

Background

Retinol (Vitamin A)



- **Important for the development of intestinal adaptive immunity**
  - Recruitment of lymphocytes to intestine
  - IgA production
- **Its transport requires retinol binding proteins** that shield it from the aqueous environment
- **Intestinal myeloid cells:** enzymatically convert retinol to its bioactive metabolite retinoic acid (RA)

→ **central to vitamin A-dependent immune regulation**

Unanswered Question

: How do intestinal myeloid cells acquire retinol to convert RA?

Serum Amyloid A (SAA)

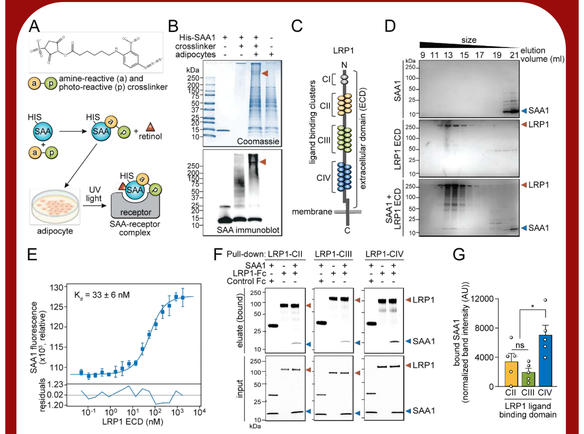
- **Retinol binding protein**
- **Expression in intestinal epithelium are induced by microbiota & vitamin A**

Hu and Bang *et al.* PNAS 2019

Hypothesis

: SAAs transport retinol into the intestinal immune cells and regulate vitamin A-dependent immune development

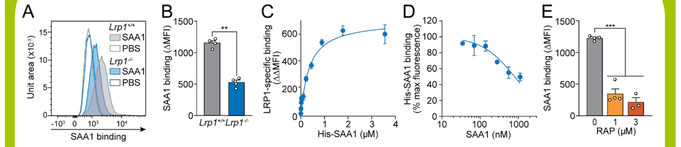
1 LRP1 is a cell surface receptor for SAAs



(A) Strategy to identify the SAA receptor. **(B)** Protein complexes were resolved by SDS-PAGE. Red arrowheads indicate the protein complex subjected to mass spectrometry for receptor identification. **(C)** Domain organization of LRP1 showing the ectodomain (ECD) and ligand binding clusters (CI-CIV). **(D)** Formation of a complex between SAA1 and LRP1-ECD confirmed by size exclusion chromatography. **(E)** Binding affinity between SAA1 and the LRP1-ECD as determined by microscale thermophoresis. **(F and G)** Binding of SAA1 to the LRP1 ligand binding clusters (LRP1-CII, LRP1-CIII, and LRP1-CIV) by pull-down assay.

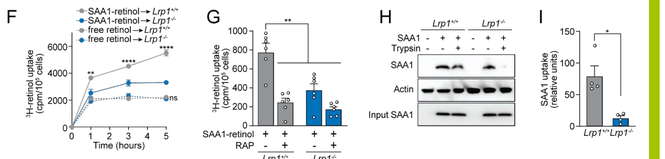
2 LRP1 mediates cell surface binding and cellular uptake of SAA1-retinol complexes

LRP1-dependent binding of SAA to cell surface



His-SAA1–retinol complex was incubated with fibroblasts (*Lrp1*$^{+/+}$ and *Lrp1*$^{-/-}$) at 4°C, and binding of SAA to cell surface is analyzed by flow cytometry. **(A,B)** Binding of SAA. **(C)** LRP1-specific SAA1 binding to the cell surface indicating a saturable binding. **(D)** Competitive binding assays by adding to the cells increasing concentrations of non-His tagged SAA1 together with His-SAA1. **(E)** Receptor-associated protein (RAP) was added to fibroblasts together with 1 μM of His-SAA1.

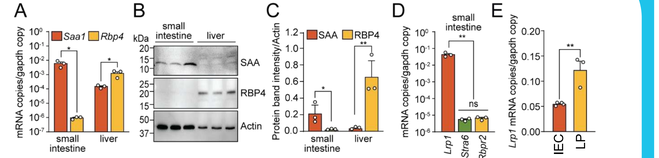
LRP1-dependent uptake of SAA-retinol complex



(F, G) Cellular ^{3}H-retinol uptake at 37ºC. **(F)** ^{3}H-retinol was added to the cells either as free retinol or in complex with SAA1. Counts per minute (cpm) in the cell pellets were determined and normalized to cell number. **(G)** RAP was added to cells together with SAA1–^{3}H-retinol complex for competition analyses. **(H, I)** Cellular SAA uptake. Cells were incubated with His-SAA1–retinol complex and subsequently treated with trypsin to remove surface-bound SAA1. Cells were lysed and cellular SAA uptake was analyzed by immunoblotting using an anti-His antibody.

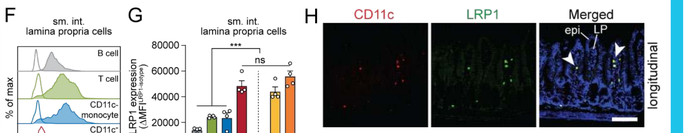
3 LRP1 is expressed on intestinal CD11c⁺ myeloid cells

SAA and LRP1 are the major retinol transporter-receptor in sm. Int.



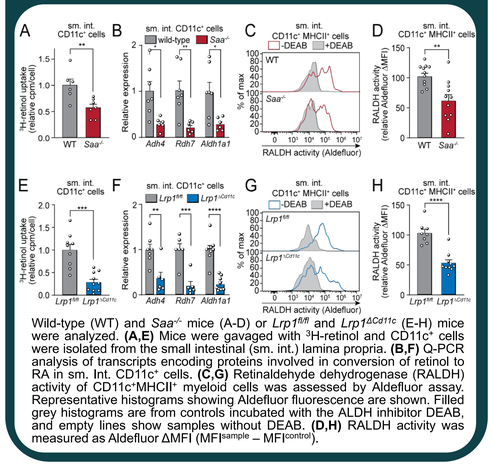
(A) Q-PCR analysis of *Saa1* and *Rbp4* transcripts in ileum and liver. **(B,C)** Immunoblotting of SAA and RBP4 in mouse ileum and liver. **(D)** Q-PCR analysis of *Lrp1*, *Stra6*, and *Rbpr2* transcripts in mouse ileum. **(E)** Q-PCR analysis of *Lrp1* transcripts in small intestinal epithelial cells (IEC) and immune cells isolated from the small intestinal lamina propria (LP).

LRP1 is expressed on intestinal CD11c⁺ myeloid cells



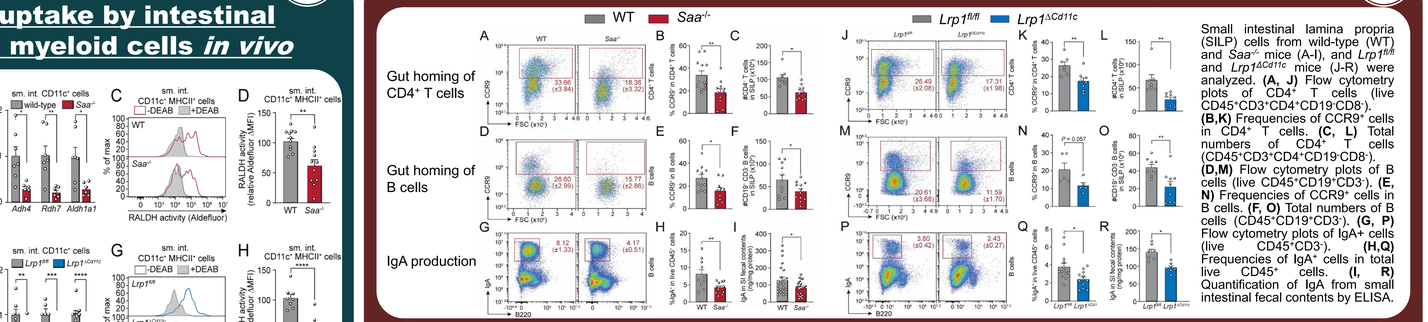
(F,G) Flow cytometry analysis of LRP1 expression on LP immune cells. **(F)** Representative histograms showing LRP1 expression. **(G)** LRP1 expression (ΔMFI) is shown as the difference in MFI between isotype-matched control and anti-LRP1 antibody staining (MFI$^{\text{LRP1}}$-MFI$^{\text{isotype}}$). **(H)** Immunofluorescence detection of CD11c (*red*), LRP1 (*green*), and nuclei (*blue*) in the mouse sm. Int.

4 SAA and LRP1 promote retinol uptake by intestinal CD11c⁺ myeloid cells *in vivo*



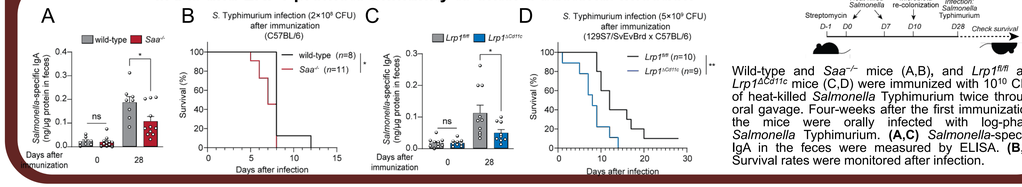
Wild-type (WT) and *Saa*$^{-/-}$ mice (A-D) or *Lrp1*$^{fl/fl}$ and *Lrp1*$^{\Delta Cd11c}$ (E-H) mice were analyzed. **(A,E)** Mice were gavaged with ^{3}H-retinol and CD11c$^+$ cells were isolated from the small intestinal (sm. int.) lamina propria. **(B,F)** Q-PCR analysis of transcripts encoding proteins involved in conversion of retinol to RA in sm. Int. CD11c$^+$ cells. **(C,G)** Retinaldehyde dehydrogenase (RALDH) activity of CD11c$^+$MHCII$^+$ myeloid cells was assessed by Aldefluor assay. Representative histograms showing Aldefluor fluorescence are shown. Filled grey histograms are from controls incubated with the ALDH inhibitor DEAB, and empty lines show samples without DEAB. **(D,H)** RALDH activity was measured as Aldefluor ΔMFI (MFI$^{\text{sample}}$ – MFI$^{\text{control}}$).

5 SAA and LRP1 promote vitamin A-dependent immunity in intestine

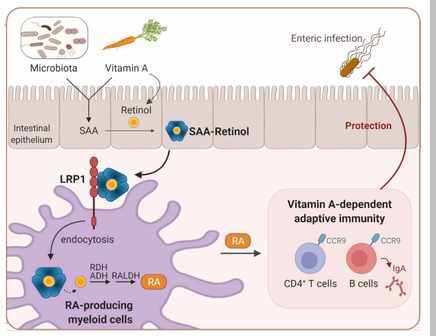


Small intestinal lamina propria (SILP) cells from wild-type (WT) and *Saa*$^{-/-}$ mice (A-I), and *Lrp1*$^{fl/fl}$ and *Lrp1*$^{\Delta Cd11c}$ mice (J-R) were analyzed. **(A, J)** Flow cytometry plots of CD4$^+$ T cells (live (CD45$^+$CD3$^+$CD4$^+$CD19$^-$CD8$^-$). **(B,K)** Frequencies of CCR9$^+$ cells in CD4$^+$ T cells. **(C, L)** Total numbers of CD4$^+$ T cells (CD45$^+$CD3$^+$CD4$^+$CD19$^-$CD8$^-$). **(D,M)** Flow cytometry plots of B cells (live CD45$^+$CD19$^+$CD3$^-$). **(E, N)** Frequencies of CCR9$^+$ cells in B cells. **(F, O)** Total numbers of B cells (CD45$^+$CD19$^+$CD3$^-$). **(G, P)** Flow cytometry plots of IgA+ cells (live CD45$^+$CD3$^-$). **(H,Q)** Frequencies of IgA$^+$ cells in total live CD45$^+$ cells. **(I, R)** Quantification of IgA from small intestinal fecal contents by ELISA.

SAAs and LRP1 promote immunity to enteric bacterial infection



Wild-type and *Saa*$^{-/-}$ mice (A,B), and *Lrp1*$^{fl/fl}$ and *Lrp1*$^{\Delta Cd11c}$ mice (C,D) were immunized with 10^{10} CFU of heat-killed *Salmonella* Typhimurium twice through oral gavage. Four-weeks after the first immunization, the mice were orally infected with log-phase *Salmonella* Typhimurium. **(A,C)** *Salmonella*-specific IgA in the feces were measured by ELISA. **(B,D)** Survival rates were monitored after infection.



SAA delivers retinol to myeloid cells by binding to LRP1 and thus promotes vitamin A-dependent adaptive immunity in the intestine. SAA delivers retinol to intestinal CD11c$^+$ myeloid cells by binding to LRP1. LRP1 facilitates endocytosis of SAA-retinol complexes, and the retinol is converted to retinoic acid (RA) through a two-step enzymatic reaction. Myeloid cell RA promotes vitamin A-dependent intestinal adaptive immune responses. These include the induction of intestinal homing receptors, such as CCR9, on CD4$^+$ T cells and B cells, and IgA production by B cells. (Illustration was created in BioRender.)

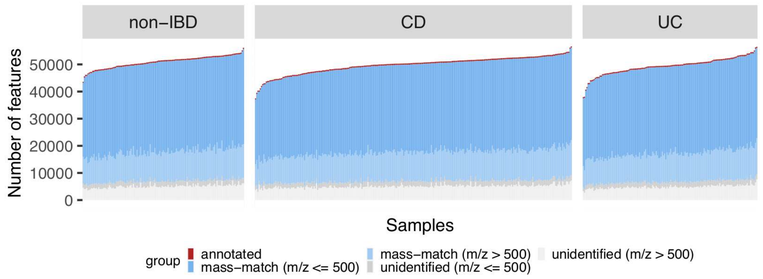
Prioritization and annotation of novel bioactive small molecules from the microbiome

Amrisha Bhosle[1,2], Sena Bae[3], Yancong Zhang[1,2], Eunyong Chun[3], Julian Avila-Pacheco[1], Ludwig Geistlinger[4], Levi Waldron[4], Clary Clish[1], Ramnik Xavier[1], Hera Vlamakis[1], Eric A. Franzosa[1,2], Wendy S. Garrett[1,3], Curtis Huttenhower[1,2,3]

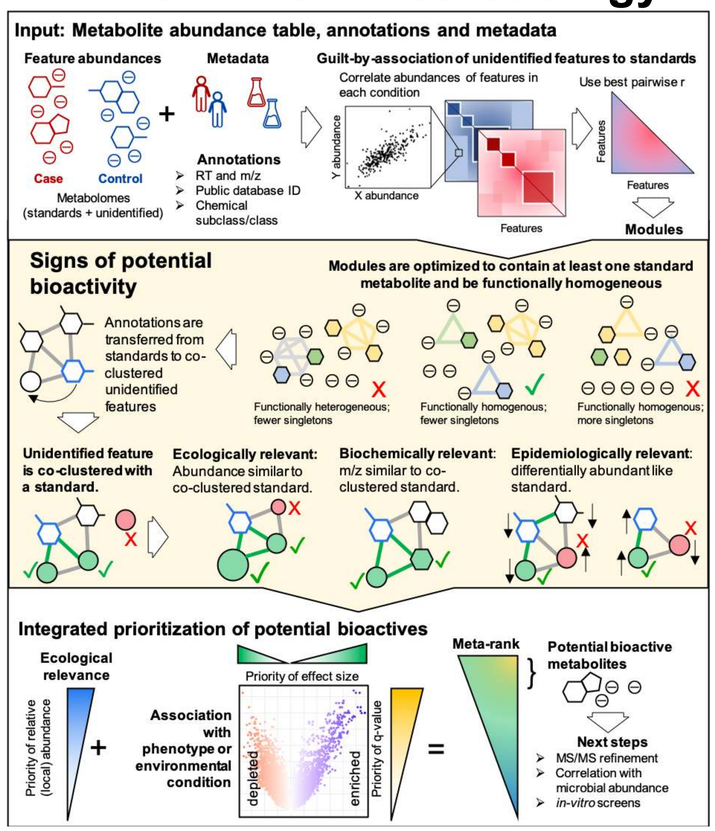
[1]Broad Institute of MIT and Harvard, [2]Department of Biostatistics, Harvard T.H. Chan School of Public Health,
[3]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, [4]Department of Epidemiology and Biostatistics, City University of New York

Chemical dark matter in IBD

Thousands of metabolites have been assayed from microbial communities, the gut microbiome in particular, but as yet with minimal biochemical characterization or knowledge of their therapeutic potential. Here, we developed a new approach, MACARRoN (**M**etabolome **A**nalysis and **C**ombined **A**nnotation **R**anks for p**R**ediction **o**f **N**ovel bioactives), for identifying potential bioactives by integrating knowledge of annotated/standard compounds with phenotypic or environmental indicators of bioactivity to annotate and prioritize small molecules from microbiomes. We have applied this approach to identify novel bioactives from the inflammatory bowel disease (IBD) metabolomes in the Integrative Human Microbiome Project (HMP2).

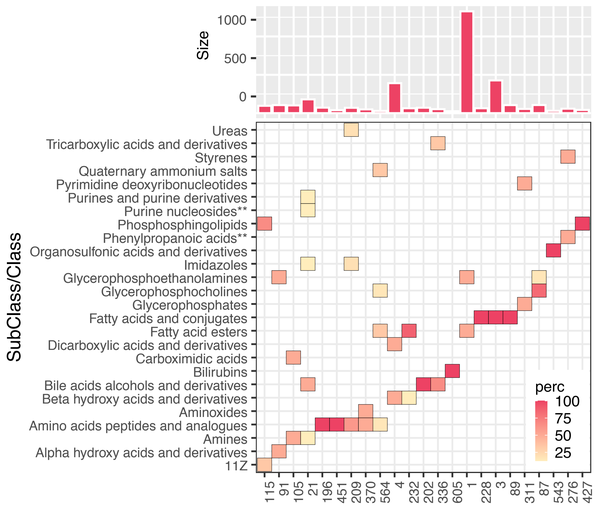


MACARRoN methodology



Quantitative metabolite annotations

To associate unannotated compounds with chemically annotated metabolites i.e. standards, we clustered features based on co-varying abundances in the different phenotypes. Of the 44,757 high-quality features, 43,498 features were distributed into 606 modules whereas 1,259 were singletons.

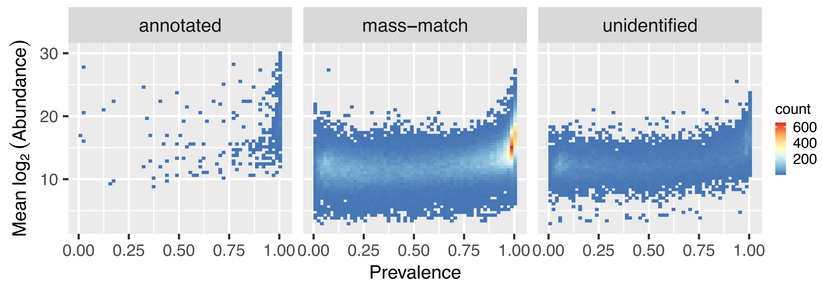


Modules of different sizes were obtained. 153 modules contained at least one standard.

Modules were broadly functionally homogeneous and contained chemical subclasses/classes** which co-vary as a result of co-occurence in a biochemical pathway or co-synthesis by a microbe.

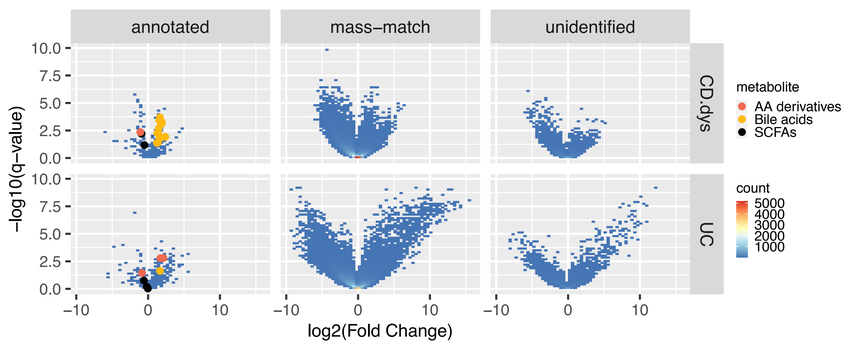
Prioritizing features based on ecological importance

Unannotated (mass-match and unidentified) features are as abundant and prevalent as annotated features. We prioritize features that are similarly abundant as compared to a co-clustered annotated feature.



Prioritizing features on strength and significance of IBD association

Unannotated features are differentially abundant in dysbiotic CD and UC with effect size similar to IBD-linked molecules such as bile acids and SCFAs.



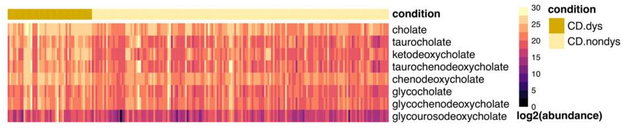
Prioritized metabolites

Metabolites highly prioritized by MACARRoN include classes previously implicated in IBD as well as novel potential bioactives

Enriched metabolites

Primary bile acids

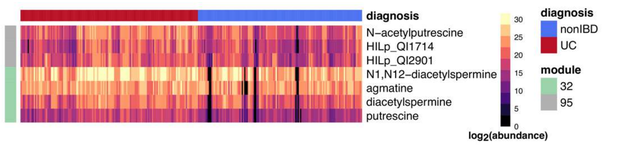
Accumulation of primary bile acids due to loss of bacterial diversity is well understood in IBD.



dys: dysbiosis (active disease); nondys: non-dysbiosis

Putrescine derivatives

Gut bacteria encode enzymes that synthesize and degrade polyamines: putrescine and spermidine. (Nakamura, 2019)



Depleted metabolites

Short chain fatty acids

SCFAs are produced via bacterial fermentation of dietary fiber and are typically found to be reduced in IBD.

Medium chain fatty acids and B vitamins

Production of MCFAs has been attributed to bacterial thioesterases.

Conclusions and future work

1. MACARRoN integrates ecological, biochemical and epidemiological annotations to prioritize metabolites in the microbiome.
2. Novel highly prioritized compounds covary with known metabolites, have a high relative abundance and are significantly differentially abundant in dysbiosis and IBD.
3. MACARRoN prioritizes known IBD-linked classes such as bile acids and SCFAs which serves to validate the workflow.
4. It also prioritizes lesser-understood classes such as putrescine derivatives, medium chain fatty acids and B vitamins whose roles in IBD will require further study.
5. MACARRoN is generalizable to microbial community metabolomes and is being developed as an open-source R package.

Acknowledgements

This work has been funded by NIH NIDDK grant R24K110499.

@hutlab

https://huttenhower.sph.harvard.edu
https://huttenhower.sph.harvard.edu/MACARRoN

Gut Microbial Enzymes Drive the Dose-Limiting Toxicity of the Immunosuppressant Mycophenolate Mofetil

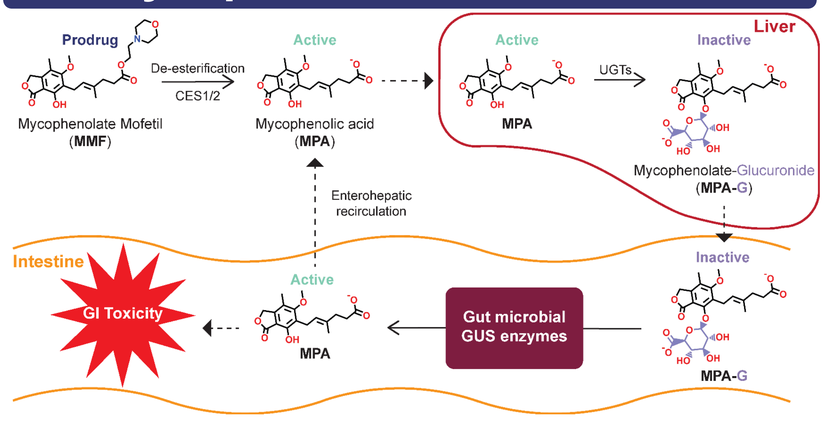
Marissa M. Bivins[1], Lindsay E. Bass[2], Amanda L. Graboski[1], Michelle E. Fiamingo[2], Rebecca L. Johnson[3], Joshua B. Simpson[2], William G. Walton[2], Nathan I. Nicely[1], John R. Lee[7,8], Matthew R. Redinbo[2,4,5,6]

Departments of Pharmacology[1], Chemistry[2], Chemical Biology & Medicinal Chemistry[3], Biochemistry & Biophysics[4], Microbiology & Immunology[5], and the Integrated Program from Biological & Genome Sciences[6], University of North Carolina, Chapel Hill, NC
Division of Nephrology & Hypertension[7], Department of Medicine, Weill Cornell Medicine, New York, NY
Department of Transplantation Medicine[8], New York Presbyterian Hospital - Weill Cornell Medical Center, New York, NY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

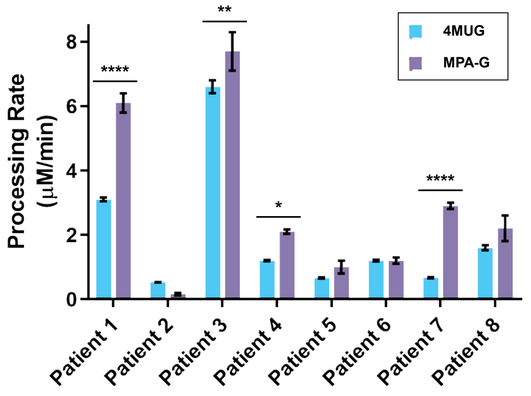
Redinbo Lab
in fimo we trust

Mycophenolate Metabolism



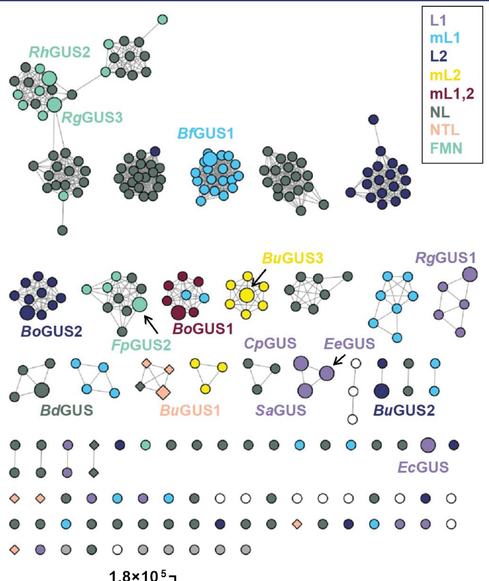
Goal: Which GUSs reactivate MPA-G?

In fimo GUS Activity

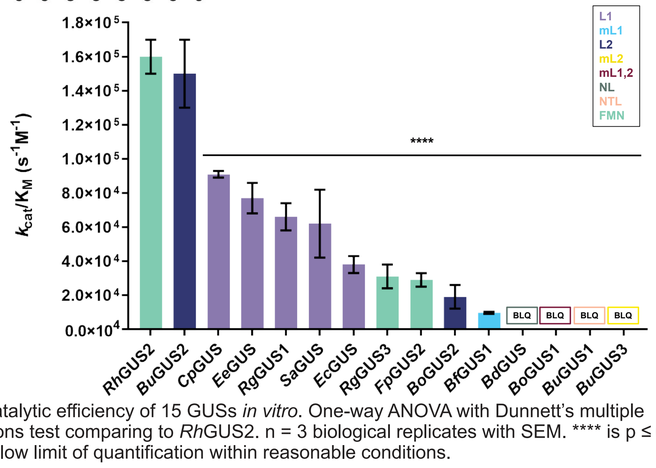


Microbial β-glucuronidase (GUS) activity was measured in purified fecal slurry (*in fimo*) with reporter substrate 4MUG and MPA-G. n = 3 biological replicates with SEM. Two-way ANOVA with Sidak's multiple comparisons test comparing one individual's sample rates where **** is $p \le 0.0001$, ** is $p \le 0.01$, * is $p \le 0.05$.

In vitro MPA-G Reactivation

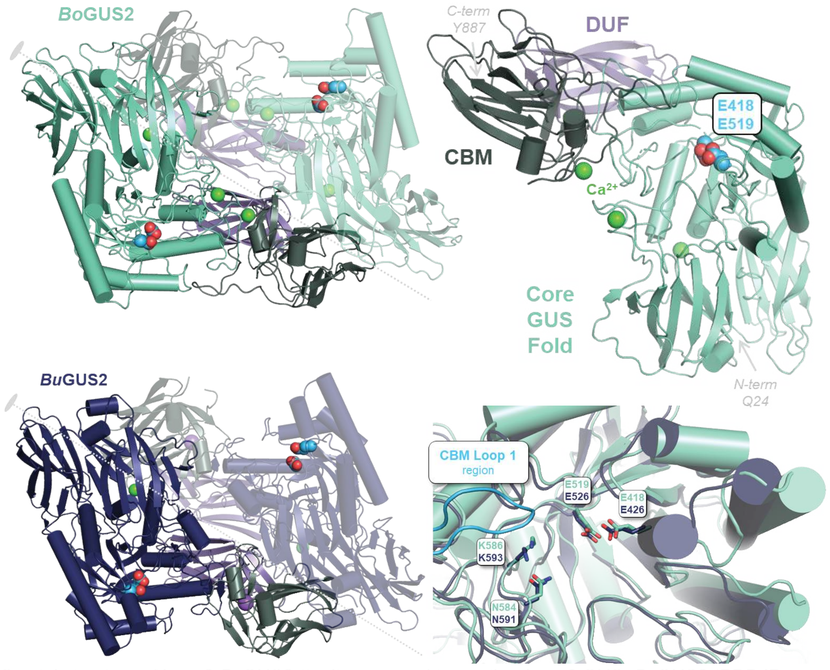


A sequence similarity network of 265 GUSs and β-galacturonidases (GalAses) found in the Human Microbiome Project fecal samples sequences. Colors indicate different loop classes of enzymes, with white denoting GalAses and gray are unclear in loop region. Diamonds indicate GUS/GalAse hybrids and large icons are investigated further.

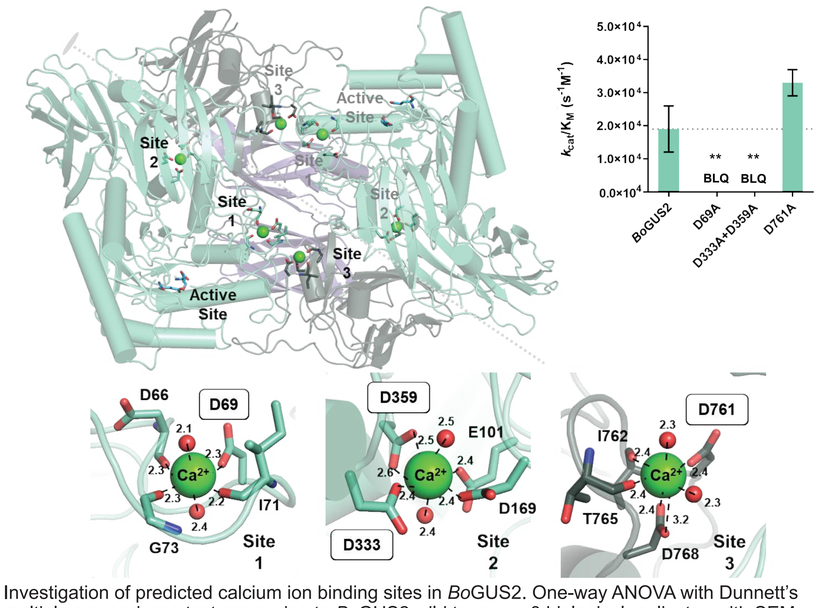


MPA-G catalytic efficiency of 15 GUSs *in vitro*. One-way ANOVA with Dunnett's multiple comparisons test comparing to *Rh*GUS2. n = 3 biological replicates with SEM. **** is $p \le 0.0001$. BLQ is below limit of quantification within reasonable conditions.

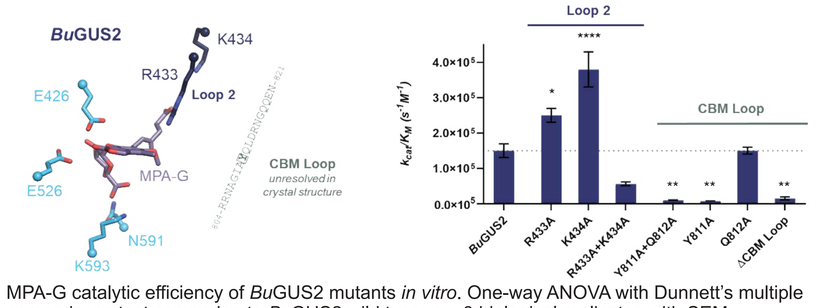
Structure Guided Analysis



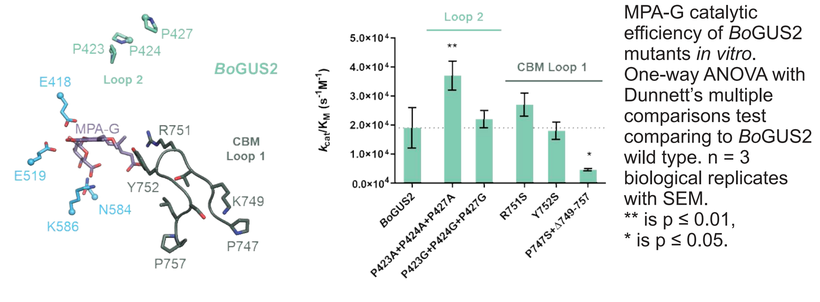
Crystal structure of loop 2 *Bo*GUS2 (teal) compared to another loop 2 GUS, *Bu*GUS2 (PDB ID: 5UJ6, navy). Carbohydrate binding modules (CBM) are in dark green and domain of unknown functions (DUF) are in purple.



Investigation of predicted calcium ion binding sites in *Bo*GUS2. One-way ANOVA with Dunnett's multiple comparisons test comparing to *Bo*GUS2 wild type. n = 3 biological replicates with SEM. ** is $p \le 0.01$. BLQ is below limit of quantification within reasonable conditions.

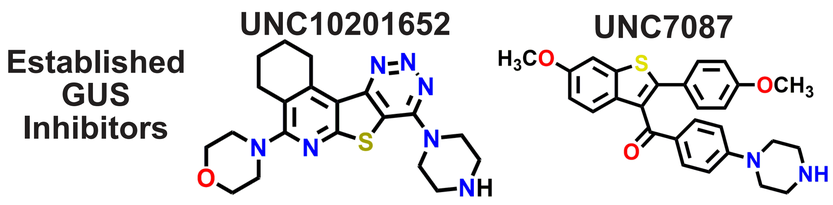


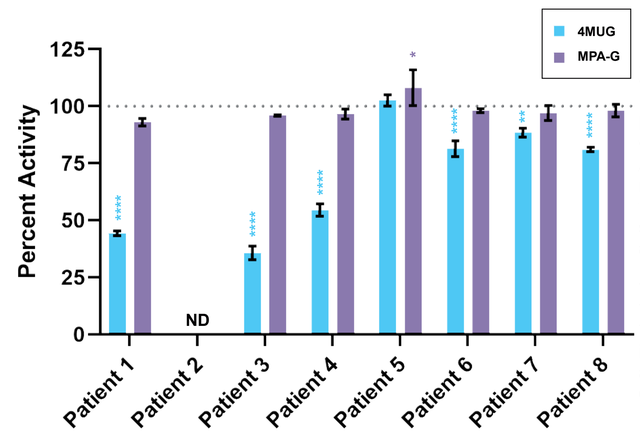
MPA-G catalytic efficiency of *Bu*GUS2 mutants *in vitro*. One-way ANOVA with Dunnett's multiple comparisons test comparing to *Bu*GUS2 wild type. n = 3 biological replicates with SEM. **** is $p \le 0.0001$, ** is $p \le 0.01$, * is $p \le 0.05$.



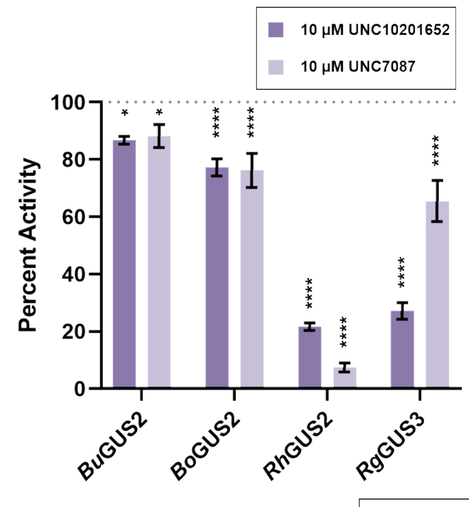
MPA-G catalytic efficiency of *Bo*GUS2 mutants *in vitro*. One-way ANOVA with Dunnett's multiple comparisons test comparing to *Bo*GUS2 wild type. n = 3 biological replicates with SEM. ** is $p \le 0.01$, * is $p \le 0.05$.

Impact of Inhibitors

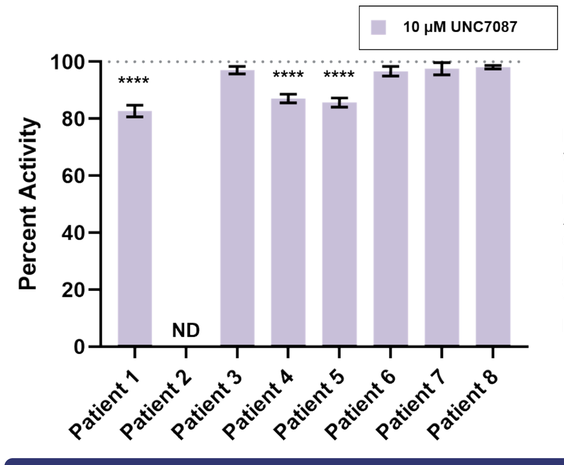




In fimo percent activity with 10 μM UNC10201652. n = 3 biological replicates with SEM. Two-way ANOVA with Dunnett's multiple comparisons test comparing sample percent activity with inhibitor to 100% where **** is $p \le 0.0001$, ** is $p \le 0.01$, * is $p \le 0.05$. ND indicates percent activity not determined.



In vitro percent activity with MPA-G by four GUS isoforms. n = 3 biological replicates with SEM. Two-way ANOVA with Dunnett's multiple comparisons test comparing percent activity to 100% for each GUS, where **** is $p \le 0.0001$, * is $p \le 0.05$.



MPA-G percent activity *in fimo* when incubated with 10 μM UNC7087. n = 3 biological replicates with SEM. Two-way ANOVA with Sidak's multiple comparisons test comparing percent activity to 100% for each patient sample, where **** is $p \le 0.0001$. ND indicates percent activity not determined.

Conclusions

- **GUS enzymes in fecal microbiota can reactivate MPA-G, but are not inhibited by previously established inhibitors.**
- **A wider range of GUS subtypes can reactivate MPA-G compared to other drug-glucuronides.**
- **GUS active site loop composition and C-terminal domain motifs are important for MPA-G processing.**

Acknowledgments

The rest of the Redinbo Lab: Parth Jariwala, Morgan Walker, Kacey Davis, Mark Kowalewski, and Adam Lietzan, along with many undergrads and rotation students.

This work is supported by NIH grants #GM135218 and #GM137286, UNC Biological and Biomedical Sciences Program, UNC Pharmacology Department T32 training grant, UNC Eshelman Institute for Innovation Discovery grant.

Contact email: Marissa Bivins (bivinsma@email.unc.edu); Matt Redinbo (redinbo@unc.edu)

UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCE, CENTER FOR BASIC METABOLIC RESEARCH

Discovering novel antibiotics through microbiome metabolome integration

Brejnrod A[1,2], Qing Fang[1], Manimozhiyan Arumugam[1], Pieter Dorrestein[2]

1: Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark
2: Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, United States

Summary of collected datasets

Dataset$	Reference	# Paired samples	Microbiome type	Species/ASVs	# Identified metabolites	# Metabolite MS2
Fermented foods	Taylor et al, 2019	79	Metagenome	1046	225	3992
Agp3k	Unpublished	1810	16S	3682	371	3367
Agp500	Mcdonald et al, 2018	482	Metagenome		531	10215
CRC (Healthy subset)	Yachida et al, 2019	127	Metagenome	623	450	NA
Franzosa	Lloyd-Price et al, 2019	220	Metagenome	201	467	NA

Table 1: Summary of collected datasets

Candidate's that are novel and known in the literature selected

Metabolite	Dataset	Producers	Bacterial prc	Literature
oleanolic acid.1	Franzosa	15	Yes	5
xylose	Franzosa	1	Plausibly	4
deoxyinosine.1	Franzosa	1	Yes	4
Coreg CR	AGP3k amplicon	1	No	3
Salpha-cholestanol	Franzosa	5	Yes	3
C20 carnitine	Franzosa	1	Yes	3
Tomatidine__3791_416.351841685994_4.41547568728522	FF	1	No	3
(E)-1-(2,4-dihydroxy-5-(3-methylbut-2-enyl)phenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	AGP3k amplicon	1	No	2
Chenodeoxycholyltyrosine	AGP3k amplicon	1		2
cis-14-hydroxy-10,13-dioxo-7-heptadecenoic acid	AGP3k amplicon	1	Yes	2
2-hydroxymyristic acid	Franzosa	1	Yes	2
imidazolelactate	Franzosa	1	Yes	2
pyridoxamine	Franzosa	11	Yes	2
suberate	Franzosa	2	No	2
cholestenone.2	Franzosa	2	Yes	2
geranyl acetoacetate*.1	Franzosa	1	No	2
(2S)-2-[[(2S)-2-acetamido-4-phenylbutanoyl]amino]-4-methyl-N-[2-[[(E,3S)-5-methyl-1-methylsu	AGP3k amplicon	1	No	1
F4-Neuroprostane (4-series)	AGP3k amplicon	1	No	1
Norhyoscyamine	AGP3k amplicon	1	No	1
tenacibactin A	AGP3k amplicon	2	Yes	1
4-hydroxy-3-methylacetophenone	Franzosa	1	Plausibly	1
C08306_Hexylamine	CRC healthy	1		
guanine	Franzosa	1	Yes	

Table 2: Candidate Evaluation**.** Significant mediation scores from all producers were selected and grouped to form a list of candidates for evaluation. Evaluation was done on the basis of literature score (explained below), a safety evaluation and considerations of solubility.
Literature scores: 1: No relevant literature, 2: Killing effect of a similar drug on some organisms, 3: antibacterial effect of this drug on some bacteria 4: Antibacterial effect of this drug on 5: Enterococcus Antibacterial effect of this drug on VRE

Identifying of metabolites mediating correlations

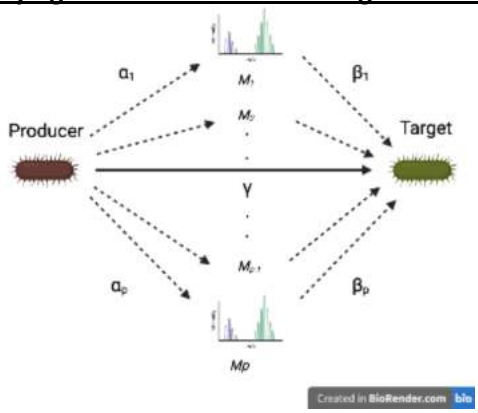


Figure 1: Schematic diagram of feature selection. For each putative producer-target relationsship a regularized linear model is fitted to identify spectra (M1 … Mp) that might explain the negative association between the bacteria. The models partitions the variance into the direct effect (gamma), the correlations between producer and spectra (alpha) and negative correlations between spectra and target (beta)

Candidate compounds show antibiotic effect *in vitro*

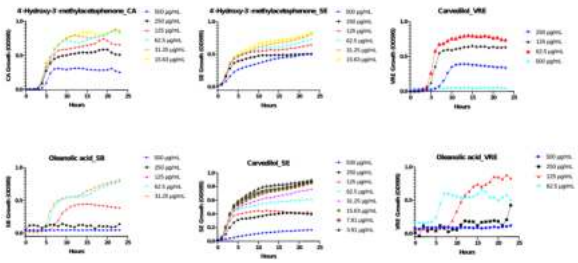


Figure 2: Growth curves of clinical relevant isolates exposed to tested compounds in different concentrationto determine MIC values of the tested compounds.

Candidate display activity against a panel of pathogens

	Minimum Inhibitory Concentration (µg/mL)						
Strain	4'-Hydroxy-3'-methylacetophenone	5-α-Cholestane	Carvedilol	Oleanolic acid	VAN	AMP	FDX
Enterococcus faecium DSM 13590(VRE)	69.57	>500	74.5	62.7	390.6	0.193	<0.098
Clostridioides difficile DSM 27543	>500	>500	>500	64.4	781.3	<9.76	0.39
C. albicans ATCC 10231	17.58	>500	163.1	98.53	1250	<9.76	>50
Escherichia coli ATCC 25922	>500	>500	>500	>500	630	4.08	50
Enterococcus faecalis ATCC 29212	471.0	>500	130.0	70.8	<9.76	<9.76	>50
Pseudomonas aeruginosa ATCC 27853	267.9	>500	214.6	>500	<9.76	0.56	1.56
Staphylococcus aureus ATCC 25923	316.4	>500	206.3	>500	<9.76	<9.76	1.56
Streptococcus B. ATCC 12386	140.4	>500	61.71	32.7	<9.76	<9.76	0.78
Staphylococcus epidermidis ATCC 35984	17.33	>500	7.76	153.1	<9.76	3.49	<0.98
Staphylococcus haemolyticus clinical isolate 8-7A	148.5	>500	>500	>500	625	3.27	>50

Table 3: Final MIC determination of a panel of bacteria. Numbers in red is moderate or better activity against the target organism. 3 of the tested compounds showed moderate activity against VRE. Oleanoloci acid displayed activity against several organisms.

Contact
Email: abrenjrod@health.ucsd.edu
Twitter: @askerbrejnrod

16S Sequencing in Pediatric Blood Detects DNA Signatures of Commensal and Pathogenic Microbes that Correlate with Subject's Medical History

Matthew Brock[1,2], Bo Zhang[1,2], Patricia Pichilingue-Reto[1], Carlos Arana[1,2], Lora Hooper[1] Nicolai S.C. van Oers[1] and Prithvi Raj[1,2]

UT Southwestern Medical Center

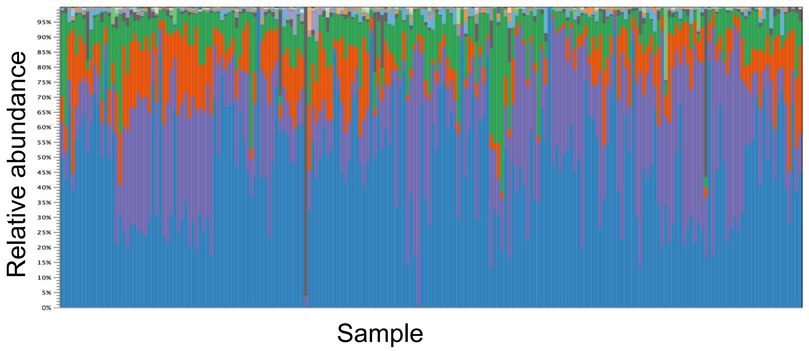
[1]Department of Immunology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[2]Microbiome Research Laboratory (MRL) at UT Southwestern Medical Center, Dallas, TX, 75235, USA

Contact: prithvi.raj@utsouthwestern.edu
matthew.brock@utsouthwestern.edu

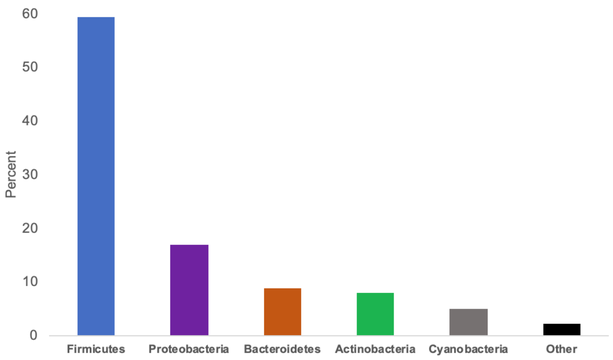
1. Abstract

We performed 16S rRNA sequencing on blood/serum specimens from 147 apparently healthy 1- and 2-year-old children to explore potential microbial signatures in circulation. Interestingly, 16S data detected some common pediatric pathogens in 6 out of 147 children. These pathogens were *Staphylococcus*, *Streptococcus, Haemophilus* and *Deinococcus*. Additionally, our sequencing assay detected DNA signatures of several commensal bacteria such as *Firmicutes*, *Bacteroides* and *Proteobacteria*. 16S data stratified samples into four major clusters differing in microbial composition. Cluster 1 was dominated by *Proteobacteria*. Cluster 2 was present in almost 50% of the samples, which comprised of 60-70% *Firmicutes*. Cluster 3 was mix in composition and presented in 18% of the samples. Cluster 4 was dominated by *Actinobacteria* and represented only 4% of the samples. Interestingly, Cluster 1 was found to be significantly (p=0.002) associated with higher BMI in children. To investigate the potential source of microbial signatures within the blood we studied resident microbiota in stool, skin and blood of germ-free and conventional mice. Our data suggest that 16S sequencing assay can rapidly detect microbial DNA signatures of commensal and pathogenic species within the blood to assist with diagnosing infectious diseases in children.

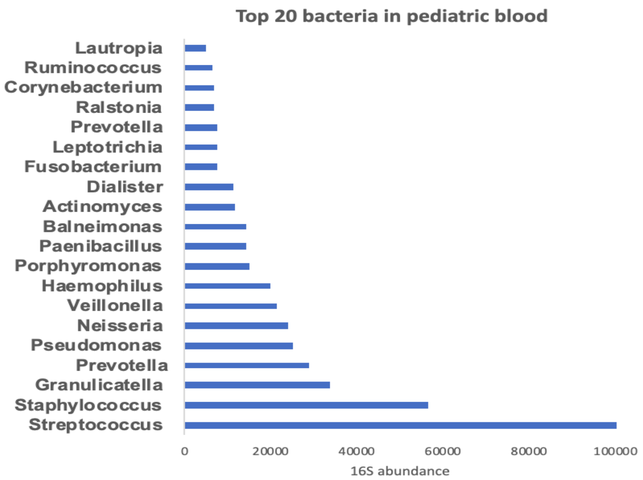
2. 16S Sequencing Results



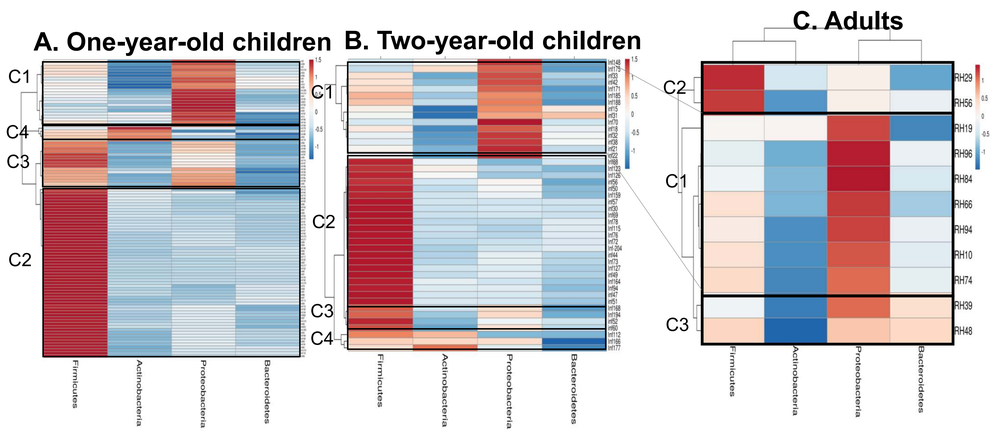
3. Phylum Level OTUs



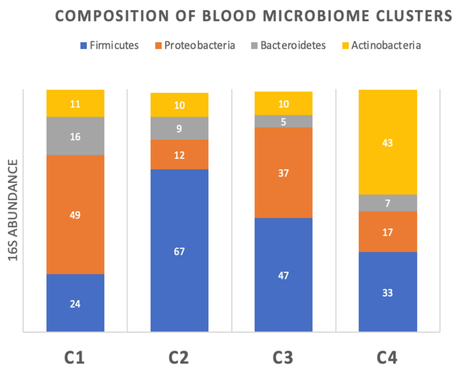
4. Genus Level OTUs



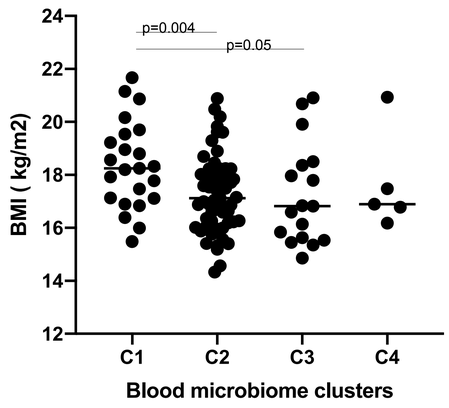
5. Blood Microbial DNA Clusters in Children and Adults



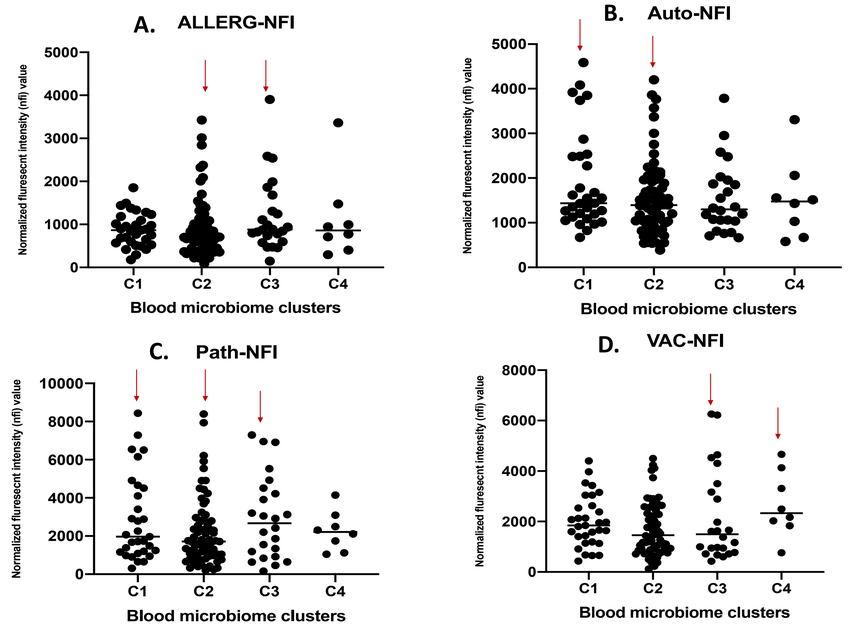
6. Composition of Clusters



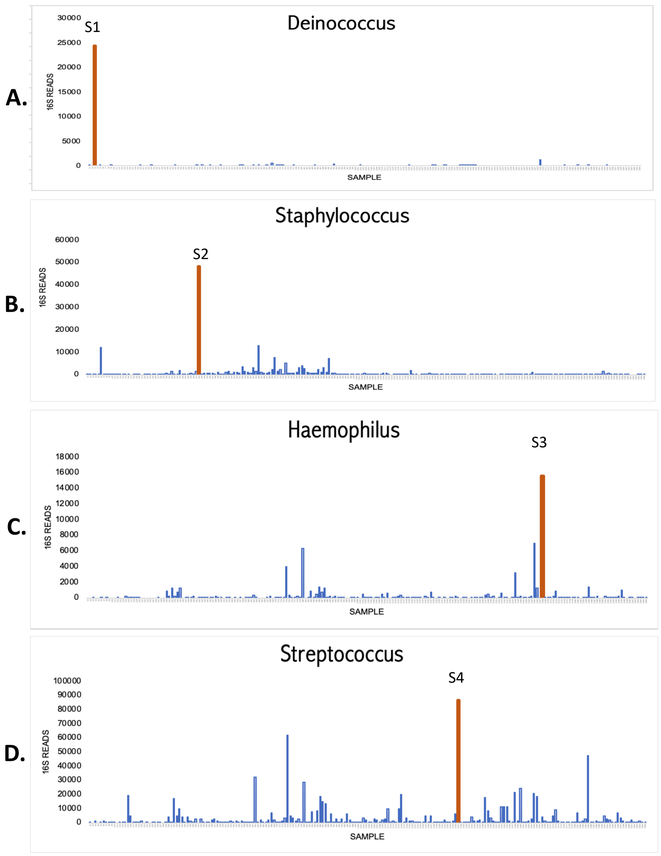
7. Association of Clusters with BMI in Children



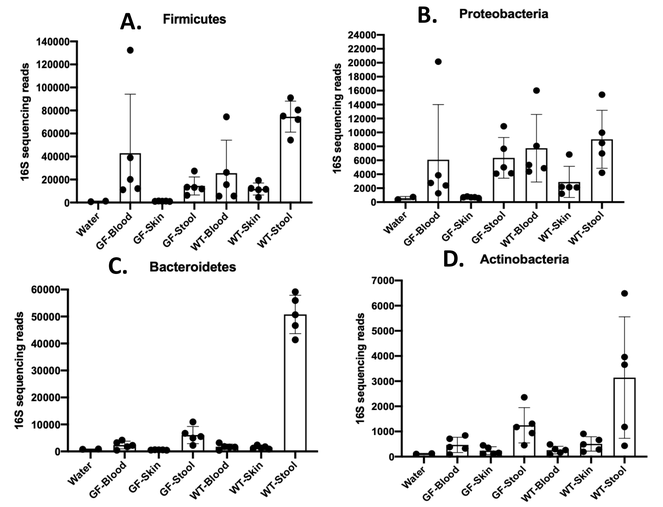
8. Microbial DNA Clusters and Immune Responses



9. Detected Pathogenic Signatures



10. Blood Microbial Signatures in Germ-free and Conventional Mice



11. Summary

- 16S rRNA gene sequencing in pediatric blood detects DNA signature of commensal and pathogenic microbes.
- About 4% (6/147) specimens show presence of significant amount of genetic material from known pathogens. However, if these signatures are from historic or active infections remains to be established.
- Signatures from commensals can be stratified into four major clusters.
- Cluster 1 show association with BMI in children.

Please email prithvi.raj@utsouthwestern.edu, if you have any question or any potential collaboration opportunity. **Thanks for visiting our POSTER !!!**

Pre-Symptomatic Detection of COVID-19 from Smartwatch Data

Tejaswini Mishra*, Meng Wang*, Ahmed A Metwally*, Gireesh K Bogu*, **Andrew W Brooks***, Amir Bahmani*, Arash Alavi*, Alessandra Celli, Emily Higgs, Orit Dagan-Rosenfeld, Bethany Fay, Susan Kirkpatrick, Ryan Kellogg, Michelle Gibson, Tao Wang, Erika M Hunting, Petra Mamic, Ariel B Ganz, Benjamin Rolnik, Xiao Li**, Michael P Snyder**
Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
Nature Biomedical Engineering[3] - *Co-First Authors - **Corresponding Authors

Stanford MEDICINE

COVID-19 Wearables

Snyder Lab - Stanford Genetics

Introduction

The COVID-19 pandemic caused by the SARS-CoV-2 virus has resulted in 32 million infections and 577,041 deaths in the US alone[1]. This paramount microbiological crisis of our lifetimes is fueling unprecendented investment in traditional avenues of treatment such as vaccines, but also warants the use of novel technologies toward early public detection and monitoring of COVID-19. Our lab previously demonstrated the utility of wearable health trackers in predicting other respiratory infections[2].

Project Aim: Implement algorithms to predict COVID-19 prior to symptom onset using data from wearable devices.

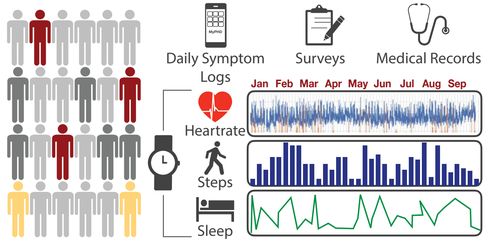


Figure 1: Phase I Study overview. In the phase I (N=7,492) we obtained data and developed algorithms to predict COVID-19. Some individuals developed COVID-19 (red), some developed other illnesses (yellow), and many remained healthy.

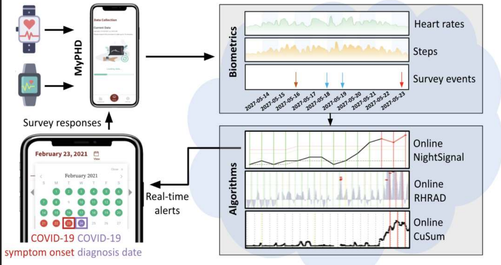


Figure 2: Phase II study returning real-time COVID-19 predictions. In phase II wearable data from 3,528 participants is analyzed twice daily. Data is retreived to cloud based algorithm system using our lab developed My Personal Health Dashboard (MyPHD) phone application. Analysis algorithms based on activity, heart-rate, and sleep generate alerts of elevated signals related to COVID-19. Alerts are then returned to participants through the MyPHD application along with surveys to assess prediction accuracy.

COVID-19 Detection Algorithms

In Phase I we refined algorithms using clear "Gold COVID-19" cases defined by robust wearable data, survey responses, and verified test results from 32 individuals. During phase II (in preparation for publication), these algorithms were implemented to send red (elevated signal) or green (normal signal) alerts to thousands of study participants twice a day. Algorithms developed include:

- **RHR-Diff:** Focuses on periods of Resting Heart Rate (RHR).
- **HROS:** Compares Heart-Rate Over Steps (HROS).
- **Night Signal:** Detects elevated heart rate during sleep.

Requirements: For Phase II implementation of alerts.

- Individually adaptive
- Train on single dataset
- Work within one month
- Tuneable parameters
- High sensitivity
- Reduce false positive alarms

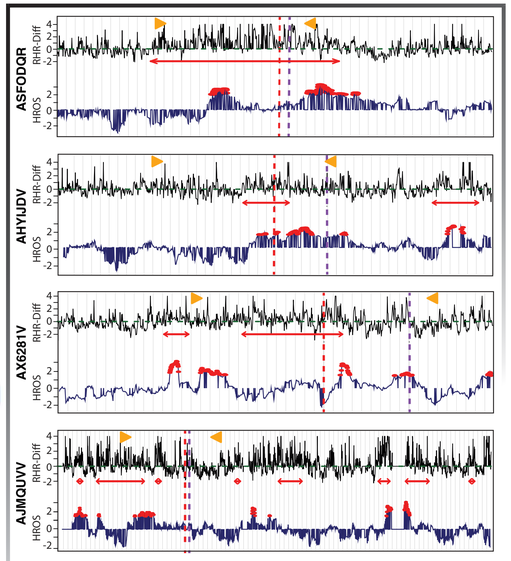


Figure 3: Four participant COVID-19 detection examples. Vertical lines represent one day (grey solid), symptom onset (dotted red), and diagnosis date (dotted purple). For RHR-Diff the black lines represent normalized RHR residuals, orange arrows encompass the 28-day baseline window, and red arrows represent periods of alarm. For HROS dark blue lines represent normalized heart rate, and red dots timepoints when anomalies are detected.

Predicting COVID-19

Algorithms were extensively refined using 32 gold COVID-19 cases, 15 individuals with other respiratory illnesses, and 79 healthy control subjects.

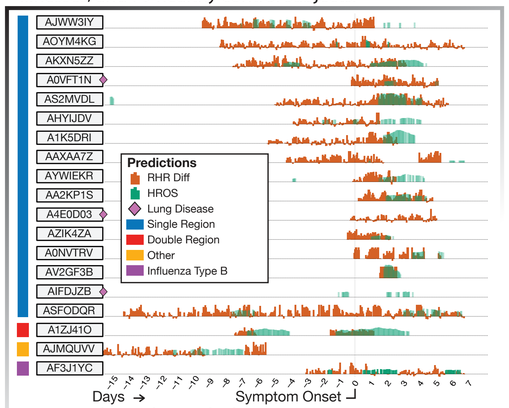


Figure 4: Alarms detected relative to COVID-19 symptom onset. Each row represents 21 days around an individuals symptom onset, with RHR-Diff alarms in green and HROS alarms in orange. Colored groups highlight alarm accuracy as a single clear alert around symptom onset, as well as repetetive, and other alerts patterns. One individual with flu is shown for comparison.

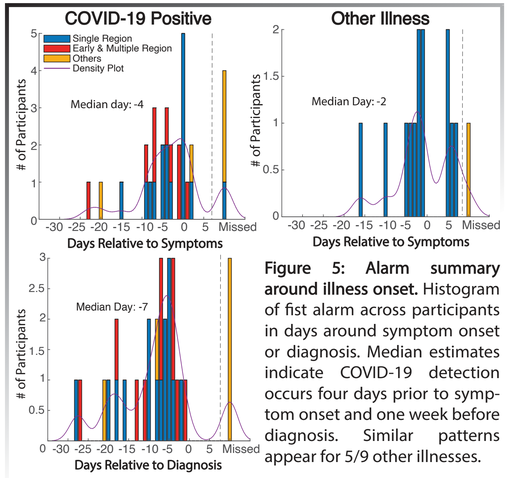


Figure 5: Alarm summary around illness onset. Histogram of fist alarm across participants in days around symptom onset or diagnosis. Median estimates indicate COVID-19 detection occurs four days prior to symptom onset and one week before diagnosis. Similar patterns appear for 5/9 other illnesses.

Real-Time COVID-19 Detection

Algorithms were adapted for detection of COVID-19 from real-time wearable data, where Phase II alarms are sent to participants based on daily uploads.

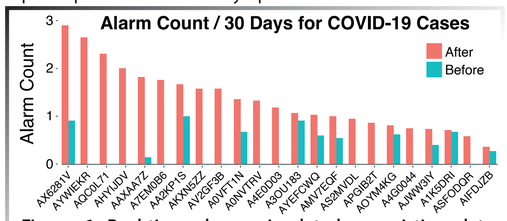


Figure 6: Real-time alarms simulated on existing data. Alarms appeared far more frequently after COVID-19 onset within 30 day averages around illness periods.

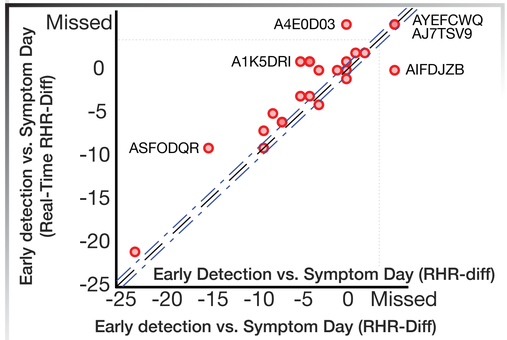


Figure 7: Real-time vs RHR-Diff around symptom onset. Real-time adaption does not limit algorithm accuracy.

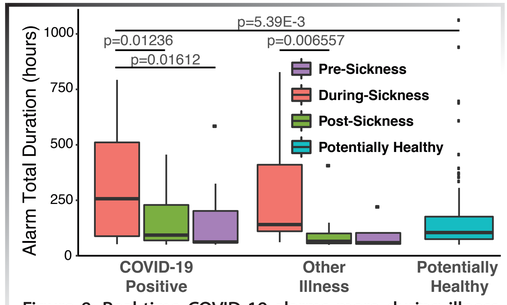


Figure 8: Real-time COVID-19 alarms more during illness periods, and significantly more than healthy rates.

Other Wearable Measures

Other wearables measures reflect COVID-19 as well.

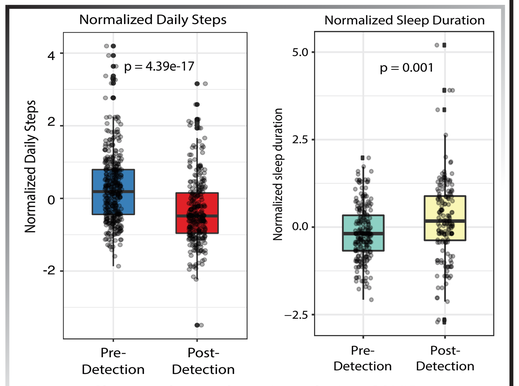


Figure 9: Sleep and step alterations during COVID-19.

Conclusions

1. COVID-19 is associated with changes in wearable measures for 80% of infection cases examined.
2. Alarms raised before symptom onset in 88% of cases.
3. Real-time detection is effective at COVID-19 detection at or before symptom onset in 63% of cases.

Wearable technologies will provide a useful approach for personalized management of epidemics.

Future Directions

Phase II data will be published, and results compiled across two study phases to seek FDA approval of algorithms. We will also further investigate detection of other infectious diseases, and roles of activity and lifestyle in false alerts.

References

1. CDC COVID-19 Data Tracker. October 2020. https://covid.cdc.gov/covid-data-tracker

2. Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information. 2017. https://doi.org/10.1371/journal.pbio.2001402

3. Pre-symptomatic detection of COVID-19 from smartwatch data. 2020. https://www.nature.com/articles/s41551-020-00640-6

Gut metabolites predict *Clostridioides difficile* recurrence

Jennifer J. Dawkins[1,3], Jessica R. Allegretti[2,4], Travis E. Gibson[1], Lynn Bry[1,2], Georg K. Gerber[1,2,3]

[1]Dept. Pathology, Brigham & Women's Hospital; [2]Massachusetts Host-Microbiome Center; [3]Harvard-MIT Health Sciences & Technology , Harvard Medical School; [4]Division of Gastroenterology, Brigham and Women's Hospital, Harvard Medical School

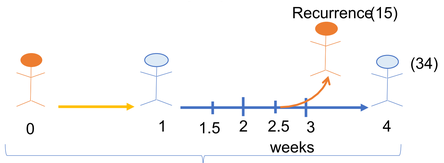
Introduction

Clostridioides difficile infection (CDI) is the most common hospital acquired infection in the U.S., causing ~450,000 cases and 29,000 deaths annually. CDI recurrence in patients is high: ~25% for the first recurrence and increasing with each episode. The initial development of CDI and its recurrence are mechanistically tied to disruption of the normal gut microbiota. Metabolites reflect functional activities of the microbiome and pathways common to multiple bacterial species, and thus may provide a clearer picture than microbial compositional data alone. In order to gain insights into gut-related factors contributing to CDI recurrence, we analyzed stool samples from 53 participants at diagnosis of CDI, directly after cessation of treatment, and weekly or bi-weekly for 4-6 weeks or until recurrence occurred. Each sample was interrogated using 16S rRNA amplicon sequencing and liquid-chromatography/mass-spectrometry (LC/MS) untargeted metabolomics. Using lasso-penalized logistic regression on these data, we developed predictors of CDI recurrence. Our predictor achieved a median cross-validated area-under-the-curve (AUC) of 0.771 with a 95% interval (0.753, 0.790) when using only metabolome data, compared to a median 0.601 AUC (0.550, 0.662) when using only microbial composition data. The combined data achieved a median AUC of 0.760 (0.689, 0.833) and moreover selected only metabolite covariates, suggesting no gain in predictive capability from the microbial composition data. We found several metabolites that predict recurrence, including a host inflammatory biomarker, a metabolite reported to affect permeability of the intestinal lumen, and a metabolite highly associated with microbial-host co-metabolism.

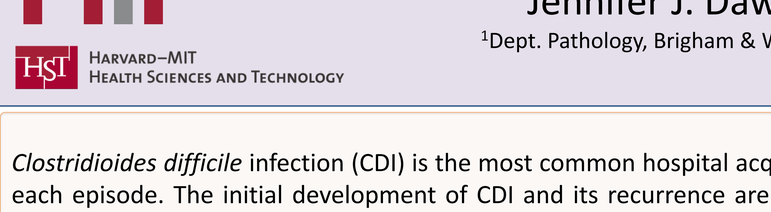
Study Design & Data Collection

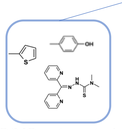
Eligibility Criteria	Exclusion Criteria
• Primary CDI • CDI symptoms (diarrhea) • Positive *C. difficile* test -by either glutamate dehydrogenase and enzyme immunoassay or polymerase chain reaction • Currently or imminently undergoing CDI treatment	• Inflammatory bowel disease • Immunodeficiencies • Severe / fulminant CDI • Ongoing non-CDI antibiotic use • Recurrence in week 1 of study

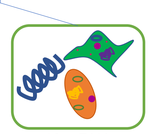
49 participants
- 15 Recurred
- 34 Asymptomatic



Samples collected:
- before treatment
- immediately after treatment
- every week or ½ week for 4 weeks or until recurrence




LC/MS untargeted metabolomics


16s rRNA amplicon sequencing

Samples analyzed by LC/MS untargeted metabolomics and 16s rRNA amplicon sequencing

Data Characteristics & Pre-Processing

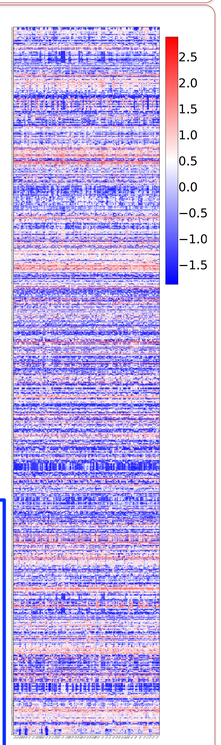
16S rRNA amplicon analysis

Raw sequencing results processed through dada2 pipeline

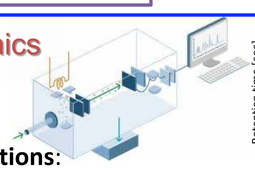
Subsequent filtering and transformations:
- Remove OTUs present in fewer than 15% of participants of each class at less than the limit of detection (10 counts)
- Transform counts to proportions & log transform counts
- Remove OTUs with less than a 5% coefficient of variance across all participants' week 1 data

123 OTUs
49 participants

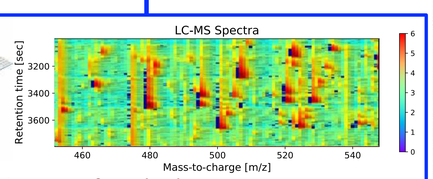
1120 metabolites
49 participants



LC/MS untargeted metabolomics



Samples analyzed and annotated by Metabolon



Subsequent filtering and transformations:
- Remove metabolites present in fewer than 25% of participants of each class
- Standardize values & log transform standardized values
- Remove OTUs with less than a 5% coefficient of variance across all participants' week 1 data

Recurrence Prediction

Methods
- Lasso logistic regression was used to predict recurrence from each participant's week 1 sample using the 16s data, metabolic data, and joint data
- Leave one out nested cross validation (CV) was used to optimize λ.

Lasso logistic regression performs L1 logistic regression (eqns 1 & 2) and then shrinks the less important feature coefficients to zero

If we have S training instances $(\mathbf{x}^{(i)}, y^{(i)}), i = 1, ..., S$
And our data contains M metabolites and B OTUs:

For metabolic data $\mathbf{x}_m^{(i)} \in \mathbb{R}^M$ For 16S data $\mathbf{x}_b^{(i)} \in \mathbb{R}^B$ For both data sources $\mathbf{x}_j^{(i)} \in \mathbb{R}^{M+B}$

For all 3 types of input data $x^{(i)} \in \{x_m^{(i)}, x_b^{(i)}, x_k^{(i)}\}$,

$$P(y = \text{"RECUR"}|\mathbf{x}^{(i)}, \vec{\beta}) = \frac{1}{1 + \exp(-(\beta_0 + \vec{\beta}^T \mathbf{x}^{(i)}))} \quad (1)$$

$$\min_\beta \sum_{i=1}^{S} -\log\, p(y^{(i)}|\mathbf{x}^{(i)}; \beta) + \lambda||\beta||_1 \quad (2)$$

Results & Discussion

Prediction accuracy

- Metabolomic data is a significantly stronger predictor of CDI recurrence than 16s rRNA amplicon data
- Combining data sources achieved a very similar AUC to predicting from metabolites alone
- All significant features found in the combined model were metabolites.

Input Features	Over 49 Folds	
	Median AUC	95% Interval
1. 16s rRNA OTUs	0.601	(0.550, 0.662)
2. LC-MS Metabolomics	**0.771**	**(0.753, 0.790)**
3. 16s rRNA + LC-MS Metabolomics	0.760	(0.689, 0.833)

Biomarkers selected

Metabolites that predict recurrence include:
- A host inflammatory biomarker
- A metabolite reported to affect permeability of the intestinal lumen
- A metabolite highly associated with microbial-host co-metabolism

The metabolite that predicts protection against CDI recurrence has been implicated in antimicrobial activity and cell cycle regulation .

Significant biomarkers	Over 49 folds and 50 random seeds	
	Median Odds Ratio	90% Interval
Metabolite 1	1.474	(1.338, 1.630)
Metabolite 2	1.394	(1.301, 1.572)
Metabolite 3	1.162	(1.036, 1.249)
Metabolite 4	1.113	(1.0, 1.221)
Metabolite 5	0.9289	(0.884, 1.0)

Conclusions & Future Work

- Our results indicate that gut-metabolites can accurately predict CDI recurrence and may provide mechanistic insights into CDI; we did not find that microbial composition data could predict CDI with a simple logistic regression model.
- These gains in prediction and better understanding CDI recurrence could enable prompt, targeted treatments to short-circuit the vicious cycle of recurrence

To build on this work, we plan to:
- Create a novel computational model that uses prior biological knowledge with the aim of higher predictive accuracy and discovery of a broader array of metabolomic features
- Create joint models of microbial and metabolomic data to capture data dependencies expected to improve predictive accuracy and interpretability
- Incorporate temporal information into models, including non-stationarity of the microbiome

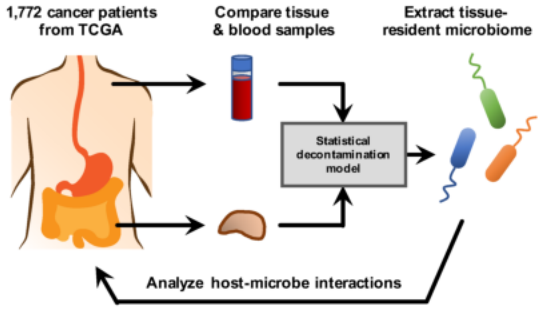
Acknowledgements This work was supported by the NSF GRFP, the BWH Precision Medicine Initiative, BWH President's Scholar Award, Harvard Catalyst and NIGMSR01GM130777.

The Cancer Microbiome Atlas (TCMA): A Resource for Querying Host-Microbe Interactions

Anders B. Dohlman[1], Diana Arguijo Mendoza[1], Shengli Ding[1], Michael Gao[2], Holly Dressman[3], Iliyan D. Iliev[4], Steven M. Lipkin[4], Xiling Shen[1#]

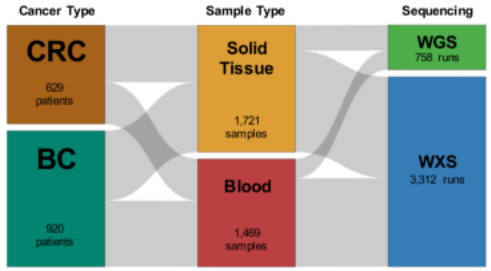
[1] Department of Biomedical Engineering, Center for Genomics and Computational Biology, Duke Microbiome Center, Duke University, Durham, NC 27708, USA
[2] Duke Institute for Health Innovation, Duke University, Durham, NC 27701, USA
[3] Department of Molecular Genetics and Microbiology, Director of Duke Microbiome Center, Duke University, Durham, NC 27708, USA
[4] Department of Medicine, Weill-Cornell Medical College, Cornell University, New York City, NY 10065, USA

Abstract



Studying the microbial composition of internal organs and their associations with disease remains challenging due to the difficulty of acquiring clinical biopsies. We designed a statistical model to analyze the prevalence of species across sample types from The Cancer Genome Atlas (TCGA), revealing that species equiprevalent across sample types are predominantly contaminants, bearing unique signatures from each TCGA-designated sequencing center. Removing such species mitigated batch effects and isolated the tissue-resident microbiome, which was validated by original matched TCGA samples. Mixed-evidence species can be further distinguished by gene copies and nucleotide variants. We thus present The Cancer Microbiome Atlas (TCMA), a collection of curated, decontaminated microbial compositions of oropharyngeal, esophageal, gastrointestinal, and colorectal tissues. This led to the discovery of prognostic species and blood signatures of mucosal barrier injuries and enabled systematic matched microbe-host multi-omics analyses, which will help guide future studies of the microbiome's role in human health and disease.

Study Design



We examined TCGA whole-genome (WGS) and whole exome (WXS) sequencing of solid tissue and blood samples from colorectal cancer (CRC) and brain cancer patients (BC).

Results

Distinguishing tissue-resident taxa from contamination. We identified two species populations: one "tissue-resident" (unique to CRC tissue) and one "equiprevalent" (equally prevalent across sample types) (Fig. 1). We then isolated the tissue-resident population, by decomposing observed metagenomic data into tissue-resident and contaminant fractions (Fig. 2).

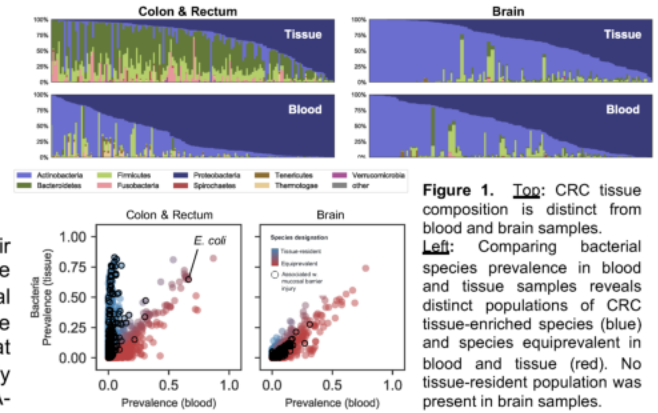


Figure 1. Top: CRC tissue composition is distinct from blood and brain samples. Left: Comparing bacterial species prevalence in blood and tissue samples reveals distinct populations of CRC tissue-enriched species (blue) and species equiprevalent in blood and tissue (red). No tissue-resident population was present in brain samples.

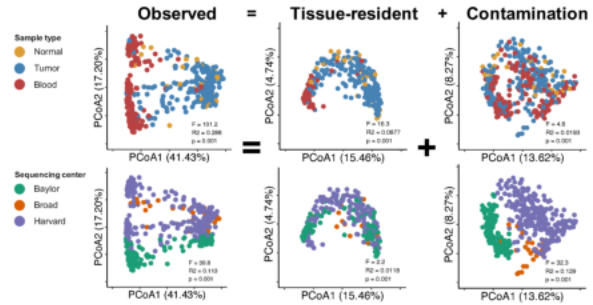


Figure 2. Observed metagenomic data (K) contains both biological and technical variation but can be decomposed into tissue-resident (T) and contaminant (C) fractions using a mixture model of the form $K = mT + nC$. The tissue-resident fraction retains biological variation (top), while the contaminant fraction retains technical variation related to sequencing center (bottom).

Resolving "mixed-evidence" species. For some species, reads counts come from some unknown combination of endogenous and contaminant bacteria. One such species is *E. coli.*, which is equiprevalent (Fig. 1) but ubiquitous among human microbiomes. We identified several *E. coli* genes that were enriched in tissue and can be used to resolve mixed-evidence cases (Fig. 3).

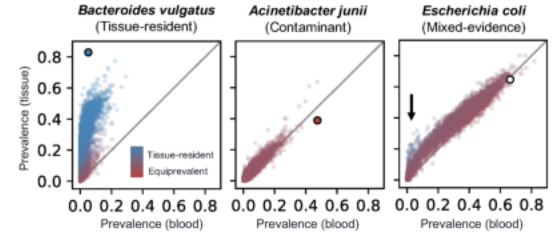


Figure 3. Prevalence of genes belonging to *B. vulgatus* (left; tissue-resident), *A. junii* (center; contaminant), and *E. coli* (right; mixed-evidence) in blood vs. tissue. Large dots indicate species-level prevalence. Arrow indicates tissue-enriched *E. coli* genes.

Experimental validation. To validate decontaminated TCGA microbiomes, we obtained original tissue and blood samples from five TCGA patients and performed 16S rRNA sequencing. Decontaminated profiles were consistent with 16S results of matched tissue.

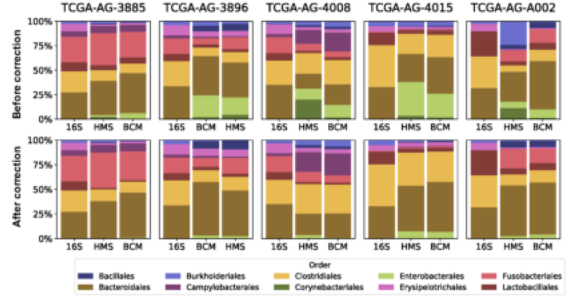


Figure 4. Relative abundances in 16S results of original tissue compared to TCGA sequencing of matched tissue samples at Harvard and Baylor, before and after decontamination.

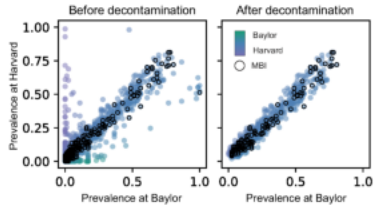


Figure 5. Prevalence of bacterial species in tissue samples sequenced at Baylor vs. Harvard before and after removing contamination Center-specific contaminants have been removed.

TCMA can be compared with multi-omic molecular data from TCGA to analyze host-microbe interactions. Decontaminated metagenomic profiles of TCGA tissues are matched to clinical metadata as well as epigenetic, genomic, transcriptomic, and proteomic profiling, allowing an integrated multi-omic analysis. We identified several known and novel taxa associated with CRC tumors compared with matched normal tissue (Fig. 6). Tissue microbiomes clustered into "Bacteroides" and "Fusobacterium" coabundance groups, which were predictive gene expression (Fig. 6).

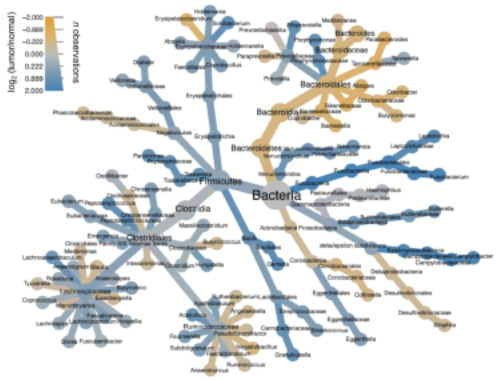


Figure 6. Heat-tree showing genera associated with tumor samples (blue) and matched normal tissue (yellow).

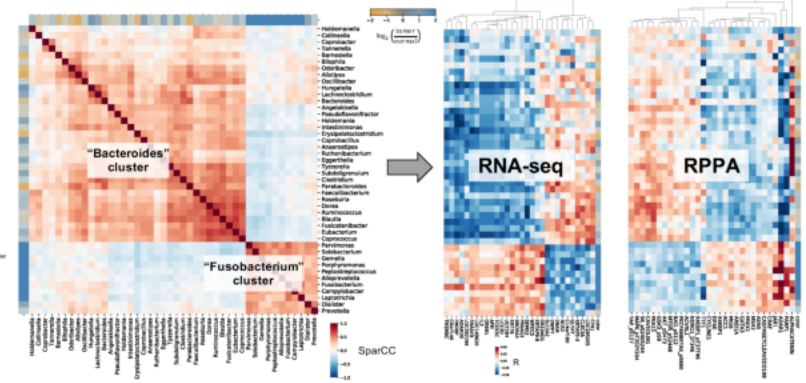


Figure 7. Left: We identified two coabundance groups of genera in CRC tumors. Right: These coabundance groups were predictive of gene transcription (RNA-seq) and protein expression (RPPA).

Reference

Dohlman AB, Arguijo Mendoza D, Ding S, Gao M, Dressman H, Iliev ID, Lipkin SM, Shen X. *The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants.* Cell Host Microbe. 2021 Feb 10;29(2):281-298.e5. doi: 10.1016/j.chom.2020.12.001. PMID: 33382980; PMCID: PMC7878430.

Explore the TCMA website!

Interactively examine host-microbe interactions using the TCMA database.

https://tcma.pratt.duke.edu/

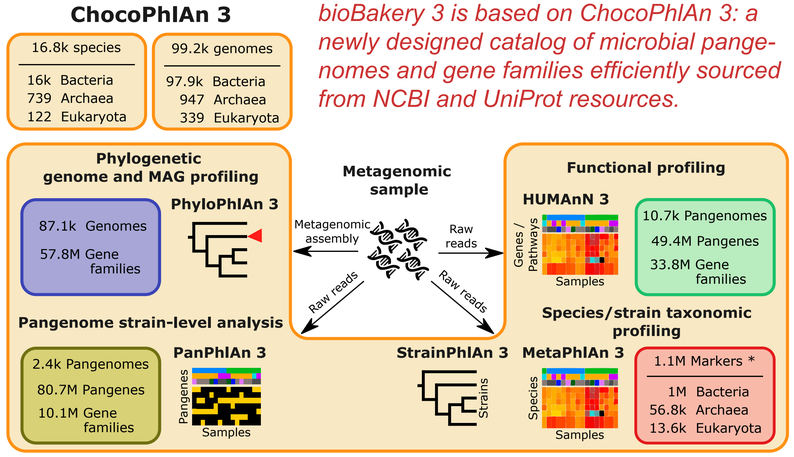
Integrating taxonomic, functional, and strain profiling of microbial communities with bioBakery 3

F. Beghini[1], L.J. McIver[2], A. Blanco-Míguez[1], L. Dubois[1], F. Asnicar[1], S. Maharjan[2], A. Mailyan[2], A.M. Thomas[1], P. Manghi[1], M. Valles-Colomer[1], G. Weingart[2], Y. Zhang[2], M. Zolfo[1], C. Huttenhower[2], E.A. Franzosa[2], N. Segata[1]

[1]Dept. CIBIO, University of Trento; [2]The Harvard Chan Microbiome in Public Health Center & The Broad Institute

bioBakery 3 design and evaluation

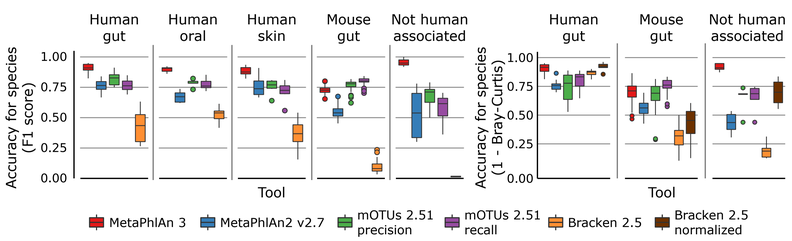
Culture-independent analyses of microbial communities have improved dramatically in the last decade, particularly due to advances in methods for biological profiling via shotgun metagenomics. Opportunities for improvement continue to accelerate given greater access to multi-omics, microbial reference genomes, and strain-level diversity. To leverage these resources, we present bioBakery 3: a set of integrated and improved methods for taxonomic, strain-level, functional, and phylogenetic profiling of metagenomes and metatranscriptomes developed using the largest set of reference sequences now available.



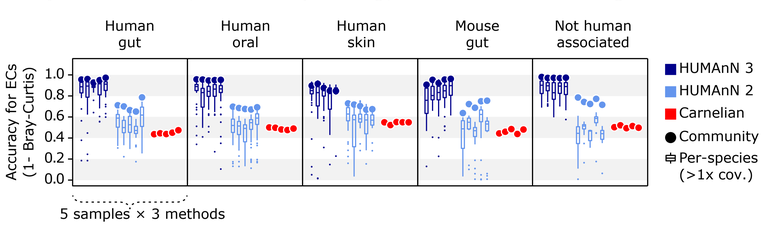
bioBakery 3 is based on ChocoPhlAn 3: a newly designed catalog of microbial pangenomes and gene families efficiently sourced from NCBI and UniProt resources.

bioBakery 3 accurately profiles synthetic metagenomes

We compared MetaPhlAn 3 with other taxonomic profiling methods in the task of identifying and quantifying species from synthetic metagenomes. MetaPhlAn 3 displayed high and often superlative accuracy, including in murine and non-human-associated communities.

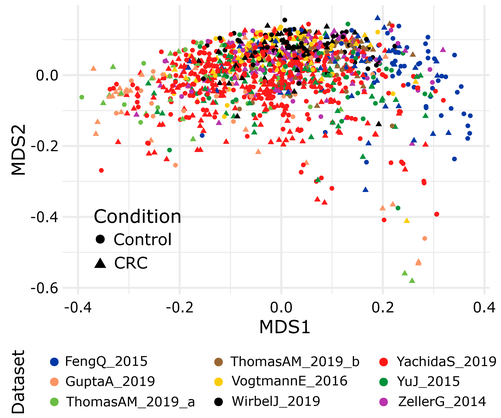


Below, HUMAnN 3 was similarly highly accurate in quantifying microbial enzyme families (functional profiling) from the same metagenomes.



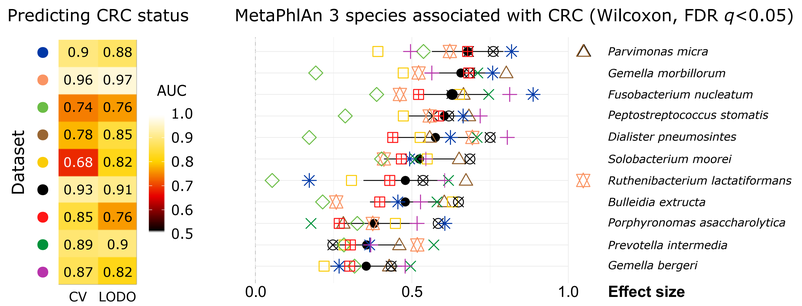
Expanding our understanding of the gut microbiome in CRC and IBD

Meta-analyzing the CRC gut microbiome with MetaPhlAn 3

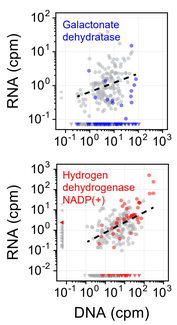


Left, Nine CRC datasets totaling >1.2K samples broadly overlapped in species composition (weighted UniFrac).

Below, MetaPhlAn 3 species enabled strong cross-validated (CV) and leave-one-dataset-out (LODO) prediction of CRC status; 11 species were individually and consistently differentially abundant during CRC.



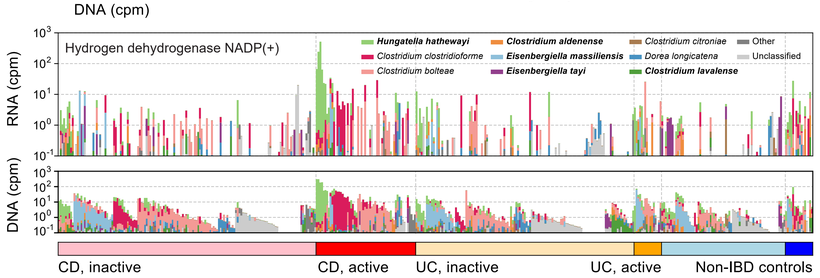
Microbial differential expression in IBD with HUMAnN 3



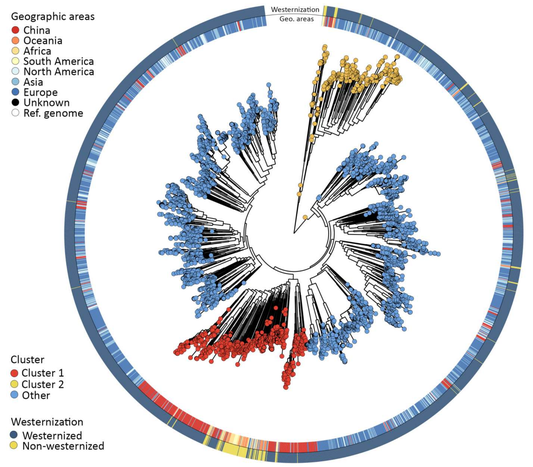
We applied HUMAnN 3 to profile microbial enzymes from 1,603 metagenomes (DNA) and 794 metatranscriptomes (RNA) from subjects in the IBDMDB cohort (HMP2).

Left, examples from >500 enzyme families that were differentially expressed (DE'ed) in samples with IBD-linked dysbiosis after controlling for gene (DNA) copy number.

Below, HUMAnN 3 reveals species contributions to a representative DE'ed enzyme; contributors with bold names are new to the bioBakery 3 platform.

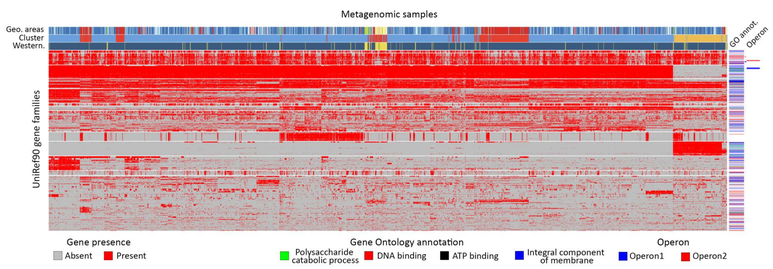


Profiling *Ruminococcus bromii* strains from >4K gut metagenomes



Above, StrainPhlAn 3 identified and compared SNV-level strains of the common human gut microbe *R. bromii* from 4,077 geographically diverse metagenomes; two distinct subspecies-level clusters were observed, one of which was enriched in individuals from China.

Below, PanPhlAn 3 compared the gene content of well-covered strains to *R. bromii* isolate genomes, highlighting functional consequences of strain differentiation (e.g. enrichment for membrane proteins in Cluster 2).



Acknowledgments

The work was supported by the European Research Council (ERC-STG project MetaPG-716575) to NS; MIUR "Futuro in Ricerca" (No. RBFR13EWWI_001) to NS; the European H2020 program (ONCOBIOME-825410 project and MASTER-818368 project) to NS; the National Cancer Institute of the National Institutes of Health (NIH; 1U01CA230551) to NS; the Premio Internazionale Lombardia e Ricerca 2019 to NS; the Harvard Chan Microbiome Analysis Core (CH); the National Institute of Diabetes and Digestive and Kidney Diseases of the NIH (R24DK110499 and U54DE023798) to CH; Cancer Research UK Grand Challenge award C10674/A27140 to Wendy Garrett (CH); the Juvenile Diabetes Research Foundation (3-SRA-2016-141-Q-R) to CH; and the National Human Genome Research Institute of the NIH (R01HG005220) to Raphael Irizarry (CH).

Discover bioBakery software and tutorials via
http://huttenhower.sph.harvard.edu/biobakery

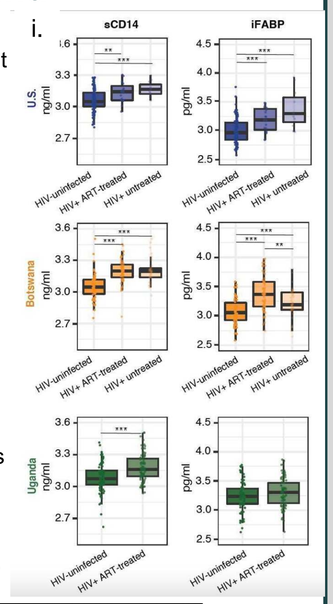
CD8+ T Cells Mediate Colon Epithelial Cell Death in an Organoid Model of HIV Pathogenesis

Upasana Das Adhikari[1], **Leah M. Froehle[1],** Alice H. Linder[1], Muntsa Rocafort Junca[1], Mark S. Ladinsky[2], and Douglas S. Kwon[1]
[1]Ragon Institute of MGH, MIT, and Harvard. Massachusetts General Hospital, Harvard Medical School, Boston, MA USA;
[2]California Institute of Technology Division of Biology and Biological Engineering

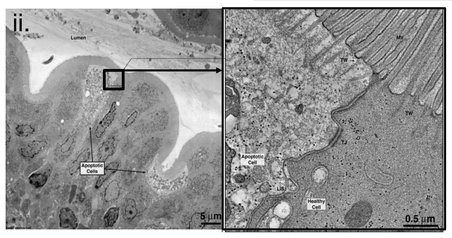
Background

- Only partial CD4 recovery in the gut is observed in HIV patients.
- Gut barrier damage persists even when viral load is suppressed by ART.
- Gut barrier disruption can lead to translocation of microbes and microbial products, leading to systemic inflammation.
- Soluble CD14, marker of systemic inflammation and iFABP, marker of intestinal barrier disruption remain upregulated in HIV positive patients with ART treatments (i).
- Tight junctions remain intact in HIV positive patients' gut epithelium (ii).

i.

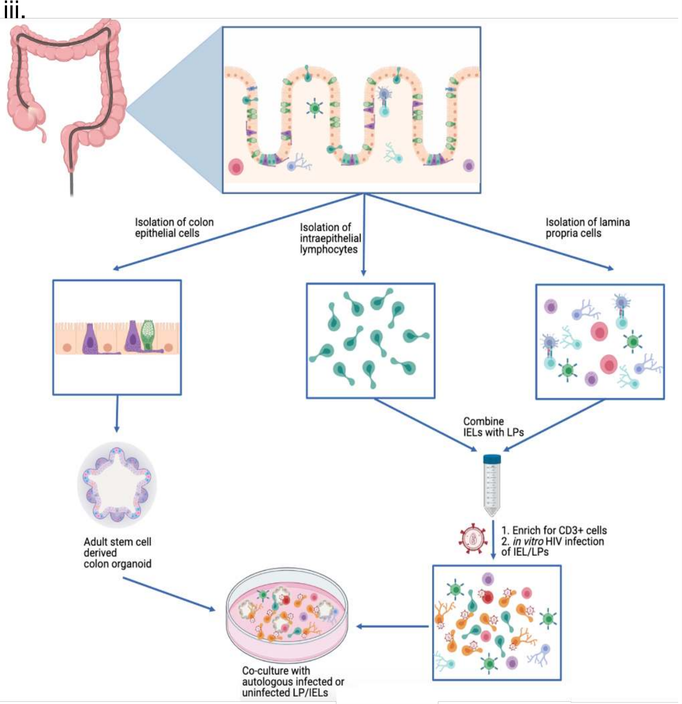


ii.



Methods

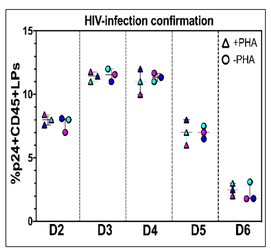
iii.

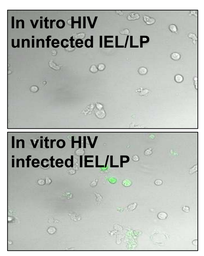


- Colonoids derived from biopsy pinches of HIV positive patients and were maintained as primary culture.
- Colonoids were passaged and maintained for further experiments.
- Colonoids were co-cultured with primary Lamina Propria (LP) derived immune cells.
- Colonoids and LP were cultured from excess surgical colon of HIV negative patients.

iv.

Infection (D1) → Change media (D2) → P24 staining (D5)

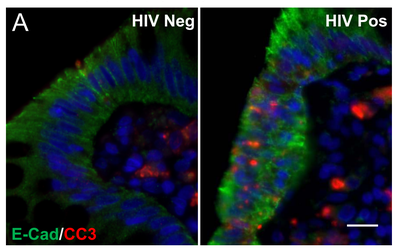




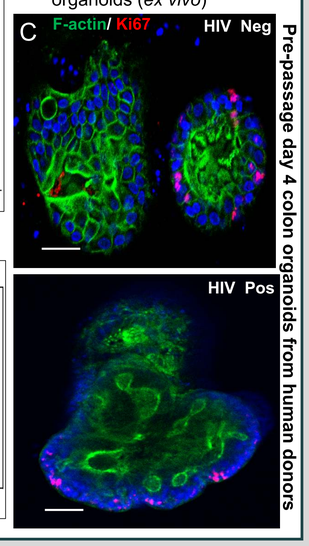
Results

Figure 1. Colon epithelial disruption during HIV pathogenesis

Apoptosis (CC3) in donor-derived colon epithelium (*in vivo*)



Deformed cytoskeleton (F-actin) in donor derived colon organoids (*ex vivo*)



Apoptosis (CC3) in donor-derived colon organoids (*ex vivo*)

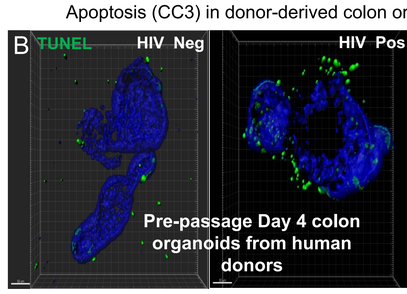
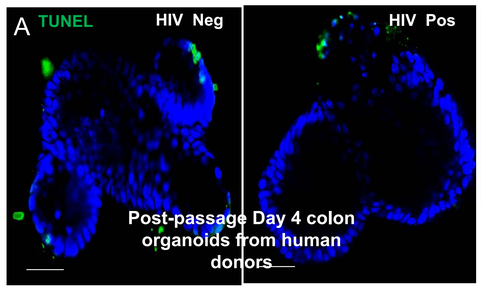
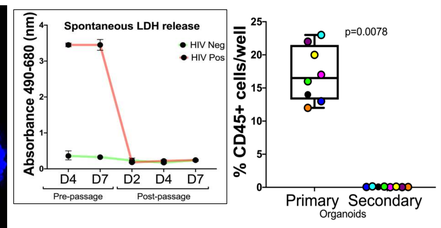
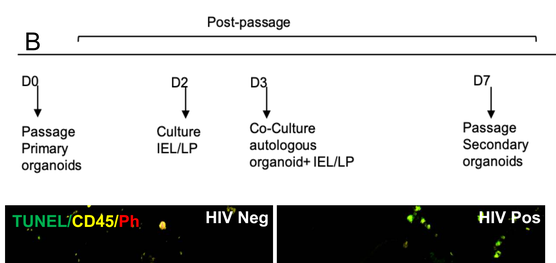


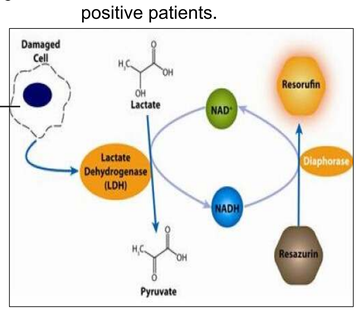
Figure 2. Colon mucosal immune cells drive epithelial cell death in organoid model.

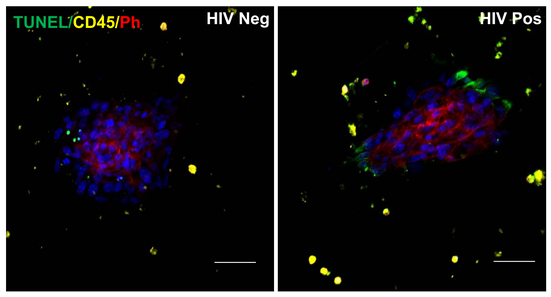


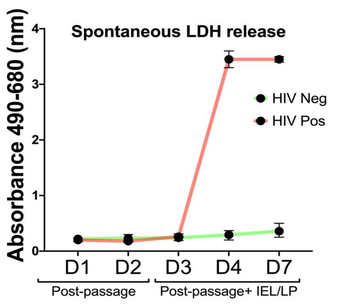


Appearance of colon epithelial disruption upon autologous LP immune cells co-culture from HIV positive patients.









C

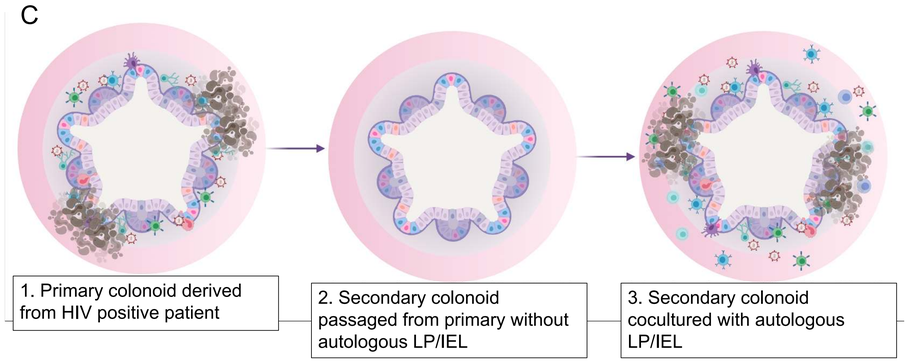
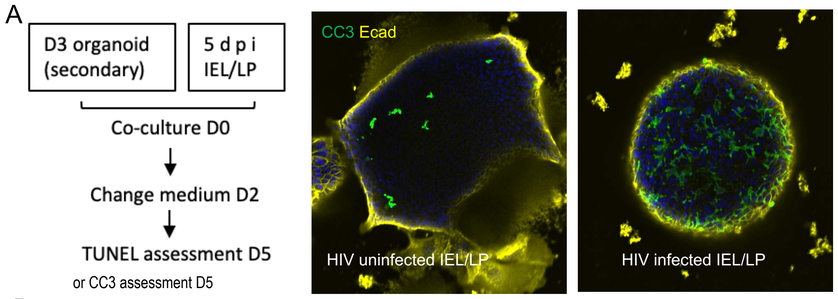
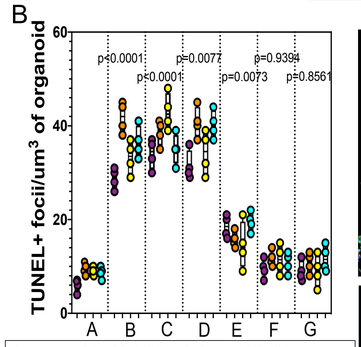


Figure 3. CD8+ T cell contact-dependent epithelial cell death

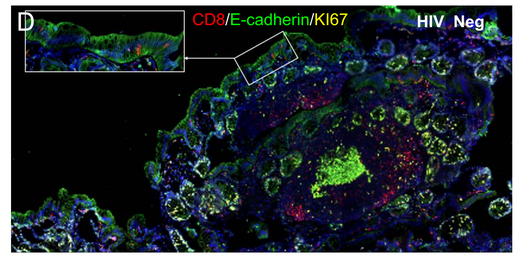
A

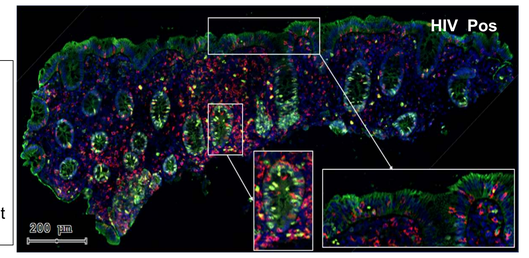


B



A: organoid with HIV uninfected IEL/LP
B: organoid with HIV infected IEL/LP
C: organoid with HIV infected+Emapulumab
D: organoid with HIV infected+Etanercept
E: organoid with HIV infected sup
F: organoid with HIV infected sup+Emapalumab
G: organoid with HIV infected sup+Etanercept

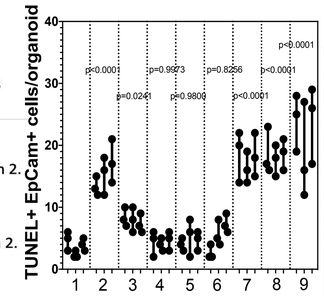




C

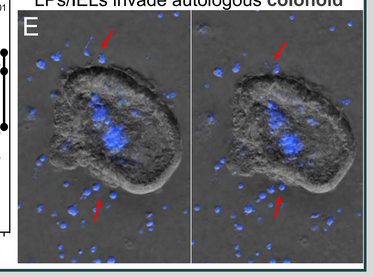
CD3+T cells → HIV infection → enrichment → CD4+T cells / CD8+T cells

1. CD3+T cells uninfected
2. CD3+T cells +HIV
3. CD4+ T cells enriched from 2.
4. 3+IFNg inhibitor
5. 3+TNF inhibitor
6. CD8+T cells stimulated.
7. CD8+T cells enriched from 2.
8. 7+IFNg inhibitor
9. 7+TNF inhibitor



CD8+ T cells from HIV infected LPs/IELs invade autologous **colonoid**

E



Summary

HIV negative patient's colon mucosa

HIV positive patient's colon mucosa

Future directions

- Characterization of CD8+ T cells responsible for epithelial apoptosis (Natural Killer T cells?).
- Metabolic profiling of CD8+ T cells in colon mucosa.
- Mechanism of CD8+ T cell mediated epithelial apoptosis.

Acknowledgments

I would like to thank my mentors **Dr. Upasana Das Adhikari** and **Dr. Doug Kwon**, the members of the Kwon lab, the Ragon Institute, the MGH Division of Gastroenterology and the MGH Surgical Pathology lab.
Funding: Ragon Institute, NIH, NIDDK, and Burroughs Wellcome Fund (DSK)
Note: all data shown are unpublished.

Contact: **Leah M Froehle**, lfroehle@mgh.harvard.edu

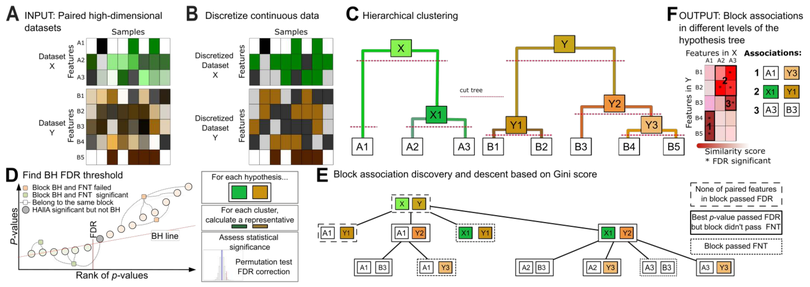
High-sensitivity pattern discovery in large, paired multi-omic datasets with HAllA

Andrew R. Ghazi[1,2], Kathleen Sucipto[1], Gholamali Rahnavard[1,2], Eric A. Franzosa[1,2], Lauren J. McIver[1,2], Jason Lloyd-Price[1,2], Emma Schwager[1], George Weingart[1], Yo Sup Moon[1], Xochitl C. Morgan[3], Levi Waldron[4], Curtis Huttenhower[1,2]

[1]Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard [3]University of Otago [4]City University of New York

Modern biological screens yield enormous numbers of measurements, and finding interpretable, statistically significant associations among features is essential. Here, we present a novel hierarchical framework, HAllA (Hierarchical All-against-All association testing), for structured association discovery between paired high-dimensional datasets. HAllA efficiently integrates hierarchical hypothesis testing with false discovery rate correction to reveal significant linear and non-linear block-wise relationships among continuous and/or categorical. We optimized and evaluated HAllA using heterogeneous synthetic datasets of known association structure, where HAllA outperformed all-against-all and other block testing approaches across a range of common similarity measures. We then applied HAllA to a series of real-world multi-omics datasets, revealing new associations between gene expression and host immune activity, the microbiome and host transcriptome, metabolomic profiling, and human health phenotypes. An open-source implementation of HAllA is freely available at http://huttenhower.sph.harvard.edu/halla along with documentation, demo datasets, and a user group.

HAllA methodology



Hierarchical All-against-All Association testing (HAllA) identifies block associations between two potentially heterogeneous datasets co-indexed along one axis. This co-indexing is referred to as the "samples" axis (columns), and the measurement axis as "features" (rows). For a pair of datasets containing measurements that describe the same set of samples and a specified pairwise similarity measure, the HAllA algorithm proceeds by 1) optionally discretizing features to a uniform representation (if required by the similarity measure), 2) finding the Benjamini–Hochberg (BH) FDR threshold, 3) hierarchically clustering each dataset separately to generate two data hierarchies, 4) iteratively dividing blocks of hypotheses according to Gini score gain in the data hierarchies and a false negative tolerance (FNT) threshold.

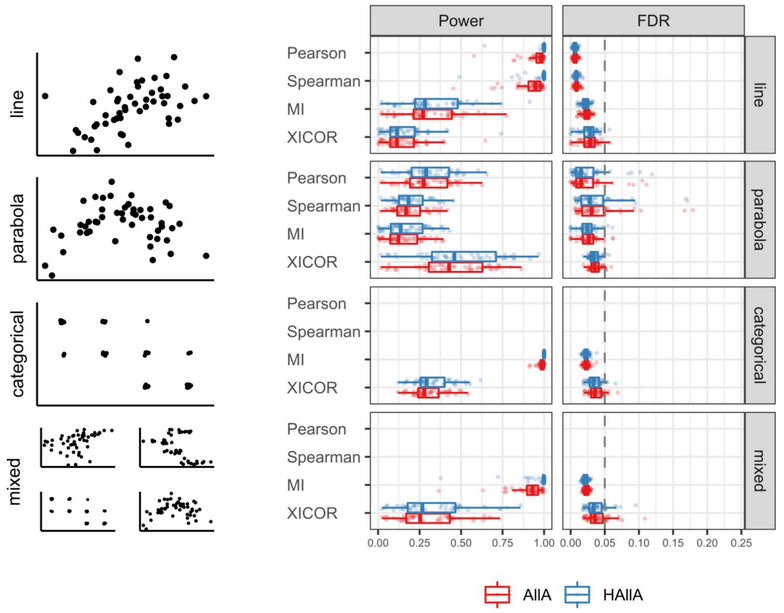
Why HAllA?

· **Broad applicability:** HAllA's methodology works on nearly all commonly found data types. A variety of user-configurable parameters such as false negative tolerance and similarity metric are set to common-sense defaults providing good performance from the outset.

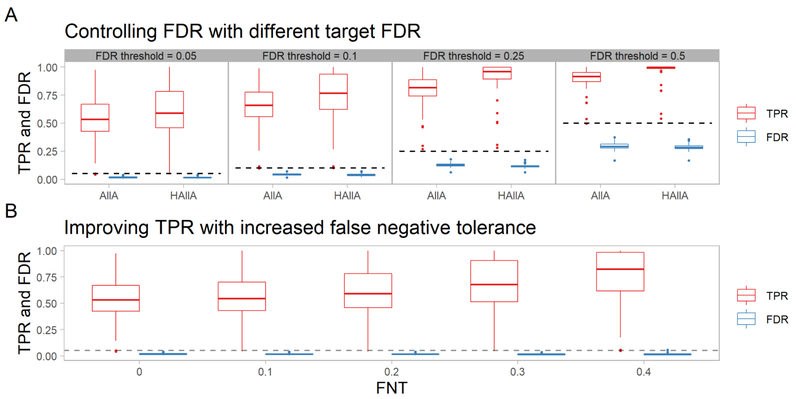
· **Well-powered:** Relative to all-against-all (AllA) pairwise association testing, HAllA consistently provides higher power.

· **Interpretable:** HAllA groups large feature sets into coherent association blocks. Built-in visualization methods make it easy to see the contents and association strength of these blocks.

HAllA is well-powered while controlling false discovery rate

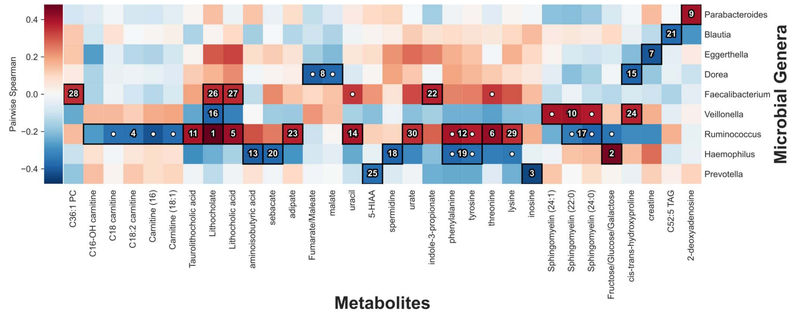


We simulated paired, 200-dimensional datasets with 50 samples and 10 association blocks across a variety of association structures, then applied HAllA and AllA with a variety of association metrics, then examined the difference in statistical power and false discovery rate between the two methods. HAllA provided superior power and comparable false discovery rate in each case.

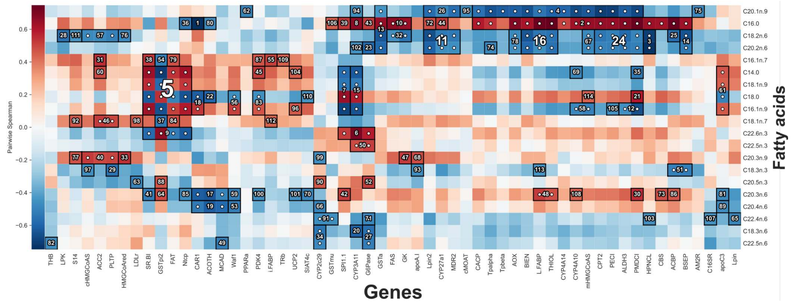


Using 50 pairs of synthetic datasets with 200 features and 50 samples containing clusters with quadratic block associations were analyzed. A) with FNT = 0.2, HAllA maintains the simulated FDR below the target (here (0.05, 0.1, 0.25, and 0.5), with associated trade-offs in statistical power. In addition, HAllA is consistently better powered than all-against-all (AllA) association testing across this range of target FDR values. Dashed lines parallel to the x-axis indicate the target FDR value in each comparison. B) By increasing the FNT, HAllA can improve the true positive rate with a comparatively minor increase in FDR..

HAllA identifies microbe-metabolite and gene-fatty acid association blocks



This "HAllAgram" shows the HAllA's result when applied to paired stool metabolomic and 16S rRNA gene sequencing data from the DIABIMMUNE cohort, in which infants were recruited at birth and sampled monthly for the first three years of life. The data comprise 104 samples and describes the abundance of 20 genera and 284 labeled metabolites. Here, we show the 30 strongest associations ranked by p-value (target FDR=0.05). Block associations are numbered in descending order of significance, and feature pairs that are marginally associated are dotted.



We applied HAllA to paired data comprising 120 hepatic transcript levels and 21 liver lipid levels in a set of 40 previously profiled mice. Each numbered block corresponding to a group of co-expressed transcripts related to a group of co-occurring lipids. A total of 114 block associations achieved significance at FDR 0.05, matching the previous study's threshold based on canonical correlation. HAllA's associations included all those found earlier by CCA. Spearman correlation was used as a similarity metric.

Acknowledgments

We thank Alex Kostic, Tommi Vatanen, and Vincent Carey for assistance obtaining and curating datasets for the applications section; Hera Vlamakis, Hector Corrada Bravo, William Shannon, A. Brantley Hall, Himel Mallick, Siyuan Ma, and Susan Holmes for helpful discussions, suggestions, and feedback. This study was supported by Army Research Office grant W911NF-11-1-0429, NSF DBI-1053486, and NIH U54DE023798 to Curtis Huttenhower.

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

Reconstruction of metagenome-scale models of the gut microbiota metabolism at species-level resolution in Inflammatory Bowel Disease

Isabella M. Goodchild-Michelman [1,2,3], Analeigha V. Colarusso [2,3,4], Ali R. Zomorrodi [2,3]

[1] Department of Molecular and Cellular Biology, Harvard Faculty of Arts and Sciences, Boston, MA. [2] Harvard Medical School, Boston, MA. [3] Mucosal Immunology and Biology Research Center, Mass General Hospital for Children, Boston, MA. [4] Department of Biology, Tufts University, Medford, MA.

Background

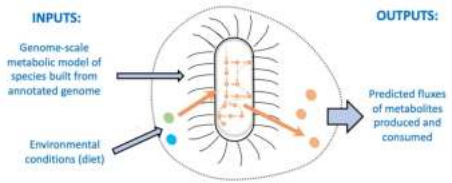
- **Inflammatory Bowel Disease (IBD)** is a chronic inflammatory condition of the intestinal tract that affects over three million Americans each year.
- **IBD has been linked to alterations in the gut microbiota** and previous studies have used amplicon and metagenomic sequencing and metabolomics to associate microbial species and microbially-derived metabolites in the gut microbiota with IBD but **underlying causal mechanisms of the disease are unknown.**
- All data used in this project is from the **Human Microbiome Project (HMP)**

Objective: Use GEMs and data from HMP to investigate metabolic interactions between the gut microbiota and host in IBD to better understand microbial and molecular mechanisms underlying these interactions.

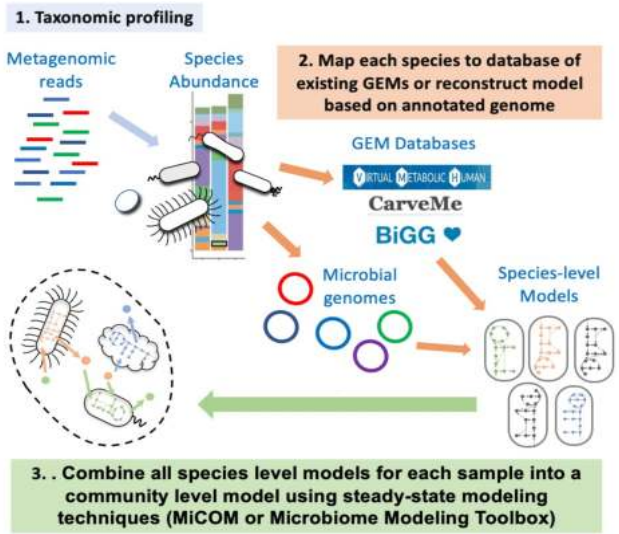
Genome-Scale Metabolic Models

- Assume a cell can be approximated by the network of its metabolic pathways and can be analyzed to trace a metabolite's production back to a specific microbial species in the gut.

Flux Balance Analysis of a Species-level GEM

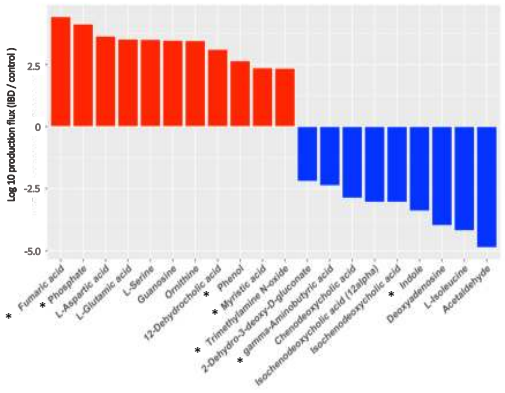


Methods



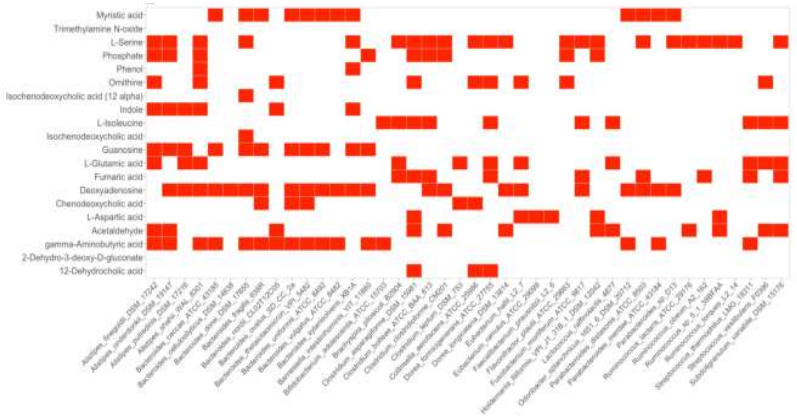
Pilot Study: Comparing the Microbiota of 4 IBD and 4 Control Subjects

Top 20 differentially produced metabolites in 4 cases vs 4 controls

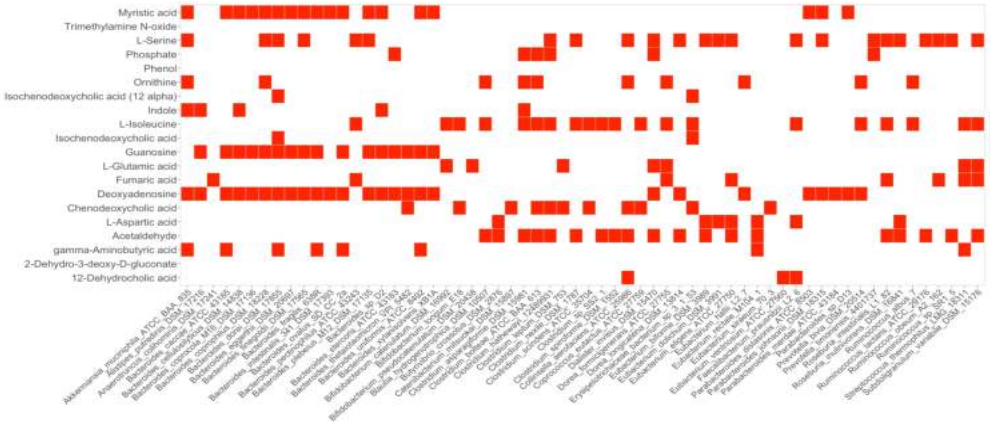


* Indicates a metabolite that has been previously implicated in IBD

Species-resolved metabolite productions in on IBD subject (MSM5LLER)



Species-resolved metabolite production in one non-IBD subject (SRS017247)



Additional differentially produced metabolites (not in the top 20) that have been previously linked to IBD:

Isobuteric acid
Glycine
Formic acid
Chorismate
Spermidine
Uridine
Proline
Histidine

Conclusion

Our preliminary results support the feasibility of this study, and they will serve as a platform for large-scale computational studies of the host-microbiota interactions

Next Steps:

- Large-scale simulations: Analyze all IBD vs control microbiomes from HMP
- Integrate models with a GEM of human intestinal epithelial cell in order to simulate microbial interactions with the human intestinal barrier.
- Analyze the results to determine relevant metabolite, inter-species, and host-microbiota cross-talk differences between the control and IBD models.

Deletion of innate effector serum amyloid A alters gut microbiome and drives metabolism in mice

Yue Sandra Yin [1], Laura J. den Hartigh[2], Xue-Song Zhang[1], Shari Wang[2], Zhan Gao[1], Abigail Armstrong[1], Jincheng Wang[3], Maria Gloria Dominguez-Bello[3], Martin J. Blaser[1]

[1]CABM, Rutgers University; [2]University of Washington; [3]SEBS, Rutgers University

RUTGERS Microbiome Program

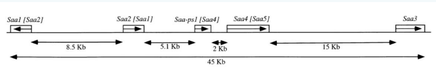
RUTGERS Center for Advanced Biotechnology and Medicine

Abstract

The serum amyloid A (SAA) proteins, mostly produced in the liver, are acute-phase reactants. SAA also is expressed in the intestinal epithelium, which is an interface between gut microbes and host immunity. SAA stimulates Th17 cells and functions as an inflammatory marker in infectious disease and metabolic disorders. Here, using two murine models, we investigated the role of SAA in mediating host responses and metabolism and the intestinal microbiota. In germ-free (GF) and conventional (CONV) C57BL/6N mice, we assessed ileal and colonic SAA expression over time. Then we examined microbiota perturbation and microbiota-mediated phenotypes comparing wildtype (WT) and SAA1/2-/- (KO) mice derived from Het/Het crosses to ensure common ancestry. WT and KO mice were either reared separately or co-housed post-weaning. Dextran sodium sulfate (DSS) colitis was induced in a separate cohort of WT and KO mice. We monitored body weight and colitis development, characterized the fecal microbiome, and measured lipid levels and expression of inflammatory marker genes. GF and CONV mice expressed differential microbiota-dependent regulation of SAA subtypes in the ileum and colon. SAA expression was depressed in the absence of microbes in the ileum but increased in the colon. SAA KO led to significantly higher post-weaning weight gain (with effects in females>males), and cohousing diminished this effect, suggesting a critical role of gut microbes. Microbial community structure differed in the WT and KO mice, with specific taxonomic differences. In the KO mice, expression of genes indicative of adipocyte differentiation and inflammation was increased in white adipose tissues. DSS treatment led to increased colonic shortening and slower recovery in the KO mice. Our results suggest that intestinal SAAs have major effects on local inflammation and microbiome characteristics, and mediate microbiota-dependent effects on host metabolism through alterations of inflammatory cytokines, adipogenesis, and lipid cycling.

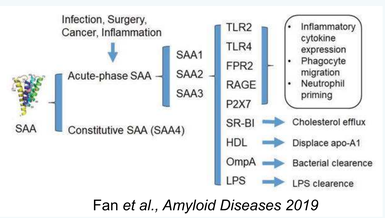
Introduction

SAA proteins comprise a family of apolipoproteins, whose gene sequences are highly conserved across vertebrates, suggesting important biological functions. The acute-phase SAA isoforms, SAA1 and SAA2, are markedly induced by diverse inflammatory stimuli. Colonizing germ-free (GF) mice with segmented filamentous bacteria (SFB) upregulated SAA1/2 expression, which then promoted Th17 cytokine production. SAA1/2 proteins also have been suggested to play a role in intestinal immune homeostasis, but their exact interaction with the microbiota remains understudied.

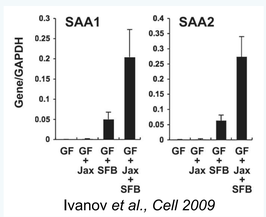


Mouse chromosome 7p

Uhlar & Whitehead, *Eur J Biochem* 1999



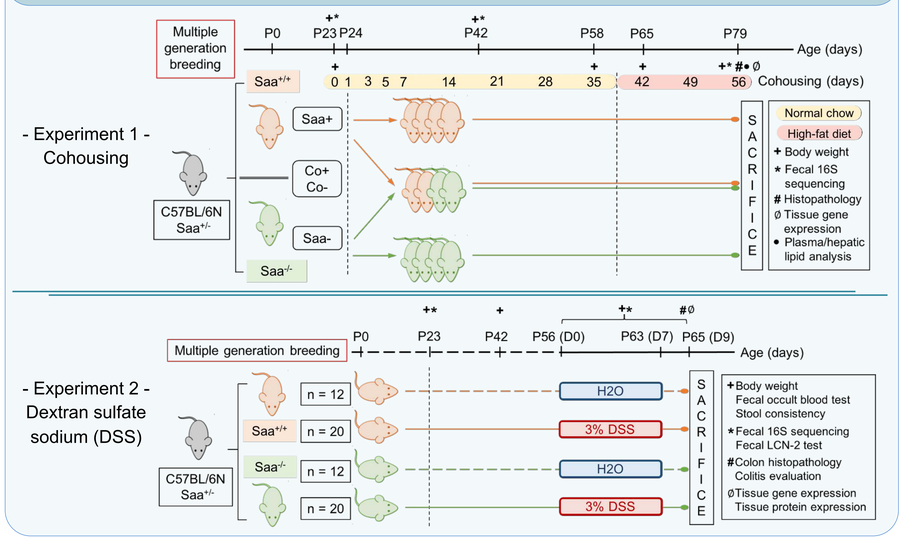
Fan *et al., Amyloid Diseases 2019*



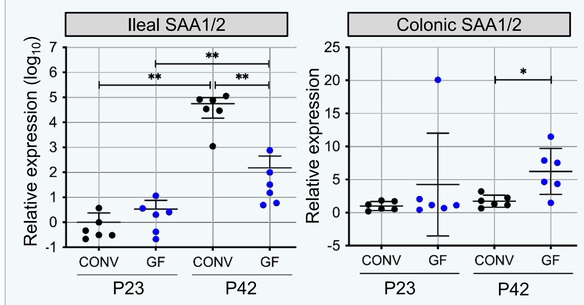
Ivanov *et al., Cell 2009*

Objective: To assess the effects of SAA1/2 deletion on mouse metabolism, gut microbiota, and disease outcome of DSS-induced colitis.

Methods

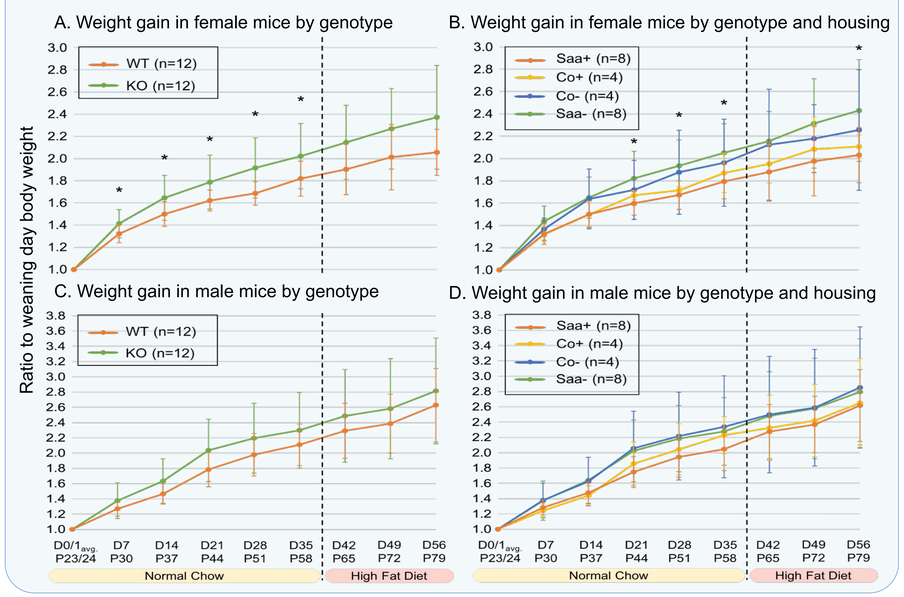


Results

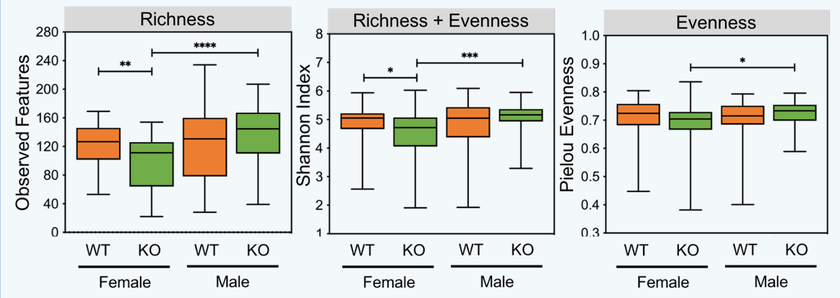


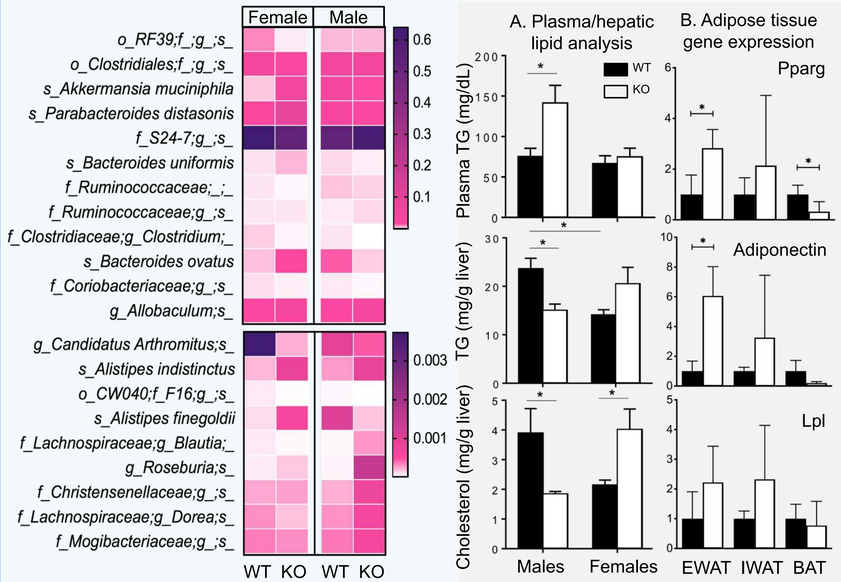
← Murine SAA1/2 expression was remarkedly depressed in the absence of microbes in the ileum but increased in the colon. The microbial regulation predominantly occurred at P42 and in the ileum.

↓ SAA KO led to significantly higher post-weaning weight gain, with greater effects in females than in males.

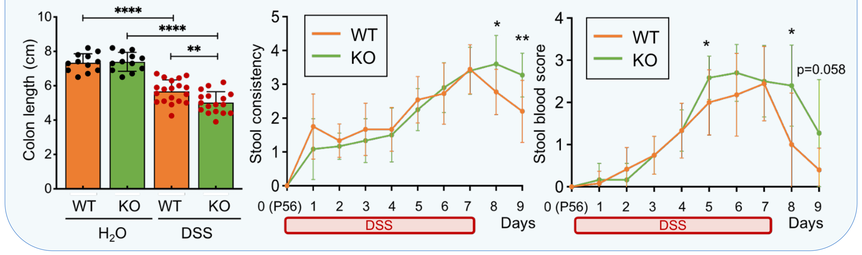


SAA KO markedly reduced species richness and evenness in females. Male KO mice showed greater microbiota alpha diversity than female KO but no genotype effects. Specific taxa changed in abundance in the KO, including *Alistipes spp.* and SFB.





DSS-treated SAA KO mice experienced slower recovery post treatment and showed significantly shortened colon length on the day of termination, compared to the WT.



Future Directions

Acute-phase SAAs are major host factors mediating microbiota-dependent effects on host metabolism through alterations of inflammatory cytokines, adipogenesis and lipid cycling. Further investigations of immune-microbiota interactions will deepen insights into the association of SAA with metabolic abnormalities and their pathophysiological impacts on infectious disease.

Send questions to **yy542@cabm.rutgers.edu**

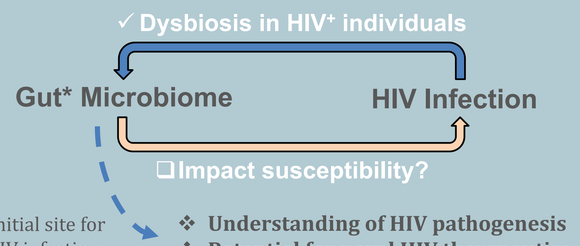
The gut microbiota is associated with HIV acquisition

Danting Jiang[1,2] and Neil Surana[1,3,4]

[1]Department of Pediatrics, [2]Program in Computational Biology and Bioinformatics, [3]Department of Molecular Genetics and Microbiology, [4]Department of Immunology, Duke University, Durham, NC 27710

Introduction

- HIV remains one of the most critical global health problems today.
- HIV infection induces a dysbiosis in the gut microbiota[1]—which plays a vital role in host physiology and alters outcomes of numerous infectious diseases[2]—but the converse of this observation is not yet clear: it remains unknown whether the gut microbiota is causally related to HIV acquisition, and, if so, the specific microbes and mechanisms involved.



Highlights

—Studied a unique NHP HIV vaccine cohort—

- ❖"MIVO6": testing a pediatric vaccine in protecting nursery-reared infant rhesus macaques against oral SHIV challenges, a nonhuman primate [NHP] model for breastfeeding transmission of HIV.
- ❖The vaccine elicited a virus-specific antibody response but conferred no protection.
- ❖There is dramatic variability in time to HIV acquisition across all animals, allowing us to investigate the role of gut microbiota in HIV susceptibility with this cohort.

—Identified 8 taxa related to HIV susceptibility—

- ❖We analyzed 16S rRNA gene sequencing data from this NHP study using a microbe-phenotype triangulation approach we previously developed[3].
- ❖8 bacterial taxa are bioinformatically associated with HIV susceptibility (one has been experimentally validated[4]).

Analyses and Results

- ❖DADA2[5] was used to process the microbiome data (n=292) generated from MIV06 (**Figure 1**).
- ❖ **Vaccination has an impact on the microbiome** from 6 weeks of age onward (**Figure 2**).

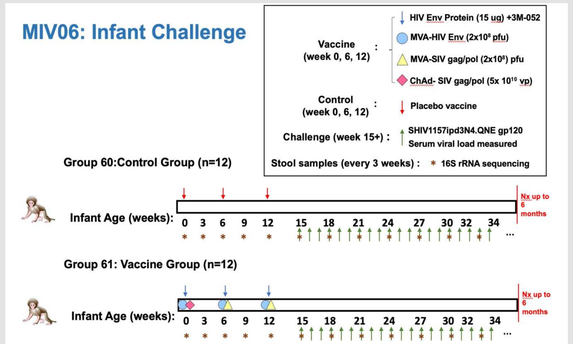


Figure 1. Study design of an NHP vaccine study that allows us to investigate the role of gut microbiota in HIV acquisition.

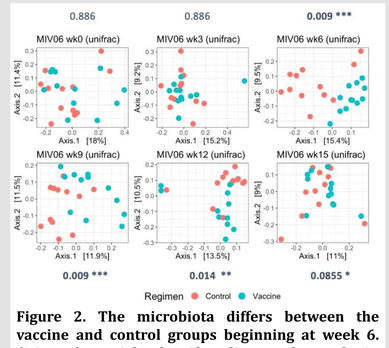


Figure 2. The microbiota differs between the vaccine and control groups beginning at week 6. (PCoA of unweighted Unifrac distance for weeks 0–15 with adjusted p-values from PERMANOVA)

- ❖We utilized **microbe-phenotype triangulation**, an approach we developed that identifies with high specificity microbes causally related to a phenotype of interest[3], based on:
  1) increase in HIV challenge dose (**Figure 3a; 3c row 1 and 2**)
  2) distribution of number of challenges to infection (**Figure 3b; 3c row 3 and 4**).
- ❖We used DESeq2[6] to detect differentially abundant taxa in each comparison and identified a total of **8 taxa that are bioinformatically associated with time to HIV acquisition** (**Figure 3c; Table 1**).
- ❖Notably, *Lactobacillus gasseri*, one of the protective taxa, has been experimentally validated as inhibiting in vitro HIV infection of human tissue [4].

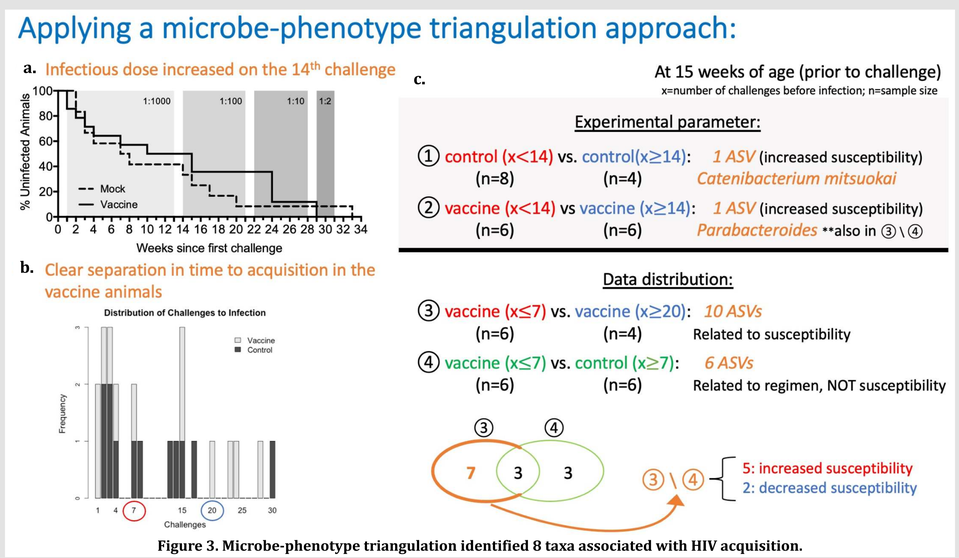


Figure 3. Microbe-phenotype triangulation identified 8 taxa associated with HIV acquisition.

Future Directions

- ➢ Determine the causal effect of these bioinformatically identified taxa on HIV infectivity (**Table 1**) using a previously established HIV infection model for human pediatric tonsillar cells.
- ➢ Compare metagenomics and targeted metabolomics data (short chain fatty acids, bile acids, etc.) to identify microbial features related to acquisition of HIV at a more granular level (i.e., species compositions and functional profiles)

HIV association	Taxa identified (resolution)
Decreased Susceptibility	*L. gasseri* (species)
Increased Susceptibility	*Lachnospiraceae* (family)
	Parabacteroides (genus)
	Bacteroides (genus)
	Colidextribacter (genus)
	Solobacterium (genus)
	Christensenellaceae (family)
	Catenibacterium mitsuokai (species)

Table 1. Taxonomy of the 8 bioinformatically identified ASVs.

Literature Cited

1. Williams B, Landay A, Presti RM. Microbiome alterations in HIV infection a review. *Cell Microbiol.* 2016;18(5):645-651.
2. Libertucci J, Young VB. The role of the microbiota in infectious diseases. *Nat Microbiol.* 2019;4(1):35-45.
3. Surana NK, Kasper DL. Moving beyond microbiome-wide associations to causal microbe identification. *Nature.* 2017;552(7684):244-247.
4. Nahui Palomino RA, Vanpouille C, Laghi L, et al. Extracellular vesicles from symbiotic vaginal lactobacilli inhibit HIV-1 infection of human tissues. *Nat Commun.* 2019;10(1):5656.
5. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13(7):581-583.
6. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

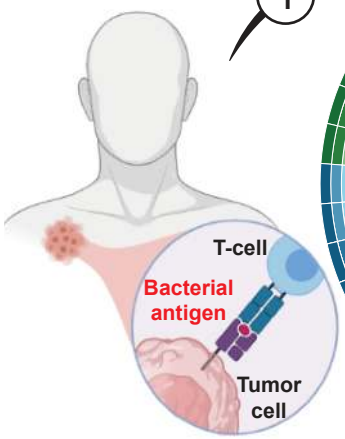
Acknowledgements

- HIV Vaccine Research and Design (HIVRAD) Program (P01)
- Neil Surana Lab, Duke University
- Sallie Permar Lab, Duke University-Weill Cornell Medicine

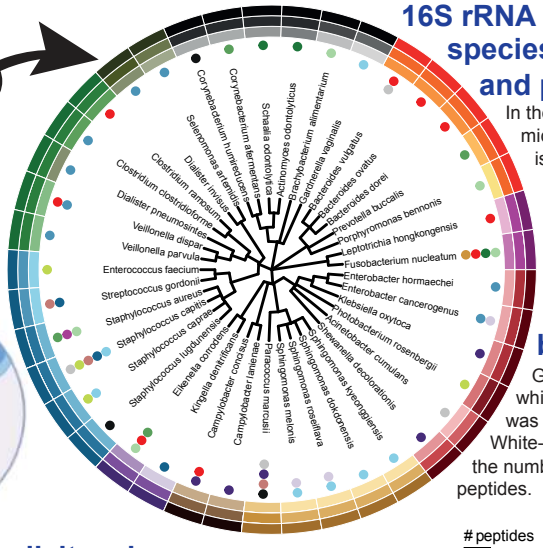
Identification of bacteria-derived HLA-bound peptides in melanoma

Shelly Kalaora, Adi Nagler, Deborah Nejman, Michal Alon, Chaya Barbolin, Eilon Barnea, Steven L. C. Ketelaars, Kuoyuan Cheng, Kevin Vervier, Noam Shental, Yuval Bussi, Ron Rotkopf, Ronen Levy, Gil Benedek, Sophie Trabish, Tali Dadosh, Smadar Levin-Zaidman, Leore T. Geller, Kun Wang, Polina Greenberg, Gal Yagel, Aviyah Peri, Garold Fuks, Neerupma Bhardwaj, Alexandre Reuben, Leandro Hermida, Sarah B. Johnson, Jessica R. Galloway-Peña, William C. Shropshire, Chantale Bernatchez, Cara Haymaker, Reetakshi Arora, Lior Roitman, Raya Eilam, Adina Weinberger, Maya Lotan-Pompan, Michal Lotem, Arie Admon, Yishai Levin, Trevor D. Lawley, David J. Adams, Mitchell P. Levesque, Michal J. Besser, Jacob Schachter, Ofra Golani, Eran Segal, Naama Geva-Zatorsky, Eytan Ruppin, Pia Kvistborg, Scott N. Peterson, Jennifer A. Wargo, Ravid Straussman & Yardena Samuels

9 melanoma patients 17 metastases



16S rRNA sequencing show similarity of bacterial species composition between metastases and patients

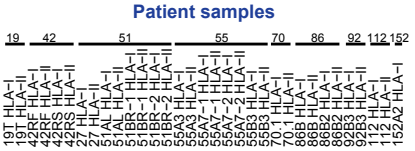


In the pylogenetic tree the identified bacterial species indicated in the middle and the circles indicate the genus, order and class. Each patient is color coded, metastases are marked with different shades.

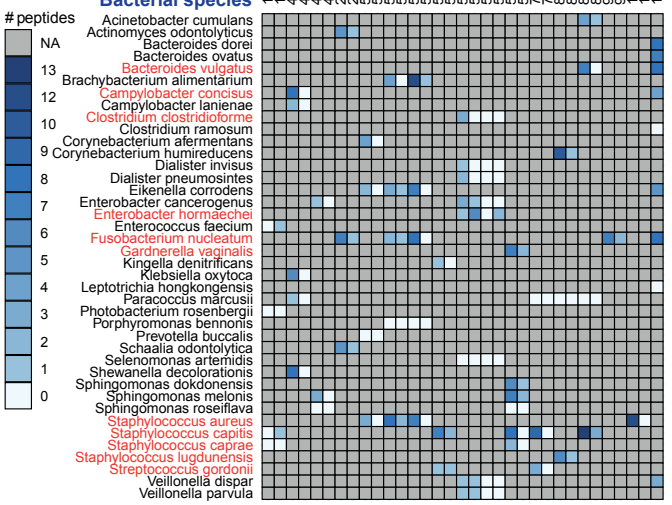
19	42	51	55	70	86	92	112	152
19T	42RF 42RS	27 51AL 51BR	55A3 55A7 55B3	70.1	86B 86B2	92B3	112	152A2

Using HLA peptidomics we were able to identify 248 HLA-I and 38 HLA-II bacterial peptides

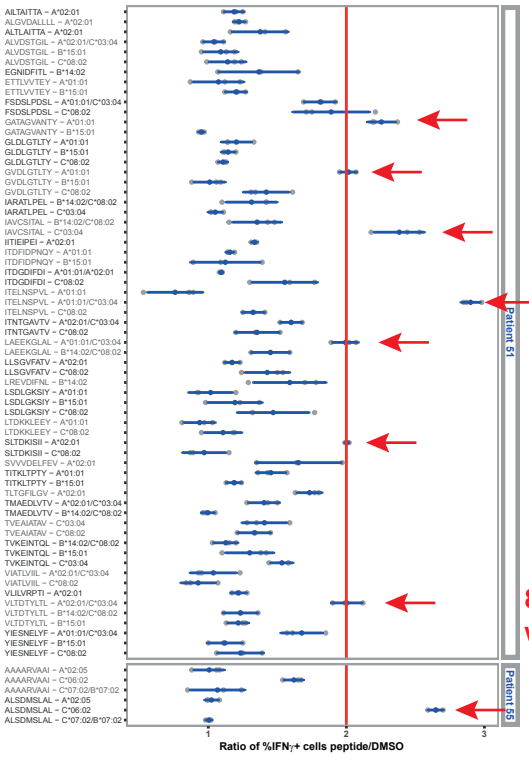
Gray squers indicat in which samples bacteria was not identified. White-blue scale indicates the number of identified peptides.

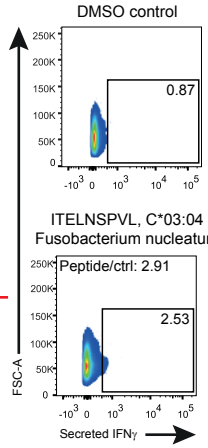


Bacterial species



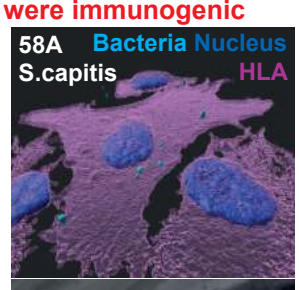
Bacterial peptides can elicit an immune responce by tumor infiltrating lymphocytes (TILs)

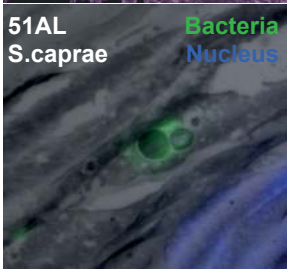


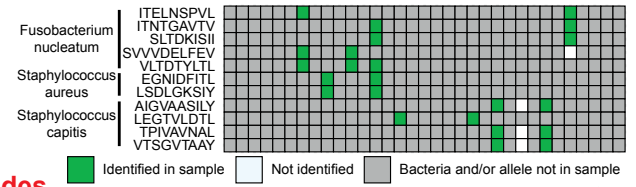


Shared bacterial peptides and peptides derived from Fusobacterium nucleatum and Staphylococcus aureus were tested for immunogenicity with the autologous TILs.

8 bacterial HLA-I peptides were immunogenic







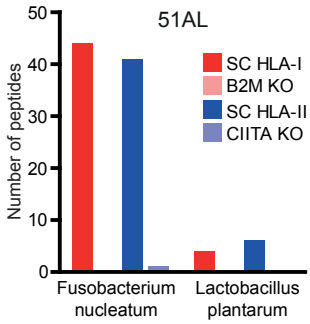
11 bacterial HLA-I peptides were shared between metastases and patients

Bacteria peptides are presented by tumor cells and antigen presenting cells

Bacterial species	HLA class	Peptide	CD45–	CD45+
Bacillus megaterium	HLA-I	EELSRQNL		■
		ILDSRGEEIY	■	
		NDQALQPLF		■
	HLA-II	PQVILRRMR	■	
Clostridium intestinale	HLA-I	LSNAKSLEL	■	■
		LSTNSIVSL	■	

As we used bulk tumors we could not indicate if bacterial peptides were derived from tumor cells or antigen presenting cells. We isolated digested single cell fresh tumor based on the CD45 marker to seperate between immune and non-immune cells. As indicated in the table, bacterial peptides were identified in both populations with some overlap.

Bacterial peptides are HLA ligands



To validate that the bacterial peptides are HLA ligands, we knocked out B2M and CIITA to reduce HLA-I and HLA-II expression. Similarly, we infected the cells with less invasive bacteria, and observed less bacterial peptides.

Bacteria isolated from tumors can invade tumor cells

We validated that the identified bacterial species, isolated from tumors can invade melanoma cell lines that were derived from the same tumors by immunoflourecence and 3D representation and electron microscopy.

Characterization of a novel microbiome marker anti-correlated with *S. aureus* carriage in the infant nasal microbiome

Madeleine C. Kline[1,4], Emma K. Accorsi[2,4], Eric A. Franzosa[2,3,4], Curtis Huttenhower[2,3,4]

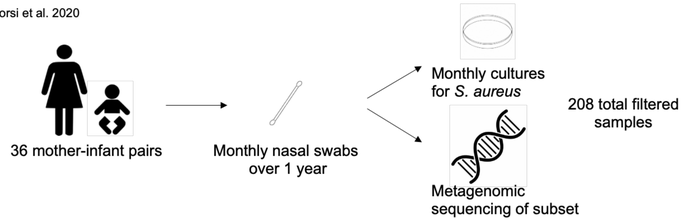
[1]Harvard Medical School [2]Harvard T.H. Chan School of Public Health [3]Broad Institute of MIT and Harvard
[4]Harvard Chan Microbiome in Public Health Center

Staphylococcus aureus carriage in the nasal microbiome is an important determinant of subsequent *S. aureus* soft tissue infection. Identifying elements of the nasal microbiome that influence carriage provides insight on how to modulate these factors to prevent progression to infection. Prior work by Accorsi et al. 2020[1] discovered an uncharacterized, taxonomically unassigned ORF that was the major predictor of whether infant microbiome samples evaluated with shotgun metagenomic sequencing contained *S. aureus.* Subsequent investigation indicated that this ORF was actually a segment of 16S rRNA gene sequence incorrectly annotated by UniProt as a protein-coding gene. Our analyses provide substantial initial support for the hypothesis that this sequence represents a phylogenetic marker for a novel clade with genomic similarity to *Streptococcus* species and *D. pigrum*, which could antagonize *S. aureus* during colonization of the infant nasal microbiome.

Study Design and Methods

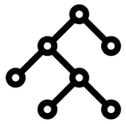
Data collection

Accorsi et al. 2020



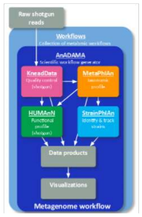
Random Forest Model

Identified sequence (UniRef90 of Interest or **UOI**) that predicted *S. aureus* carriage



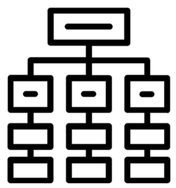
bioBakery version 3 analysis

Updated functional and taxonomic profiling tools for higher resolution classification of UOI and correlated UniRefs



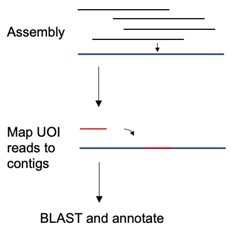
Covariation-based genome reconstruction

- MSPminer and MetaBAT try to characterize reads related to the UOI
- Analysis of *S.aureus* dataset and 267 nares samples in Expanded Human Microbiome Project (**HMP**)



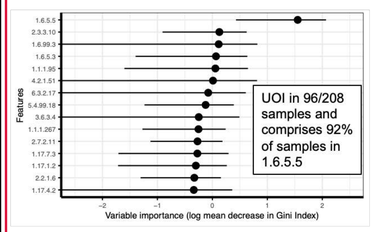
Analysis of UOI-adjacent sequences

In *S.aureus* and HMP datasets

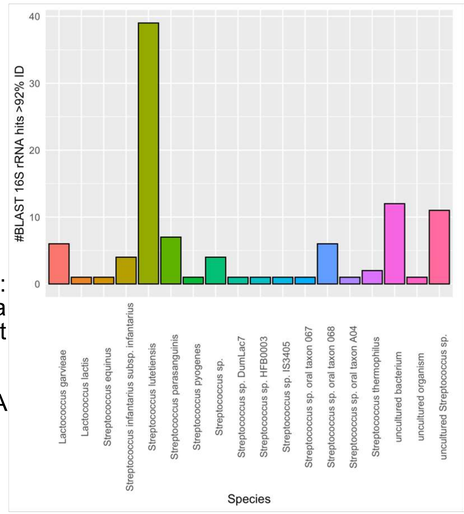


A sequence marker of a novel clade predicts *S. aureus* exclusion

Top 15 Random Forest Predictors



UOI BLAST Results

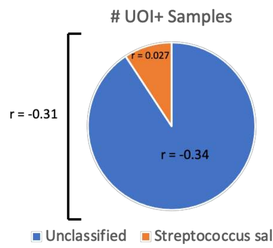


UOI (UniRef of Interest: UniRef90_X5NU12), a 42 aa sequence, is the strongest predictor of S. aureus carriage.

It is homologous to 16s rRNA from many species.

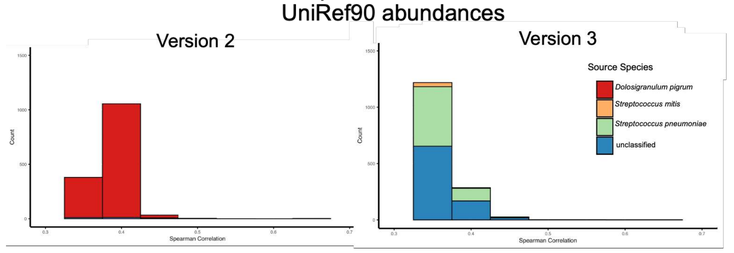
Updated taxonomic and functional profiling of *S.aureus* dataset

Spearman correlation coefficients: UOI to *S. aureus* relative abundance



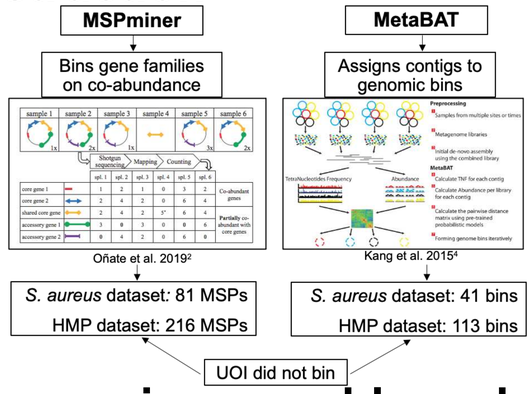
UOI partially classifies to *S. salivarius* in updated analysis, but only **unclassified UOI is signficantly negatively correlated with *S. aureus*.**

Correlations of UOI with species-stratified UniRef90 abundances



D. pigrum gene families from version 2 were still present in version 3, but r ~ 0.28, likely due to increased resolution of *Streptococcus* species in version 3.

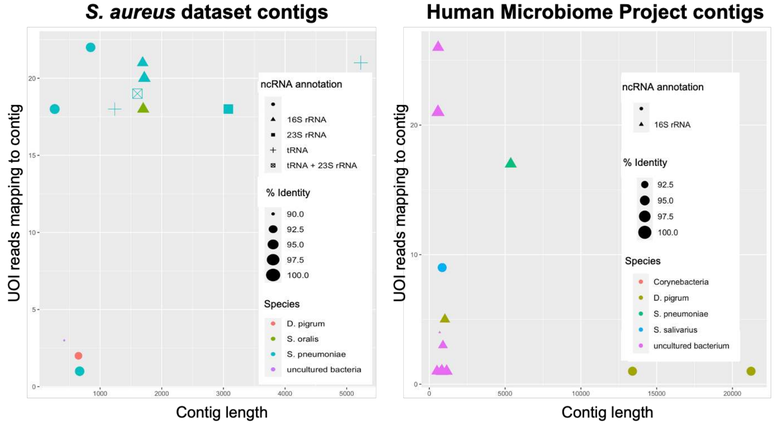
Covariation-based genome reconstruction



Metagenomic assembly and contig analysis

Assembled contigs containing UOI reads analyzed for length, # of UOI reads, BLAST taxonomic results and % identity, and rRNA or tRNA annotation.

***S. aureus* dataset contigs** | **Human Microbiome Project contigs**



Conclusions and future work

Our work supports the hypothesis that the UOI is a fragment of 16S rRNA that represents a novel species. It is genomically similar to *D. pigrum* and *Streptococcus* species. Further in vitro experimentation, including amplification the sequence from related bacteria via RT-qPCR could help to characterize this novel clade.

Sources

1. Accorsi, E. K. et al. Determinants of Staphylococcus aureus carriage in the developing infant nasal microbiome. Genome Biol. 21, 301 (2020).
2. Plaza Oñate, F. et al. MSPminer: abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data. Bioinformatics 35, 1544–1552 (2019).
3. Kenny, D. J. et al. Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. Cell Host Microbe 28, 245-257.e6 (2020).
4. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165 (2015).
Images from the Noun Project by Yu Luck, Knut M. Synstead

Acknowledgments

We appreciate the valuable contributions of Yancong Zhang and Lauren McIver. Thank you also to the Harvard-MIT MD-PhD program.

http://huttenhower.sph.harvard.edu

Instability, Heterogeneity, and Pathogenicity Reservoirs in the Skin, Oral, and Gut Microbiota of Older Adults

UCONN SCHOOL OF MEDICINE

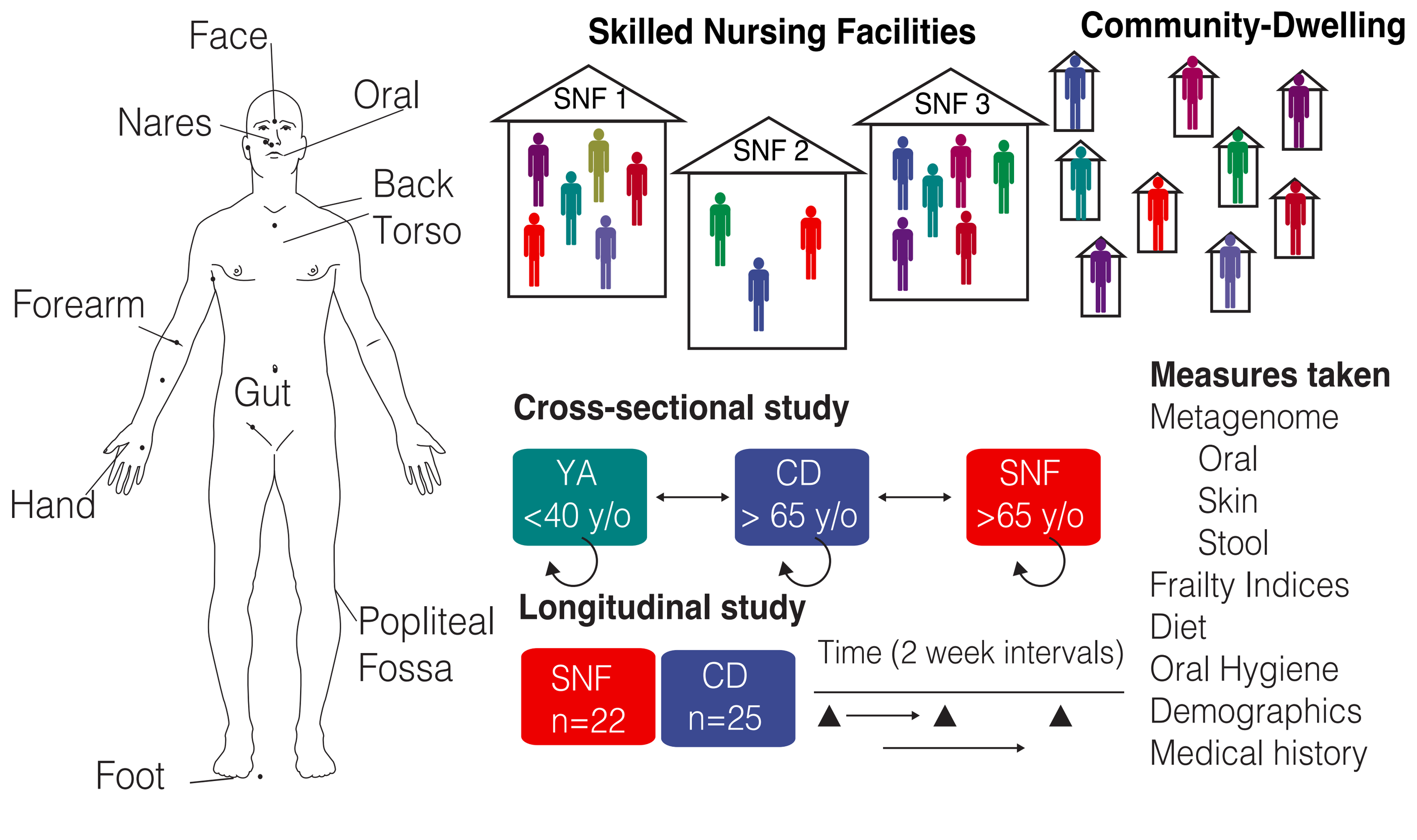
The Jackson Laboratory — Leading the search for tomorrow's cures

Peter Larson[1,2], George Kuchel MD[1], James Grady PhD[1], Julie Robison PhD[1], Julia Oh PhD[2].

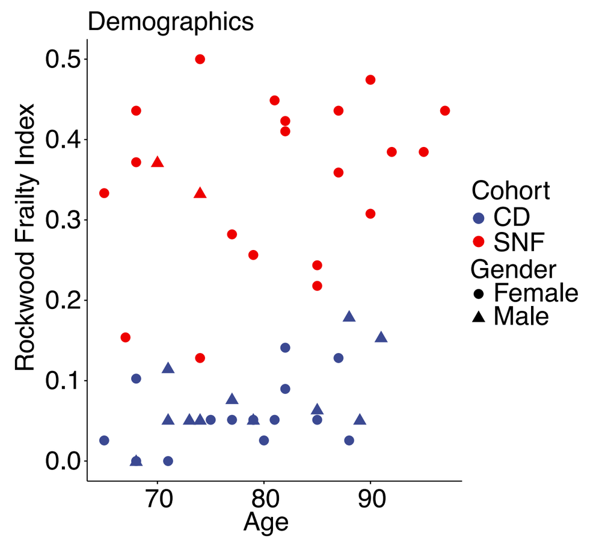
1. UCONN Health (University of Connecticut), Farmington, CT. 2. The Jackson Laboratory for Genomic Medicine, Farmington, CT.

Introduction

Despite their elevated risk for morbidity and mortality from infections, the microbiome of older adults remains understudied. While colonization resistance from resident microflora is a promising means to prevent infections, little is known about pathogenicity reservoirs and colonization resistance in this vulnerable population. We studied the skin, oral, and gut microbiome dynamics of older adults in both community and Skilled Nursing Facility (SNF) settings, investigating relationships between age, frailty, environment, microbiota, and pathogenicity reservoirs.



Demographic Characterisitics	CD	SNF
Total No.	25	22
Age range, y	65-91	65-97
Age, y, Mean (SD)	78.2(7.59)	82.9 (8.46)
Sex, No. (Self-Reported)		
Male Participant	11	2
Female Participant	14	20
Race		
American Indian or Alaska Native	1	0
White	24	20
More than one race	0	2
Ethnicity		
Hispanic or Latino	1	2
Non-Hispanic or Latino	24	18
Height, mean (SD), cm	167 (11.9)	160 (8.5)
Weight, mean (SD), kg	69.0 (15.4)	78.7 (23.9)



We conducted a longitudinal metagenomic whole genome shotgun survey of 47 adults age 65+ years of age; 22 residents of 3 different SNFs and 25 community dwelling individuals. We performed metagenomic whole genome shotgun sequencing on stool, oral, and skin samples from 8 sites, 1421 total. To correlate clinical and behavioral variables, we measured frailty, collected medical records, and interviewed participants on diet and lifestyle. We also draw comparisons with previous younger cohorts[1-3].

Results

Instability, Heterogeneity, Diversification, and Biogeography in the Aging Skin Microbiome

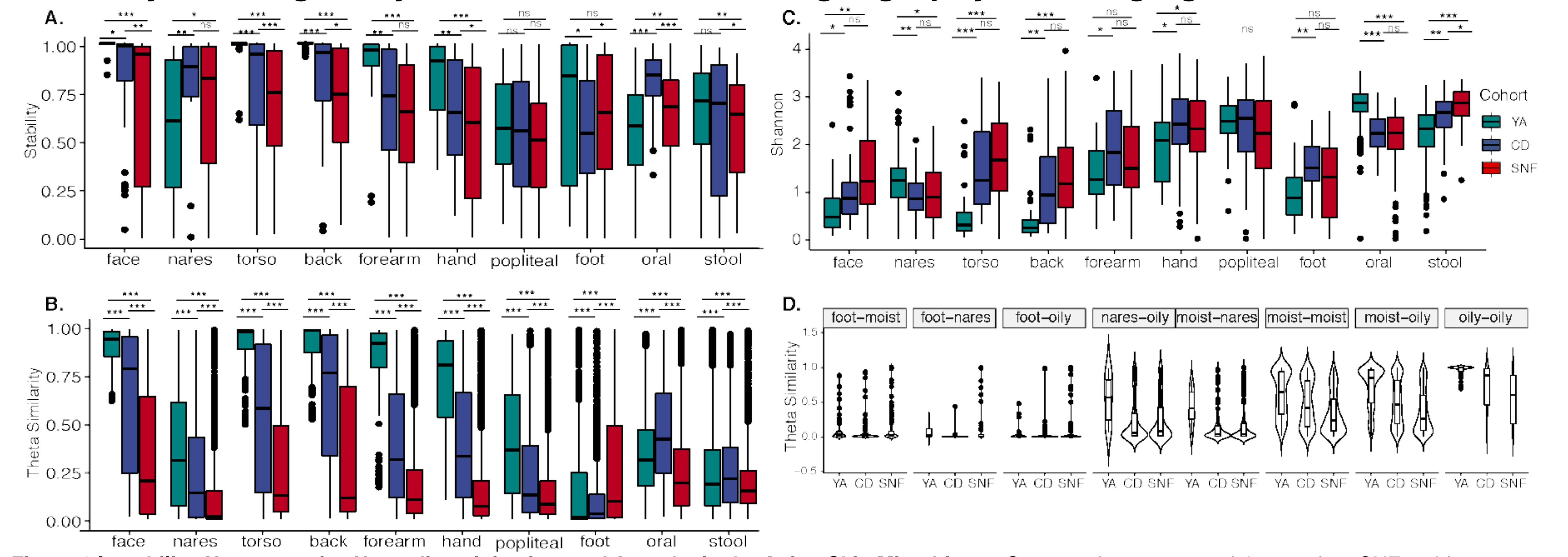


Figure 1 Instability, Heterogeneity, Hyperdiversigication, and Anarchy in the Aging Skin Microbiome. Compared to younger adults, or when SNF residents are compared to Community-Dwelling older adults the taxonomic composition of the skin microbiota was generally characterized by:

A. Decreased stability over time. Yue-Clayton Theta Index comparing samples from an individual at different timepoints. Where most younger adult skin sites are relatively constant overtime, the skin microbiota of older and frailer adults appears to vary substantially.

B. Decreased inter-individual similarity. Yue-Clayton Theta Index comparing samples between individuals in each cohort. Older and frailer adults exhibited far less skin microbiome similarity to their peers than younger adults, demonstrating higher Heterogeneity.

C. Hyper-diversification (Shannon Index of diversity representing the number and evenness of species). This trend was also observed in the gut.

D.Biogeographic divergence. Biogeographic determinism, site-specific community composition, is a hallmark of the skin microbiota in older adults, but is different in older adults. Yue-Clayton Theta Index comparing samples from different skin sites on the same individual at the same timepoint. Rather than becoming more similar with skin aging, skin sites appear to diverge in the older and frailer cohorts. v

CD=Community-Dwelling; SNF= Skilled Nursing Facility; YA=Younger Adults. Bidirectional Wilcox tests, *=p<0.05, **=p< 0.0005, ***=p<5E-8.

Taxonomic Compositional Differences in the Microbiota of Older Adults.

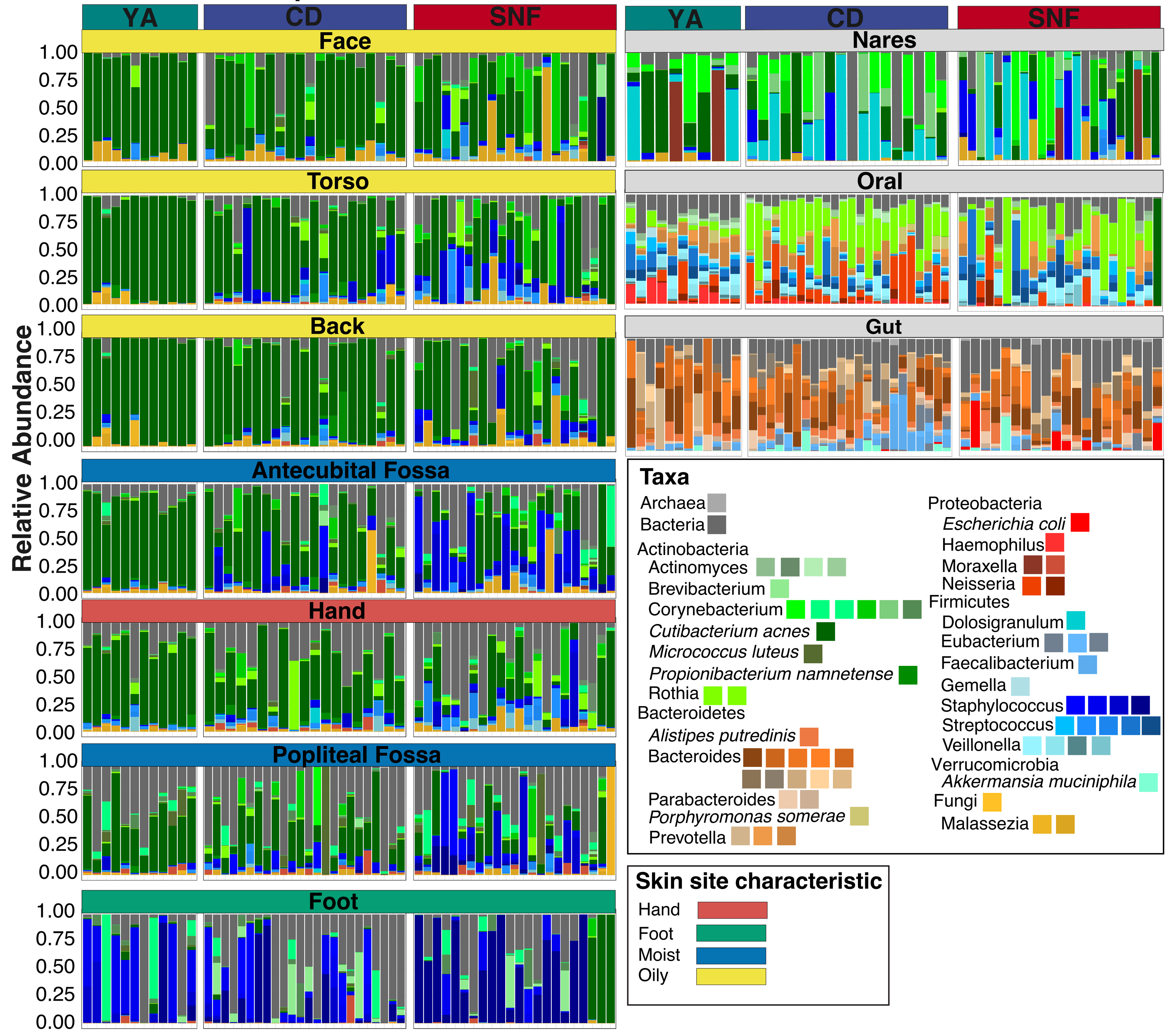


Figure 2: Taxonomic Compositional Differences in the Microbiota of Older Adults. Relative abundance of species according to MetaPhlAn 3.0[4] classification. Each bar represents 1 subject, 1 timepoint represented per subject. Older adults, especially SNF residents, exhibit marked decrease in cutaneous *Cutibacterium acnes* abundance, with a reciprocal increase in Staphylococci, Corynebacteria, and in some cases *Malassezia* and oral species. High inter-individual heterogeneity in older cohorts is also evident here. Oral (tongue dorsum) had notably higher abundance of *Rothia* species, and notably less Proteobacteria in the SNF cohort. Gut microbiota of SNF residents had a higher Firmicutes:Bacteroides Ratio, and in many cases increased Proteobacteria and decreased *Akkermansia mucinophila*.

Associations Between Age, Frailty, and Structural Differences in the Skin Microbiome

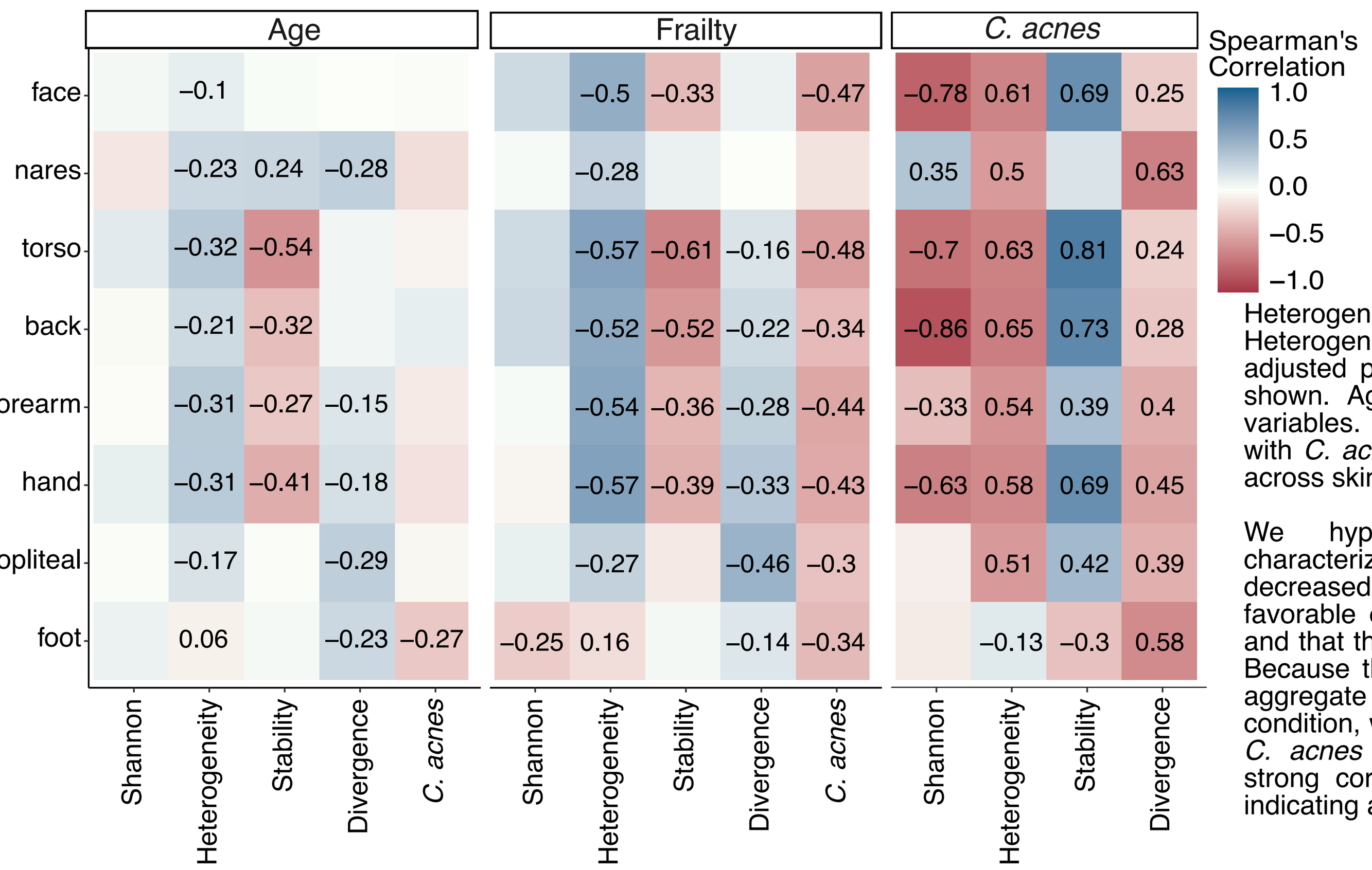


Figure 3: Associations Between Age, Frailty, and Structural Differences in the Skin Microbiome. Mixed Effects Model controlling for temporal pseudoreplication testing linear correlation by skin site between age, frailty (Rockwood Index), or *C. acnes* abundance and Shannon Diversity Index, Inter-individual Heterogeneity, Stability, or Intra-individual Heterogeneity (Anarchy). Only significant (fdr adjusted p < 0.05) Spearman's Coefficients shown. Age was a poor predictor of all variables. Frailty was negatively correlated with *C. acnes* abundance and heterogeneity across skin sites.

We hypothesized that skin aging, characterized by follicular atrophy and decreased sebum production, creates a less favorable environment for *C. acnes* growth, and that the resulting decrease is correlated. Because the Rockwowrd Frailty Index is an aggregate score not directly indicating skin condition, we tested the relationship between *C. acnes* and the other variables, finding strong correlations at nearly all skin sites, indicating a pattern of dysbiosis.

Skin a Reservoir for Specific Pathobionts in Older Adults

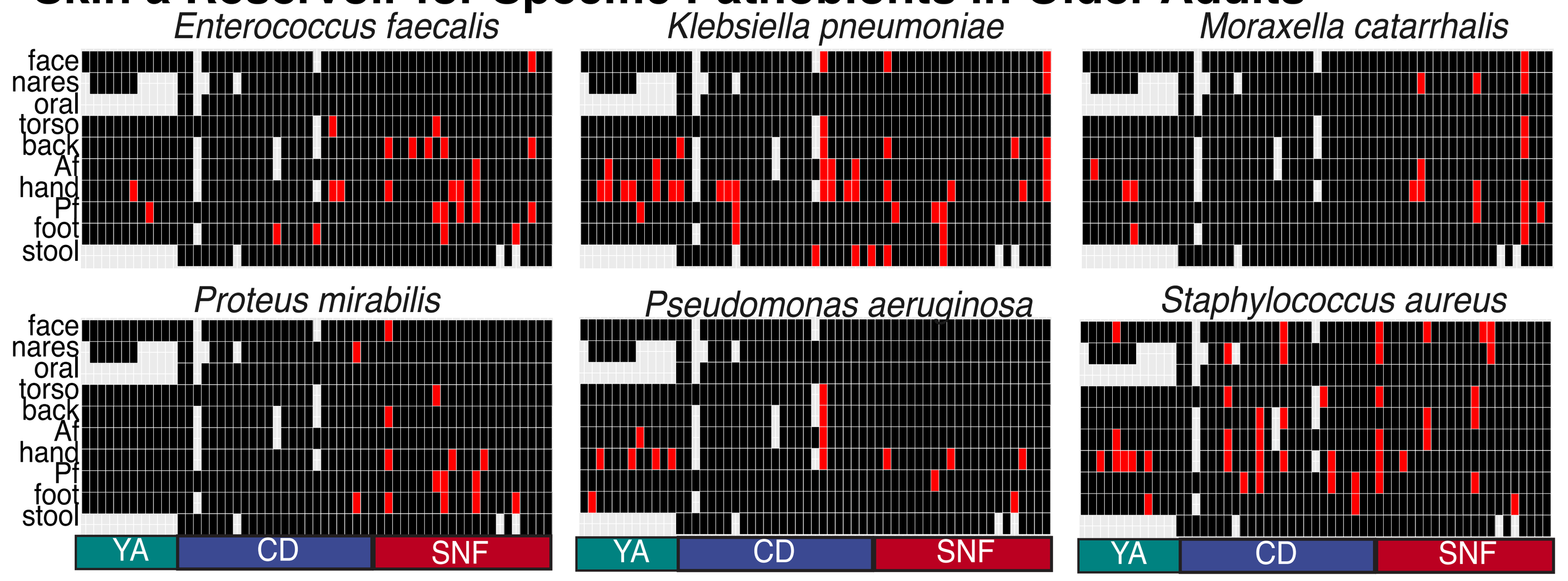


Figure 4: Skin a Reservoir for Specific Pathobionts. Presence/Absence heatmap of specific pathobionts identified with metagenomic shotgun sequencing. This method is less sensitive for low-abundance pathogens than culture-based screening, but still found clinically significant pathobiont colonization in the skin more frequently than oral or gut. Red = present; black = absent.

Skin Major Reservoir of Plasmid Anti-Microbial Resistance in Older Adults

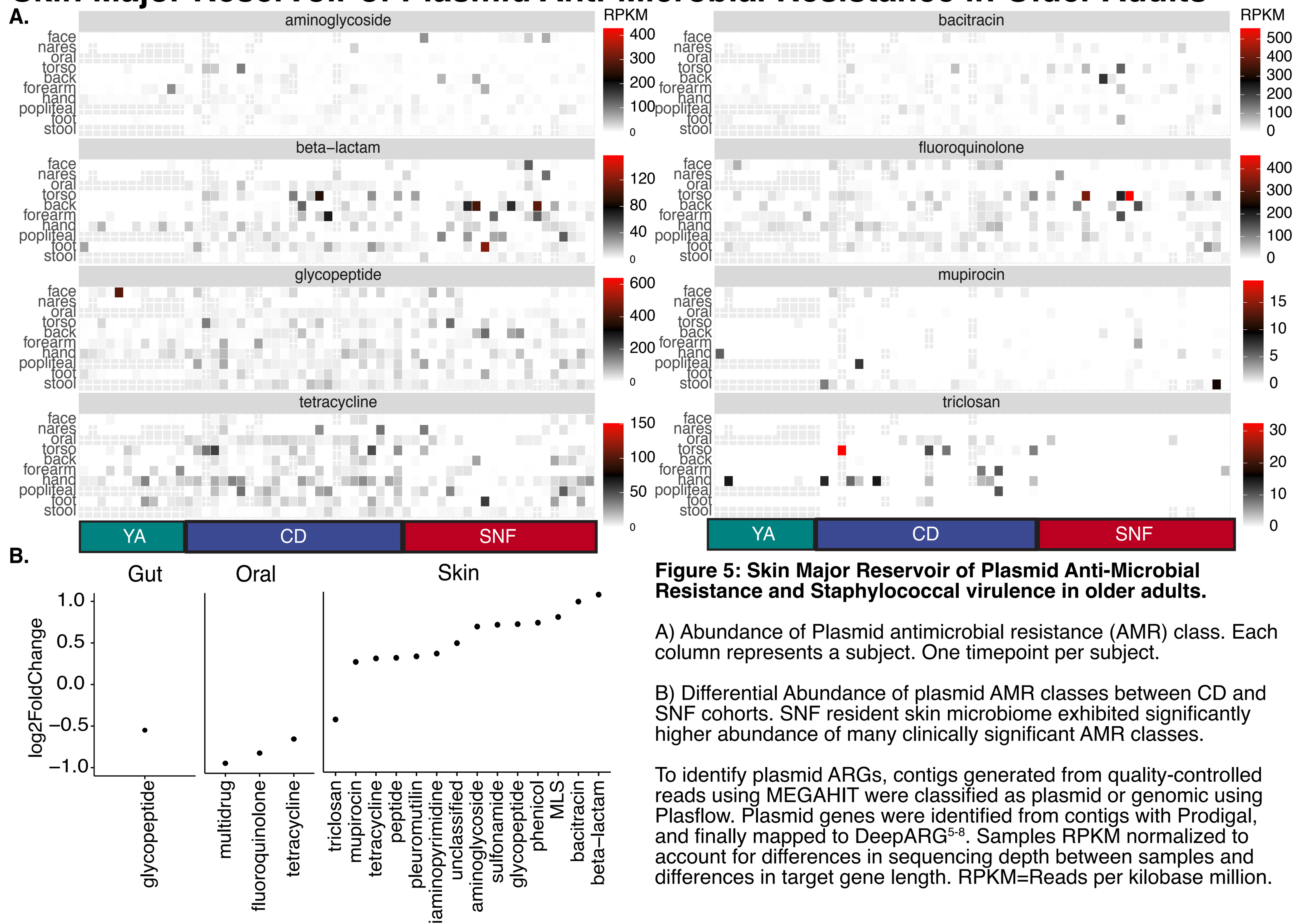


Figure 5: Skin Major Reservoir of Plasmid Anti-Microbial Resistance and Staphylococcal virulence in older adults.

A) Abundance of Plasmid antimicrobial resistance (AMR) class. Each column represents a subject. One timepoint per subject.

B) Differential Abundance of plasmid AMR classes between CD and SNF cohorts. SNF resident skin microbiome exhibited significantly higher abundance of many clinically significant AMR classes.

To identify plasmid ARGs, contigs generated from quality-controlled reads using MEGAHIT were classified as plasmid or genomic using Plasflow. Plasmid genes were identified from contigs with Prodigal, and finally mapped to DeepARG[5-8]. Samples RPKM normalized to account for differences in sequencing depth between samples and differences in target gene length. RPKM=Reads per kilobase million.

Summary

We conducted a novel, longitudinal, gut, oral, and skin metagenome whole genome shotgun study of older adults both in skilled nursing facilities and living privately in the greater community. To the best of our knowledge, this is also the largest report to date of the skin metagenome in older adults.

We found that in particular the skin microbiota of older adults are substantially different to those of younger adults. In particular, we found:

- Major compositional differences between healthy older adults and younger adults, as well as SNF residents to Community-Dwelling older adults including:
  - o Decreased relative abundance of *C. acnes*
  - o Increased Staphylococci, Corynebacteria, fungi, and oral species
- Substantially decreased stability of the skin microbiota
- High inter-individual heterogeneity
- Biogeopgrahic divergence t
- Age alone is a poor predictor of these changes.
- There are strong correlations between decreased *C. acnes* abundance and instability, hyper-diversification, and hyper-heterogeneity. This indicates a pattern of dysbiosis.
- The skin microbiome in older adults, and particularly SNF residents, serves as a major reservoir of clinically important pathobionts and antimicrobial resistance.

Conclusions

Although preliminary, we believe that these results represent foundational findings in our understanding of the microbiota of older adults. In particular, they demonstrate dramatic differences in the skin microbiome among older adults. We suspect that skin aging is a key driver in these changes, adversely affecting C. acnes and leading to a breakdown in community structure, although this possibility cannot be directly addressed by our dataset and must be a subject of future research. Most importantly, our findings draw attention to the skin as potentially a more important reservoir than the oral and gut microbiota for clinically relevant pathogens and antimicrobial resistance.

Acknowledgements

We would like to express our deep appreciation for the contributions of:
- Alba Santiago and Sarah Driscoll from the UConn Center on Aging for their assistance in subject recruitment and sampling.
- Member of the Oh lab, especially Elizabeth Fleming for their support in the preparation of sample collecting kits and the preparation of samples for sequencing.
- The Jax-GM Microbial Genomic Services and Genomic Technologies Cores.
- Our partner SNFs: Riverside Health and Rehabilitation, Skyview Center, and Manchester Manor Health Care Center, including all of the staff who assisted in coordination and implementation.
- All of our research subjects who volunteered their time and microbiota samples for the advancement of science.

References

1. Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA. Biogeography and individuality shape function in the human skin metagenome. Nature. 2014;514(7520):59-64. Accessed Apr 1, 2018. doi: 10.1038/nature13786.
2. Oh J, Byrd AL, Park M, Kong HH, Segre JA. Temporal stability of the human skin microbiome. Cell. 2016;165(4):854-866. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4860256/. Accessed Jul 14, 2020. doi: 10.1016/j.cell.2016.04.008.
3. Peterson J, Garges S, Giovanni M, et al. The NIH human microbiome project. Genome Res. 2009;19(12):2317-2323. Accessed Nov 30, 2017. doi: 10.1101/gr.096651.109.
4. Segata N, Waldron LD, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods. 2012;9(8):811-814. http://nrs.harvard.edu/urn-3:HUL.InstRepos:11369874. Accessed Jul 18, 2018.
5. Li D, Luo R, Liu C, et al. MEGAHIT v1. 0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016;102:3-11.
6. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: Predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res. 2018;46(6):e35-e35. https://academic.oup.com/nar/article/46/6/e35/4807335. Accessed Sep 2, 2020. doi: 10.1093/nar/gkx1321.
7. Hyatt D, Chen G, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics. 2010;11(1):119. http://www.ncbi.nlm.nih.gov/pubmed/20211023. doi: 10.1186/1471-2105-11-119.
8. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 2018;6(1):1-15.

Probiotic Supplementation and Marathon Runners: there are any effect up to Gut Microbiota and neutrophil function?

Geovana S F Leite[1], Ayane S Resende[5], Edgar Tavares[3], Helena A P Batatinha[2], Ricardo A Fock[4], José C R Neto[2], Ronaldo V T dos Santos[3], Antonio H Lancha Junior[6].

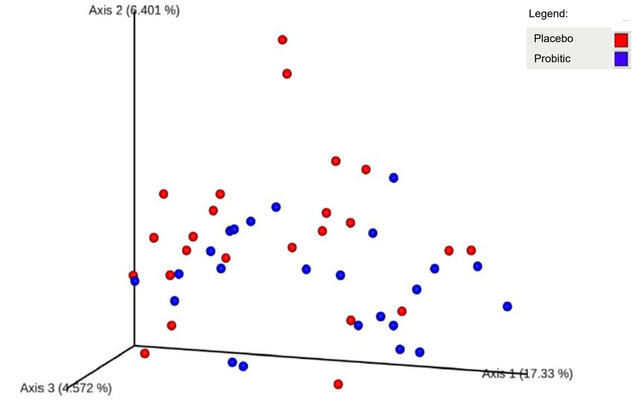
1.Department of Biodynamics of the Movement of the Human Body, School of Physical Education and Sports, University of São Paulo, São Paulo-SP, Brazil. 2- Department of Cell and Tissue Biology, Institute of Biomedics Science, University of São Paulo, São Paulo-SP, Brazil ; 3- Department of Psychobiology, Federal University of São Paulo, Brazil; 4- Departament of Clinical and Toxicological Analysis, Faculty of Pharmaceutical Sciences, University of São Paulo, São Paulo-SP, Brazil; 5- Department of Health Science, Federal University of Sergipe, Aracaju, Sergipe – Brazil; 6- Laboratory of clinical Iinvestigation: Experimental surgery, LIM26, Hospital das Clínicas, Faculdade de Medicina, University of são Paulo

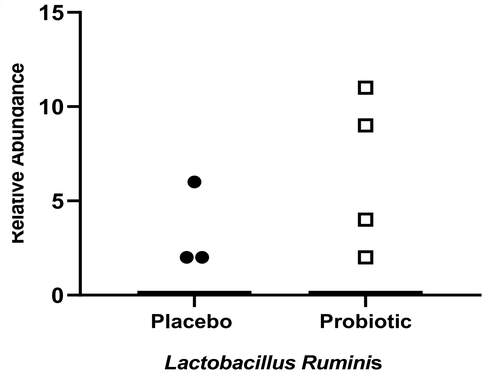
INTRODUCTION Probiotic supplementation can induce positive alterations in intestinal environment, however the effect of a month period short of probiotic supplementation on gut microbiota and neutrophil function of endurance athletes is not known. .

PURPOSE: Investigate the effect of thirty days of probiotic supplementation up on gut microbiota composition and neutrophil function in marathon athletes. .

METHODS : Twenty-seven marathon runners were double-blind randomly assigned to either a Probiotic (PR) (35,96 ± 5,81 years,79,30 ±10,99Kg) or Placebo (PL) group (PL= 40,46 ± 7,79 years, 72,67 ±10,20Kg). PR consumed *Lactobacillus Acidophilus* and *Bifidobacterium Lactis (*10x109UFC + maltodextrin) during 30 days in a sachet form, while PL received a sachet with maltodextrin (5g/day). The gut microbiota composition was evaluated before (BASELINE) and after the supplementation period (POS-SUP). Fiber consumption was evaluated using one-day diet record at the baseline and Pos-sup. Blood collection was realized (BASELINE and POS-SUP) to verify neutrophil function, after blood cell neutrophil isolation peroxide and cytokine production (IL-1-β; TNF-α; IL-6; IL-8) was analyzed. The Bacterial DNA were extracted using QIAamp Fast DNA Stool Mini Kit® and faecal microbiota composition was assessed by 16S rRNA sequencing, V3-V4 regions, with Illumina® MiiSeq plataform. Operational taxonomic units (OTUs) and diversity indices were obtained after bioinformatic treatment on Qiime2® software. β-diversity was computed considering the sampling of 1,800 sequences per sample, which was based on the rarefaction curve. To test differences among groups and time, it was performed a pairwise PERMANOVA for beta-diversity and ANCOM for OTUs relative abundance. Data analyses were conducted using SAS Statistical Software version 9.3® (p< 0.05) and multiple tests corrected when necessary). For neutrophil function was used the of repeated measures statistical test mixed Model (with 'group' and 'time' as factors) being used with Tukey's post hoc - GraphPad Prim8 ®.

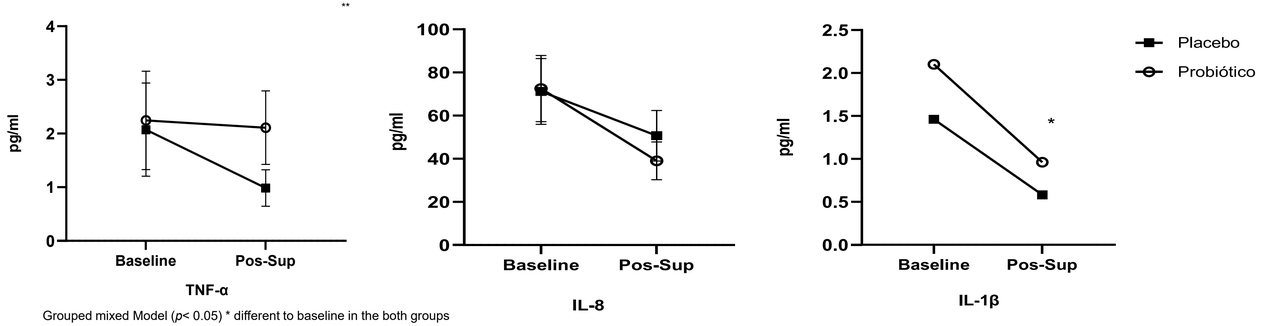
GUT MICROBIOTA COMPOSITION





Beta-diversity, UniFrac *weighted* index was different between groups (p=0,04; PERMANOVA)

NEUTROPHIL CYTOKINE PRODUCTION



Grouped mixed Model (p< 0.05) * different to baseline in the both groups

RESULTS Fiber consumption was similar between groups with no statistical difference [BASELINE: 23,65 ± 10,53 (PL); 24,92± 19,15 (PR); POS-SUP= 25,63± 15,61 (PL); 15,61±15,25(PR)]. The peroxide and cytokines production by neutrophils were no different between groups. From the gut microbiota analyses, it was identified 2.634 OTUs based on 173.096 final sequences. Regarding beta-diversity, UniFrac *weighted* index was different between groups (p=0,04; PERMANOVA) and the relative abundance of *Lactobacillus ruminis* was significant different between groups, in which PR exhibited significant high levels after supplementation period. .

DISCUSSION: Probiotic induced changes in the intestinal environment or increased interaction among specific bacterial species leading to an increase in the relative abundance of lactic acid bacteria, such as *Lactobacillus ruminis*. This effect seems to be a positive change from the supplementation toward athletes' health, since this specie is a probiotic bacteria known for its immunomodulatory activity

CONCLUSION: Without fiber consumption influence, 30 days of *Lactobacillus Acidophilus* plus *Bifidobacterium Lactis* (10x109UFC/day) supplementation not modify neutrophil peroxide and cytokine (IL-1-β; TNF-α; IL-6; IL-8) production however cause specific modification in gut microbiota composition increasing relative abundance of *Lactobacillus ruminis*. .

SUPPORT CNPq 302706/2020 8437801/2018-7 CAPES

The Harvard T.H. Chan School of Public Health Microbiome Collection Core

Chengchen Li[1], Jeremy E. Wilkinson[1], Curtis Huttenhower[1,2,3]

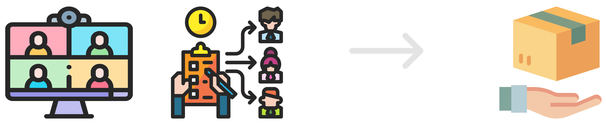
[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard

[3]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health

HCMPH MICROBIOME CENTER

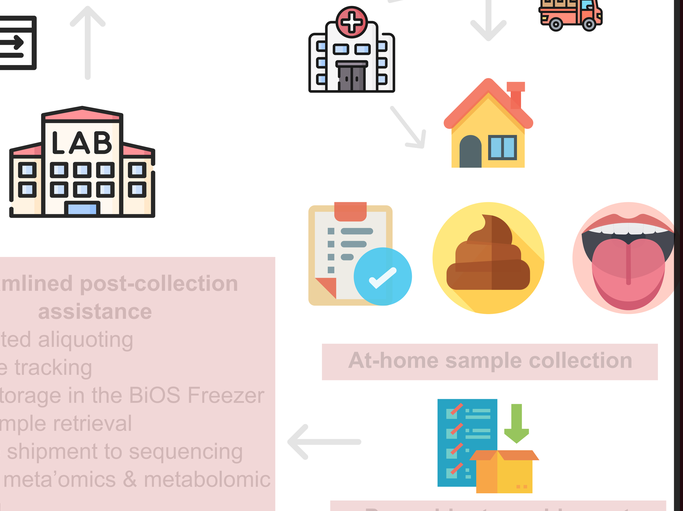
The Microbiome Collection Core at the Harvard T.H. Chan School of Public Health (HCMCC) was established in response to a strong demand among the research community for validated microbiome sample collection kit configurations and easy usability for in-home sampling. Under the umbrella of the Harvard Chan Microbiome in Public Health Center (HCMPH), HCMCC aims to support population-scale microbiome sample collection and expand our understanding of the microbiome to improve population health. The HCMCC has developed a multi carrier-compatible home stool and oral sample collection kit that permits cost-effective multi'omic microbiome studies, leveraging the intellectual and infrastructure foundation laid by the HMP2 (the 2nd phase of the NIH Human Microbiome Project) and the MLSC (Massachusetts Life Sciences Center)-funded MICRO-N (MICRObiome Among Nurses) collection. By providing this customizable microbiome collection kit, we enable researchers to perform multiple different molecular assays and tailor collection plan to study-specific needs.

HCMCC services



Microbiome sample collection plan development
- Collection kit configuration
- Kit distribution & logistics
- Sample transport plan
- Sample handling & storage plan

Kit ordering & shipment
- Kit customization & implementation
- Ambient temperature shipping
  - to selected clinical sites
  - direct to participants



Streamlined post-collection assistance
- Automated aliquoting
- Barcode tracking
- -80˚C storage in the BiOS Freezer
- Fast sample retrieval
- Sample shipment to sequencing labs for meta'omics & metabolomic profiling

At-home sample collection

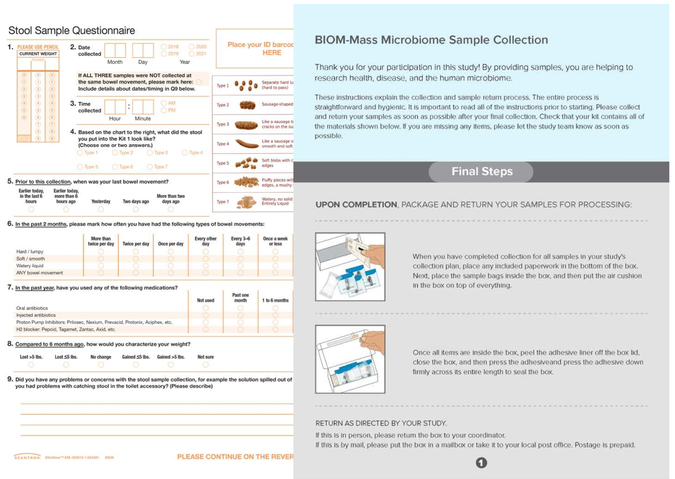
Pre-paid return shipment

THE HARVARD CHAN MICROBIOME COLLECTION CORE

The Microbiome Collection Core is a part of the Harvard Chan Microbiome in Public Health Center (HCMPH).
Want to learn more? Visit **https://hcmph.sph.harvard.edu**

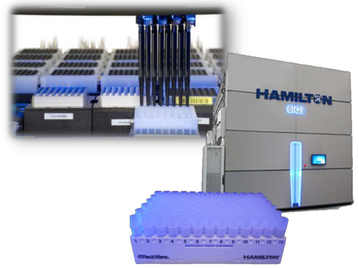
A scalable gut and oral microbiome sample collection platform

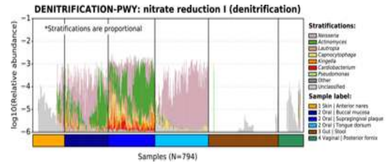
This customizable microbiome sample collection kit avoids the need for expensive, bulky, and inconvenient ice packs by providing several different room temperature storage media that are also compatible with multiple different molecular assays including **any combination of amplicon (16S), metagenomic, metatranscriptomic sequencing, metabolomics, and other molecular assays**. This kit further includes a collection method that uses anaerobic transport media that **yields live microbes for culture or gnotobiotic research**.



In addition to storage media, this sample collection kit includes **user-friendly instructions** and toilet accessories to maximumly facilitate and smooth the in-home stool sample collection experience. **Standardized questionnaires**, as companions to collected samples, are included to capture **recent medications, diet, anthropometric measurements, and gastrointestinal health status measured by the Bristol Stool Scale**. The modularity of this kit allows researchers to tailor kit components to study-specific needs and conduct cost-effective microbiome research ranging from **pilot studies to large-scale studies involving 10,000s of participants.**

HCMCC-supported study activities within the BIOM-Mass platform





Pre-collection
- Participant enrollment
- Kit ordering
- Kit distribution

Collection
- Self-collection
- Sample return through pre-paid shipment

Post-collection
- Sample aliquoting via Hamilton STAR automated liquid handler
- Long-term -80˚C storage via the BiOS Freezer Core
- Data generation
- Data analysis via the Microbiome Analysis Core

Microbiome population health research opportunities

- Accessible microbiome population studies' data on the BIOM-Mass Data Portal **https://biom-mass.org**
- Integrative microbiome informatics and analysis via the Harvard Chan Microbiome Analysis Core **https://hcmph.sph.harvard.edu/hcmac/**
- Long-term sample storage via the Harvard Chan BiOS Freezer Core
- Gnotobiotic mice experiments via the Harvard Chan Gnotobiotic Center for Mechanistic Microbiome Studies
- Course offerings on microbial communities and human microbiome research via the Harvard Chan Microbiome in Public Health Center

Special thanks to the the Massachusetts Life Sciences Center (MLSC), the Harvard Chan Microbiome Platform Steering Committee, the Harvard Chan BiOS Freezer Director Eric Rimm, the BWH/Harvard Cohorts Biorepository Laboratory Manager Christine Everett.

Project Manager: Chengchen (Cherry) Li
Microbiome Analysis Core Director: Jeremy E. Wilkinson
Scientific Director: Curtis Huttenhower

https://hcmph.sph.harvard.edu/hcmcc
https://huttenhower.sph.harvard.edu

Scalable learning of interpretable rules from microbiome dynamics

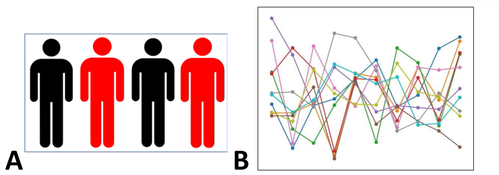
Venkata Suhas Maringanti[1,2,3], Vanni Bucci[3], Georg K. Gerber[1,4,5]

[1]Dept. Pathology, Brigham & Women's Hospital, Harvard Medical School; [2]University of Massachusetts Dartmouth; [3]University of Massachusetts Medical School; [4]Massachusetts Host-Microbiome Center; [5]Harvard-MIT Health Sciences & Technology

Abstract

The microbiome is inherently dynamic over time. Analyzing the relationship between microbiome dynamics and host outcomes is critical for defining predictive models and ultimately understanding the mechanisms through which the microbiome causes disease. Although off-the-shelf machine learning methods are widely deployed in the field, many methods do not consider temporal patterns in the data or the microbial phylogenetic structure, and moreover are not interpretable. We present a new, highly scalable model, MDITRE: Microbial Differentiable and Interpretable Temporal Rule Engine, based on our earlier fully-Bayesian method MITRE [1]. MDITRE is a supervised machine learning model that learns human-interpretable rules to predict host outcomes from microbial time-series data. Our model learns rules that capture relevant features in the phylogenetic and temporal space via domain-specific attention mechanisms. Our inference algorithm deployed on GPUs demonstrates that MDITRE achieves similar predictive performance as MITRE while running 30x – 70x faster on a suite of longitudinal microbiome datasets. Moreover, our model learns biologically meaningful relationships that our prior model did not. Our results indicate that MDITRE is highly scalable and accurate, and moreover can provide novel insights into complex and dynamic host-microbial ecosystems.

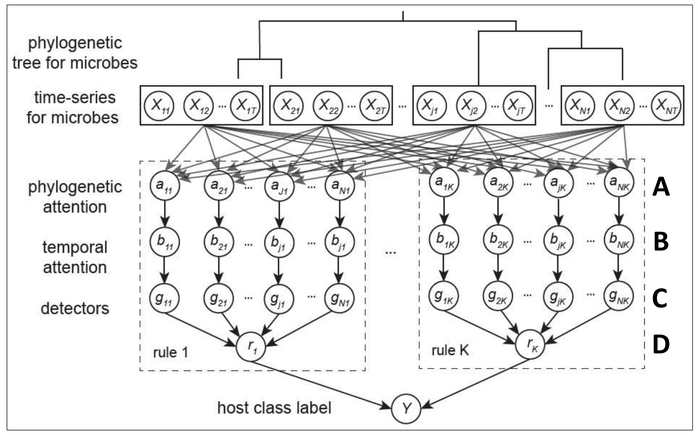
Model and Inference



Inputs

A: Binary host outcomes, e.g. disease or healthy
B: Relative abundances over time of microbes
C:Phylogenetic tree of microbes, encoding prior evolutionary relationships

Differentiable 5-layer Neural Network Model



Novel domain-specific attention mechanisms

A: Aggregate abundances of phylogenetically similar microbes

$$a_{ki}(\mathbf{X}_s) = \sum_j^N u_{kij}\mathbf{X}_{sj}$$

B: Average abundances from **A** over a relevant time window

$$b_{ki}(\mathbf{X}_s) = \sum_t^T v_{kit}a_{kit}(\mathbf{X}_s)$$

Rules as logical conjunctions of detectors

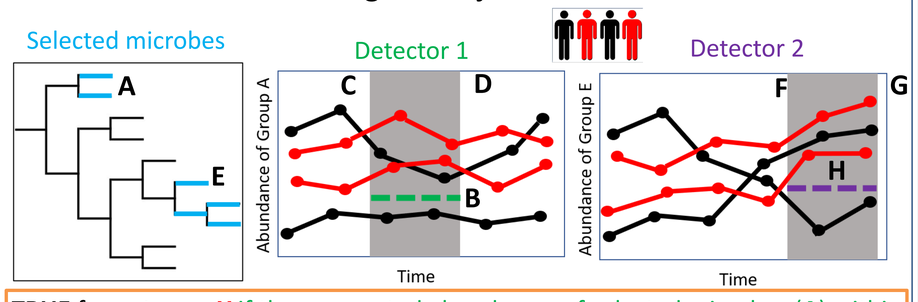
C: Compute detector response via sharp logistic function

$$g_{ki}(\mathbf{X}_s) = h(b_{ki}(\mathbf{X}_s) - c_{ki}, m)$$

D: Rule response as soft logical conjunction of detectors

$$\tilde{r}_k(\mathbf{X}_s) = \prod_i^N (1 - z_{ki}(1 - g_{ki}(\mathbf{X}_s)))$$

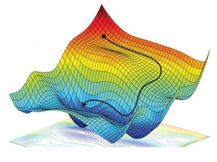
A learned rule as logical conjunction of 2 detectors



TRUE for outcome **Y** if the aggregated abundance of selected microbes (**A**) within selected time window **(C, D)** is above **B%** **AND** if the aggregated abundance of selected microbes (**E**) within selected time window **(F, G)** is above **H%** .

Inference

We perform Maximum-a-Posteriori (MAP) inference inference on our fully probabilistic model using gradient-descent optimization. We use prior distributions on number of rules and detectors to encourage sparsity and interpretability.



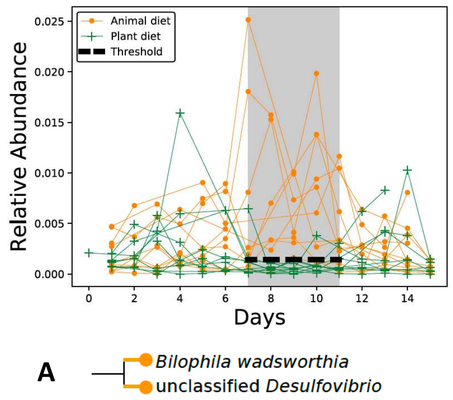
Experiments and Results

Prediction Accuracy and Scalability

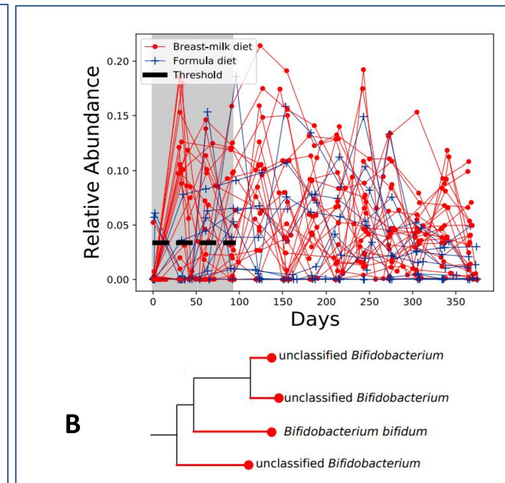
- We demonstrate the accuracy and scalability of our model on 3 longitudinal microbiome datasets David [2], Bokulich [3] and Vatanen [4] and compare to our prior method MITRE [1].
- We use F1 score as the metric to quantify the predictive performance using cross-validation

Model	F1 score (running time) for all datasets		
	David [2]	Bokulich [3]	Vatanen [4]
MITRE [1]	0.77 (23 hrs)	0.64 (30 hrs)	0.83 (64.5 hrs)
MDITRE (new method)	**0.81 (0.45 hrs)**	**0.58 (1 hr)**	**0.82 (0.83 hrs)**

Sample Learned Rules (not found by prior method MITRE [1])



- David et al. [2] studied the effect of exclusive Plant/Animal based dietary intervention (spanning 5 days) on the dynamics of 185 microbes with 20 human subjects over 16 days.
- Rule: **TRUE** for subjects on **Animal diet** if the aggregated abundance of **selected microbes A** is above **0.31%** between **day 7** and **day 11**.
- Rule is biologically relevant as the selected microbes are known to increase in presence of dietary lipids or chondroitin sulfate (component of animal cartilage)



- Bokulich et al. [3] studied the effect of Formula/Breast-milk dominant diet on the dynamics of 65 microbes in 35 infants observed for a year.
- Rule: **TRUE** for subjects on **Breast-milk dominant diet** if the aggregated abundance of **selected microbes B** is above **3%** between **day 0** and **day 93**.
- Rule is biologically relevant as selected microbes are known to express enzymes necessary to digest human milk oligosaccharides.

Conclusions

- We developed a **differentiable** model with **domain-specific attention mechanisms** that dynamically learns relevant features in phylogenetic and temporal space to predict host outcomes from microbial time-series data.
- Our model is **accurate and highly scalable**, and can provide novel insights into the complex and dynamic host-microbial ecosystems

Future Work

- Incorporate additional prior knowledge into the model such as knowledge about the taxonomy of microbes and biologically relevant time windows, to boost the model accuracy and robustness
- Add extra capability to the model to link rate of change of microbial abundances over time to host outcomes
- Explore alternate inference methods such as Variational Inference and Hamiltonian Monte Carlo

References

[1] Bogart et al., MITRE: inferring features from microbiota time-series data linked to host status, Genome Biology, 2019.
[2] David et al., Diet rapidly and reproducibly alters the human gut microbiome, Nature, 2014.
[3] Bokulich et al., Antibiotics, birth mode, and diet shape microbiome maturation during early life, Sci Transl Med, 2016.
[4] Vatanen et al., Variation in microbiome LPS immunogenicity comtributes to autoimmunity in humans, Cell, 2016.

Acknowledgements This work was supported by BWH President's Scholar Award, Harvard Catalyst and NIGMSR01GM130777.

Characterization of a novel microbiome marker anti-correlated with *S. aureus* carriage in the infant nasal microbiome

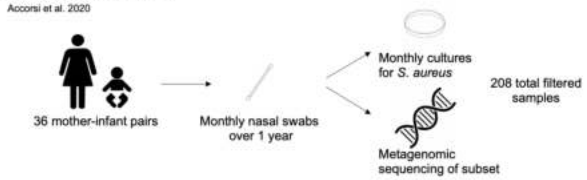
Madeleine C. Kline[1,4], Emma K. Accorsi[2,4], Eric A. Franzosa[2,3,4], Curtis Huttenhower[2,3,4]

[1]Harvard Medical School [2]Harvard T.H. Chan School of Public Health [3]Broad Institute of MIT and Harvard
[4]Harvard Chan Microbiome in Public Health Center

Staphylococcus aureus carriage in the nasal microbiome is an important determinant of subsequent *S. aureus* soft tissue infection. Identifying elements of the nasal microbiome that influence carriage provides insight on how to modulate these factors to prevent progression to infection. Prior work by Accorsi et al. 2020[1] discovered an uncharacterized, taxonomically unassigned ORF that was the major predictor of whether infant microbiome samples evaluated with shotgun metagenomic sequencing contained *S. aureus*. Subsequent investigation indicated that this ORF was actually a segment of 16S rRNA gene sequence incorrectly annotated by UniProt as a protein-coding gene. Our analyses provide substantial initial support for the hypothesis that this sequence represents a phylogenetic marker for a novel clade with genomic similarity to *Streptococcus* species and *D. pigrum*, which could antagonize *S. aureus* during colonization of the infant nasal microbiome.

Study Design and Methods

Data collection



Random Forest Model

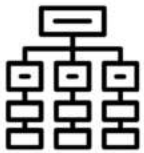
Identified sequence (UniRef90 of Interest or **UOI**) that predicted *S. aureus* carriage

bioBakery version 3 analysis

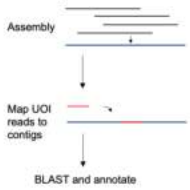
Updated functional and taxonomic profiling tools for higher resolution classification of UOI and correlated UniRefs

Covariation-based genome reconstruction

- MSPminer and MetaBAT try to characterize reads related to the UOI
- Analysis of *S.aureus* dataset and 267 nares samples in Expanded Human Microbiome Project (**HMP**)

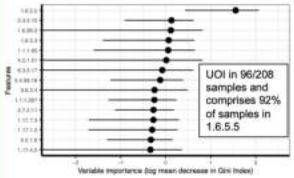


Analysis of UOI-adjacent sequences

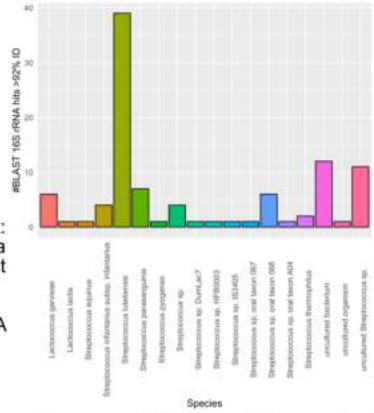


A sequence marker of a novel clade predicts *S. aureus* exclusion

Top 15 Random Forest Predictors



UOI BLAST Results

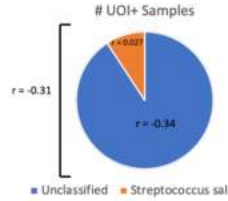


UOI (UniRef of Interest: UniRef90_X5NU12), a 42 aa sequence, is the strongest predictor of S. aureus carriage.

It is homologous to 16s rRNA from many species.

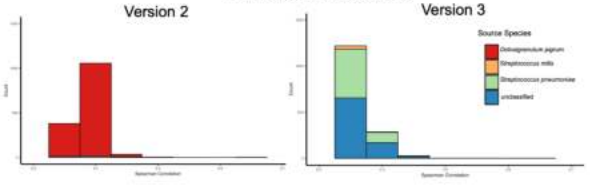
Updated taxonomic and functional profiling of *S. aureus* dataset

Spearman correlation coefficients: UOI to *S. aureus* relative abundance



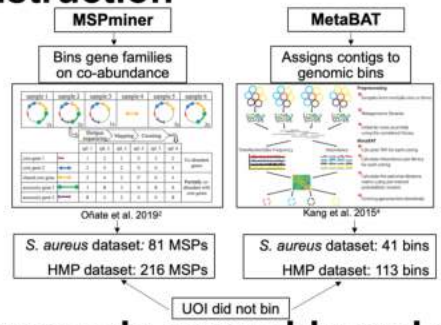
UOI partially classifies to *S. salivarius* in updated analysis, but only **unclassified UOI is signficantly negatively correlated with *S. aureus*.**

Correlations of UOI with species-stratified UniRef90 abundances



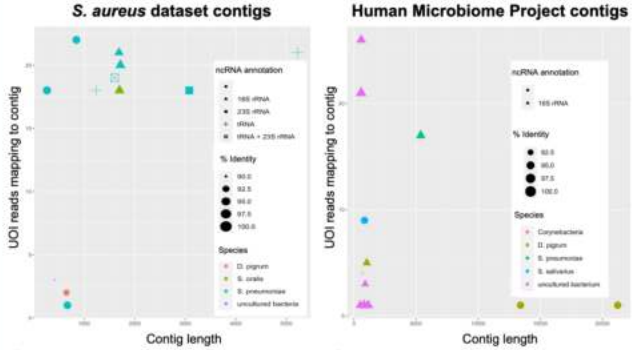
D. pigrum gene families from version 2 were still present in version 3, but r ~ 0.28, likely due to increased resolution of *Streptococcus* species in version 3.

Covariation-based genome reconstruction



Metagenomic assembly and contig analysis

Assembled contigs containing UOI reads analyzed for length, # of UOI reads, BLAST taxonomic results and % identity, and rRNA or tRNA annotation.



Conclusions and future work

Our work supports the hypothesis that the UOI is a fragment of 16S rRNA that represents a novel species. It is genomically similar to *D. pigrum* and *Streptococcus* species. Further in vitro experimentation, including amplification the sequence from related bacteria via RT-qPCR could help to characterize this novel clade.

Sources

1. Accorsi, E. K. et al. Determinants of Staphylococcus aureus carriage in the developing infant nasal microbiome. Genome Biol. 21, 301 (2020).
2. Plaza Oñate, F. et al. MSPminer: abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data. Bioinformatics 35, 1544–1552 (2019).
3. Kenny, D. J. et al. Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. Cell Host Microbe 28, 245-257.e6 (2020).
4. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165 (2015).
Images from the Noun Project by Yu Luck, Knut M. Synstead

Acknowledgments

We appreciate the valuable contributions of Yancong Zhang and Lauren McIver. Thank you also to the Harvard-MIT MD-PhD program.

http://huttenhower.sph.harvard.edu

Infant Gut Microbiome and Infections and Symptoms: A Prospective Cohort Study

Yuka Moroishi[1], Jiang Gui[1], Anne G. Hoen[1], Hilary G. Morrison[2], Emily R. Baker[1], Kari Nadeau[3], Hongzhe Li[4], Zhigang Li[5], Juliette Madan[1], Margaret R. Karagas[1]

[1]Dartmouth College, [2]Marine Biological Laboratory, [3]Stanford University, [4]University of Pennsylvania, [5]University of Florida

Background

The developing gut microbiome plays a critical role in immune maturation and infant health.

Method

Study Design	Prospective cohort data from the New Hampshire Birth Cohort Study (NHBCS)
Data Collection	• Stool samples collected at 6 weeks post partum • Telephone interviews conducted with infant's caregivers at 4, 8, 12 months postpartum
Statistical Analysis	• Generalized estimating equations with Poisson regression; log-alpha diversity and log-relative abundance of stool on repeated measures of counts of health outcomes

Takeaways:

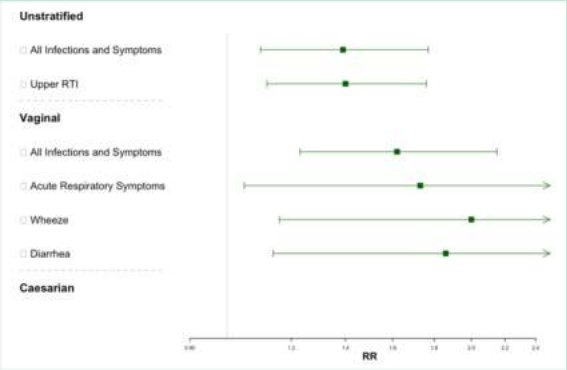
- High Alpha Diversity at 6 weeks Associated with Risk of All Infections and Symptoms, Upper RTI, Wheeze, and Diarrhea
- *Clostridium*, *Streptococcus*, and *Veillonella* Species in Gut Associated with Infant Outcomes

All analyses adjusted for maternal BMI, delivery type, sex, breast feeding at 6 weeks, perinatal antibiotic use, and gestational age. In tables, **black** represent positive association, and **red** indicates negative association

Table 1: Selected Metagenomics Species Associated with Health Outcomes (N = 185)

Outcomes	Total Sample N = 185	Outcomes	Caesarian N = 56
All Infections and Symptoms	*Veillonella unclassified*	All Infections and Symptoms	*Haemophilus influenzae*
Diarrhea	*Streptococcus peroris* *Streptococcus salivarius*	Upper RTI	*Veillonella parvula* *Corynebacterium pseudodiphtheriticum* *Streptococcus peroris* *Clostridium butyricum* *Coprobacillus unclassified*

Figure 1: Risk Ratios of Selected 16S Alpha Diversities Associated with Counts of Health Outcome (N = 464)



cILR: Taxonomic Enrichment Analysis with Isometric Log Ratios

Quang Nguyen[1 2] Anne G. Hoen[1 2] H. Robert Frost[2]

[1] *Department of Epidemiology, Dartmouth College*
[2] *Department of Biomedical Data Science, Dartmouth College*

Introduction

- Standard microbiome analyses often aggregate variables to sets, commonly Linnean taxonomic categories (e.g. Phylum) identified through sequence classification. Aggregation can help with standard challenges with microbiome relative abundance data, such as high-dimensionality and sparsity.
- However, most researchers perform aggregation through the pairwise summation of counts, preventing comparison across sets of different sizes. Count-based aggregation methods also do not preserve inter-sample distances, due to fact that microbiome data is uniquely compositional.
- Here we developed a method to aggregate variables through computing a competitive enrichment score, comparing those inside the set and those outside the set.

Acknowledgements

This research is supported by funding from the National Institutes of Health (grants NLM R01LM012723, NIGMS P20GM104416, NLM K01LM012426 NIH UG3 OD023275, NIEHS P01ES022832 and EPA RD-83544201).

Results

Sample-level significance testing

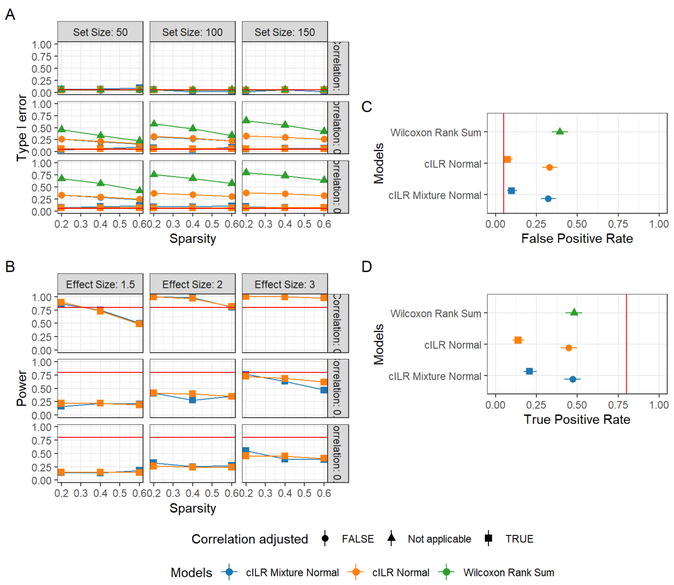
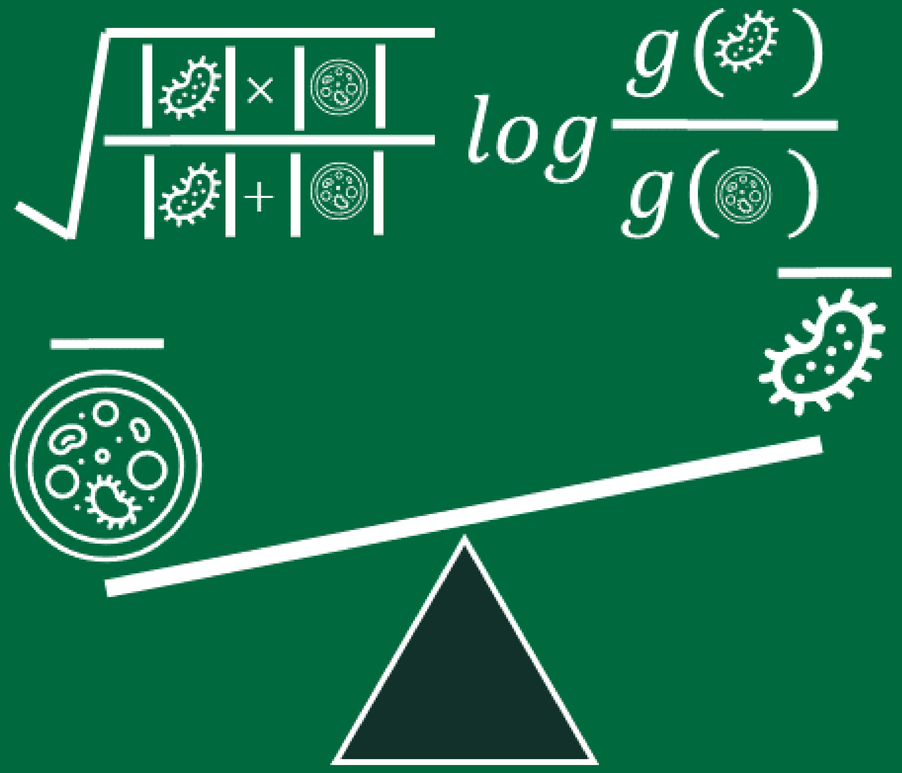


*Figure 1: Type I error and power of sample-level enrichment testing using cILR compared against a naive Wilcoxon Rank Sum test. Panel **(A)** presents type I error evaluated in parametric simulations under different set sizes, inter-taxa correlation and sparsity. Panel **(B)** presents power evaluated in parametric simulations under different effect sizes, inter-taxa correlation and sparsity. For panels **(C)** and **(D)**, we utilized the the 16S rRNA sequencing dataset of supragingival and subgingival sites from the Human Microbiome Project where supragingival sites are known to have enriched aerobic microbes. Here, we test for the enrichment of aerobic microbes across all samples, and considered a true positive is when a sample is significantly enriched for aerobic microbes and labelled as supragingival.*

Classification capacity

Our method leverages the **isometric log ratio transformation** to **generate enrichment scores for taxa sets** that can be used for standard microbiome analyses while also allowing for **sample-level significance testing** under a **competitive null hypothesis**



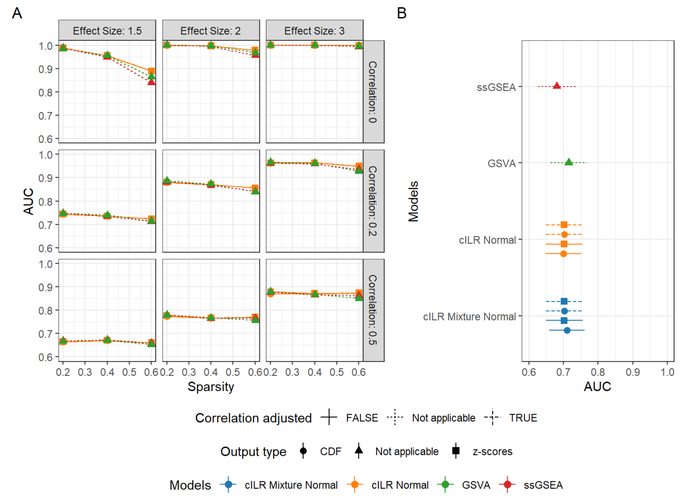


*Figure 2: Classification power of cILR scores compared against existing methods in the gene set testing literature that generates single sample enrichment scores. Area under the ROC curve (AUC) measures whether scores highly rank samples where the set of interest is known to be enriched. Panel **(A)** presents results under different parametric simulation conditions while panel **(B)** presents similar analyses on the 16S rRNA sequencing dataset of supragingival and subgingival sites from the Human Microbiome Project. In this data set, supragingival sites are known to have enriched aerobic microbes.*

Utilizing enrichment scores for disease prediction

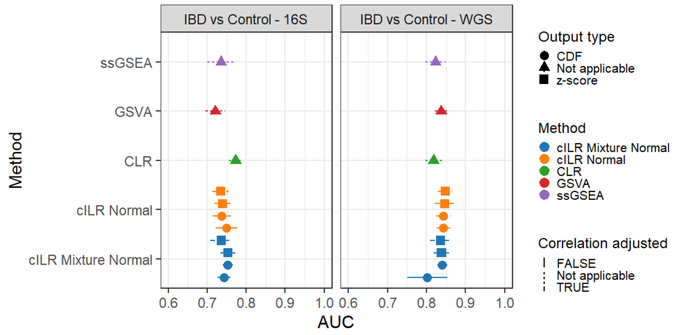
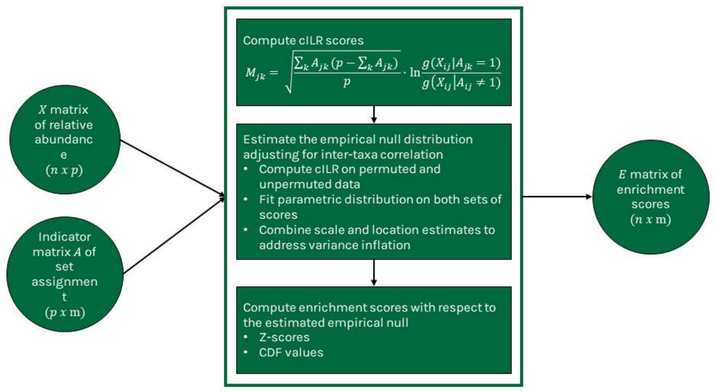


Figure 3: Classification performance of a standard random forest model using cILR scores compared against existing methods in gene set testing literature and the standard centered-log ratio transformation approach. The learning task involves predicting patients with inflammatory bowel disease (including Crohn's disease and ulcerative colitis) versus controls. Data sets used span both 16S rRNA sequencing (Gevers et al. 2014) and whole genome shotgun sequencing (Nielsen et al. 2014)

Methods



References

Gevers, Dirk, Subra Kugathasan, Lee A. Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, et al. 2014. "The Treatment-Naive Microbiome in New-Onset Crohn's Disease." *Cell Host & Microbe* 15 (3): 382–92. https://doi.org/10.1016/j.chom.2014.02.005.
Nielsen, H. Bjørn, Mathieu Almeida, Agnieszka Sierakowska Juncker, Simon Rasmussen, Junhua Li, Shinichi Sunagawa, Damian R. Plichta, et al. 2014. "Identification and Assembly of Genomes and Genetic Elements in Complex Metagenomic Samples Without Using Reference Genomes." *Nature Biotechnology* 32 (8): 822–28. https://doi.org/10.1038/nbt.2939.

Intestinal inflammation leads to changes in the blood PBMC and plasma microbiome

Emese Prandovszky O'Donnell[1 ‡], Hua Liu[1 ‡], Emily G. Severance[1], Faith Dickerson[2], Robert H. Yolken[1]

[1] Stanley division Developmental Neurovirology, Johns Hopkins University School of Medicine, Baltimore, MD, USA;
[2] Sheppard Pratt, Baltimore, MD, 21204
‡Authors equally contributed to the work.

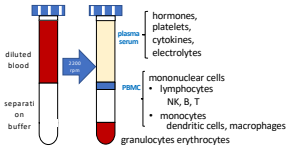
Introduction

Despite several studies having confirmed the presence of bacteria in the blood of humans *(1)*, very little is known about the distribution of microbial DNA among leukocytes in the systemic circulation *(2)*. The presence of a blood microbiome more often has been associated with chronic inflammatory diseases *(3)* due to impaired barrier function. Existing drawbacks of low biomass contamination makes the interpretation of the assay fairly challenging *(4)*.

Objectives

In this preliminary study, we aimed 1) to evaluate the peripheral blood mononuclear cells (PBMC) and plasma bacterial microbiome in humans, and 2) to investigate potential bacterial translocation into systemic circulation. To address these questions, two groups of patients with schizophrenia were selected for the microbiome analyses of both plasma and PBMC samples: one group with acute intestinal inflammation and the other without it. Each individual in the acute intestinal inflammation group had high ELISA values (≥ 1) for ASCA IgG in the blood *(5)*.

Materials and Methods



Samples:
Blood from individuals with intestinal inflammation (n=7) and without it (n=6) were separated into plasma and PMBC using gradient centrifugation.

DNA extraction: Along with microbial community standards, and negative blank extraction samples, the DNA of the PBMC and plasma samples was extracted using **Ultra-Deep Microbiome Prep Kit** (Molzym, G-020-050).

Libraries (including negative controls) were prepared using **Illumina NexteraXT** kit targeting **16S v3v4** regions and were sequenced on **MiSeq v3** Illumina platform.

Sequence analyses:

SEQs → **QIIME2** *(6)* built-in DADA2 de-noising + additional ENV contaminant removal → **BOWTIE 2** *(7)* EUK contaminant removal → **TAXONOMY ASSIGNMENT** GreenGenes v13.8 / **DIVERSITY ANALYSES**

Visualization: was completed by employing MS-Excel and q2R package *(8)* in R.

PCR confirmation: Based on the sequencing results, specific primers were designed for Enterococcus faecalis. The remaining genomic DNA of the same blood samples was tested in a SYBR Green PCR reaction.

Results

Figure 1A. – Diversity
Principal coordinate analyses of Jaccard dissimilarity metrics along with pairwise Permanova analyses showed significant differences between patients with and without intestinal inflammation in both plasma and PBMC.

Group 1	Group 2	Sample size	Permutations	pseudo-F	q-value
PBMC_wo.Infl	PBMC_Infl	10	999	3.0579	0.041
Plasma_wo.Infl	Plasma_Infl	11	999	4.4674	0.012
PBMC_wo.Infl	Plasma_wo.Infl	8	999	1.4696	0.205
PBMC_Infl	Plasma_Infl	13	999	0.7461	0.695

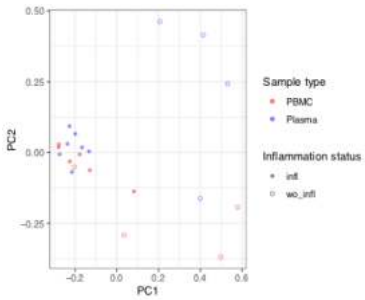


Figure 1B. - Diversity
Shannon metrics depicting species richness in the four experimental groups. Kruskall-Wallis pairwise comparison showed that inflammation leads to species enrichment in both plasma (q=0.049) and PBMC (q=0.031).

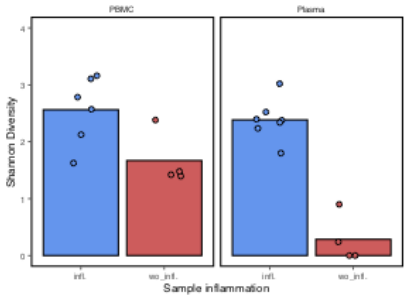


Figure 2. (heatmap below)
Relative abundance at species level
Enterococcus faecalis and Enterobacteriaceae/*Escherichia-Shigella* genus were highly abundant in the group with intestinal inflammation in both plasma and PBMC.

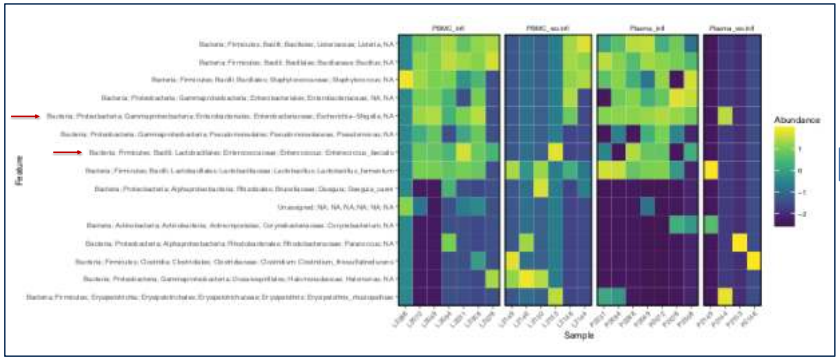
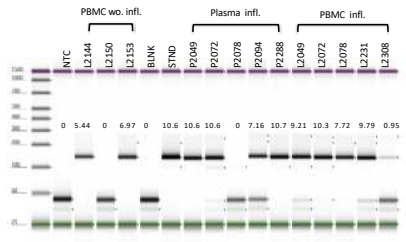


Figure 3.
***E. faecalis* PCR**



The blank (BLNK) and non-template control (NTC) did not produce any amplification, while the microbiome community standard (STND) had a positive amplification at 138 bp.

The higher relative abundance from the amplicon sequencing correlated with higher concentration of specific *Enterococcus* faecalis PCR product in the PCR reaction.

Conclusion

- Patients with intestinal inflammation are more likely to have detectable bacterial microflora than those without it in both plasma and PBMC.
- The bacterial microflora in each plasma and PBMC pair was different. This finding suggests that the results cannot be ascribed to skin flora or the contamination of reagents.
- In the intestinal inflammation group, some highly abundant species were more likely to originate from the intestine in both plasma and PBMC.
- Due to the preliminary feature of the study, the main limitations are the small sample size and lack of correction for potential cofounders.
- The measurement of the microbiome from plasma and PBMC may provide a new method for the characterization of intestinal inflammation.

References

1. Païssé S, *et al.* Comprehensive description of blood microbiome from healthy donors assessed by 16S targeted metagenomic sequencing. Transfusion. 2016;56(5):1138–4.
2. Peña-Cearra, A *et al.* Peripheral blood mononuclear cells (PBMC) microbiome is not affected by colon microbiota in healthy goats. *Anim microbiome* **3,** 28 (2021)
3. Potgieter M. et al. The dormant blood microbiome in chronic, inflammatory diseases. FEMS Microbiol Rev. 2015;39(4):567–91.
4. Glassing A *et al.* Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016;8(1):24..
5. Vermeire, Severine, *et al.* "Anti-Saccharomyces cerevisiae antibodies (ASCA), phenotypes of IBD, and intestinal permeability: a study in IBD families." *Inflammatory bowel diseases* 7.1 (2001): 8-15.
6. Bolyen, Evan, *et al.* "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2." *Nature biotechnology* 37.8 (2019): 852-857.
7. Langmead, Ben, and Steven L. Salzberg. "Fast gapped-read alignment with Bowtie 2." *Nature methods* 9.4 (2012): 357.
8. https://github.com/jbisanz/qiime2R

Maternal probiotic intake in obese mice reduces anxiety-like behaviour in offspring and increases blood and brain lactate

[1,2,3]**Radford-Smith DE.**, [2]Probert F., [3]Burnet PWJ. & [1]Anthony DC.

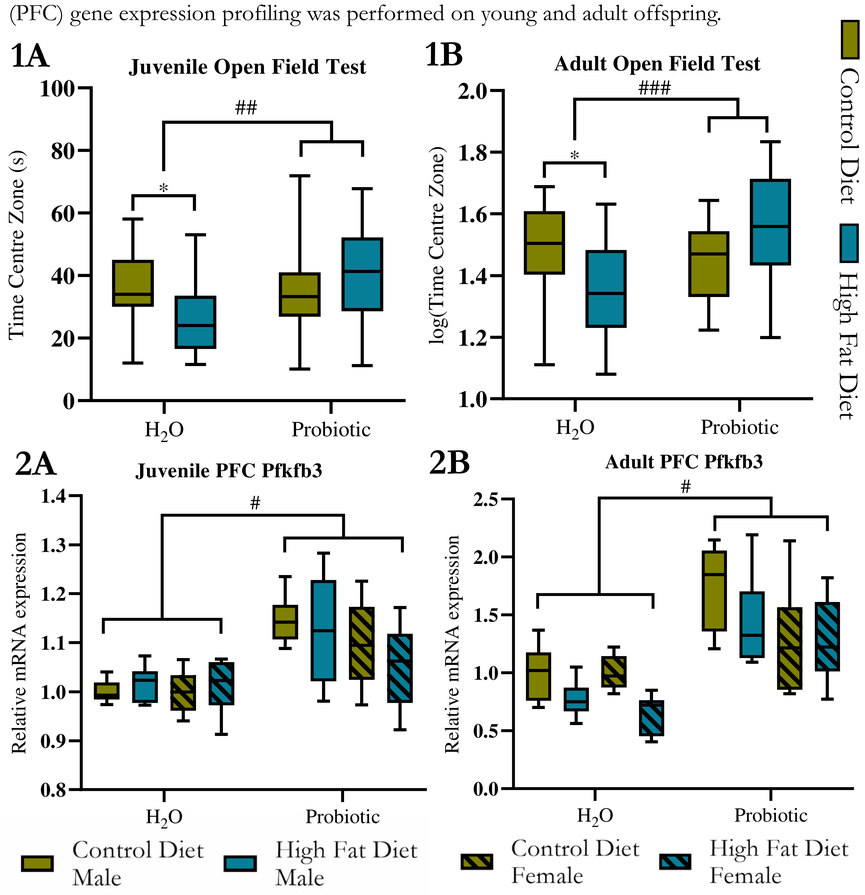
[1]Pharmacology, University of Oxford, Oxford, UK; [2]Chemistry, University of Oxford; [3]Psychiatry, University of Oxford

Aim: To investigate how maternal diet-induced obesity affects offspring brain development and behaviour, and whether the changes could be mitigated by perinatal probiotic exposure.

Background:

- Maternal obesity and depression are growing public health epidemics in Europe and the United States (1-3).
- Maternal gut dysbiosis induced by diet affects the offspring microbiome, brain, and behaviour (4).
- No studies have investigated whether maternal probiotic intake may counter the adverse neurometabolic and behavioural effects of maternal obesity on offspring.

Study Design and Methods: CD-1 female mice were randomly assigned to receive either high-fat diet (n=8) or control diet (n=8) prior to and throughout gestation and nursing. Half of each group received a multi-strain probiotic during gestation and nursing, and the other half received the vehicle. Offspring behaviour was tested at weaning age (n=27-29 per group) and at sixteen weeks old (n=19-21 per group). Untargeted brain, liver, faecal, and plasma metabolomics were performed on dams and young offspring. Prefrontal cortex (PFC) gene expression profiling was performed on young and adult offspring.



Results:

- Maternal obesity increased anxiety-like behaviour in juvenile and adult offspring, independent of sex (Figure 1). Perinatal probiotic supplementation protected against this increase in anxiety-like behaviour, and had a main effect on reducing anxiety-like behaviour.
- Maternal probiotic intake increased brain *Pfkfb3* expression (Figure 2), a marker of brain astrocyte metabolism, in offspring at postnatal day 21 (weaning) and at 16 weeks (adulthood).
- *Pfkfb3* expression correlated with brain lactate levels, which were increased in probiotic offspring (Figure 3 and Figure 4).

Figure 1 (left): Maternal obesity increases anxiety-like behaviour in the open field test (OFT) in A) juvenile and B) adult offspring in the absence of perinatal probiotic supplementation. Maternal probiotic intake reduced anxiety-like behaviour. There was a significant interaction between maternal diet and probiotic intake on increasing time in centre in juvenile ($F_{(1, 106)} = 10.04$, $p = 0.0020$) and adult ($F_{(1, 75)} = 14.20$, $p = 0.0003$) offspring.

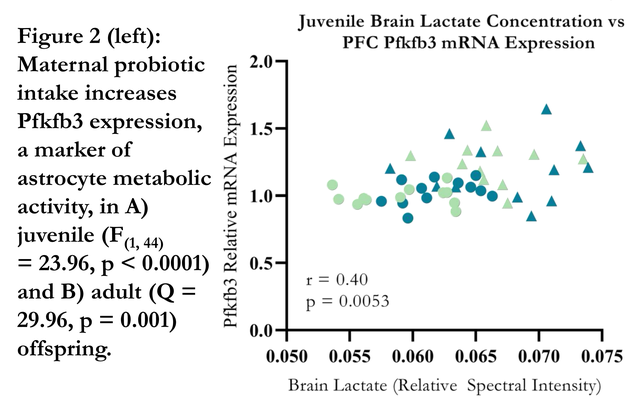


Figure 2 (left): Maternal probiotic intake increases Pfkfb3 expression, a marker of astrocyte metabolic activity, in A) juvenile ($F_{(1, 44)} = 23.96$, $p < 0.0001$) and B) adult ($Q = 29.96$, $p = 0.001$) offspring.

Figure 3 (left): A positive correlation exists between cortical lactate concentration and Pfkfb3 mRNA expression in the prefrontal vortex ($r = 0.40$, $p = 0.0053$) in juvenile offspring.

Funding: DE Radford-Smith is supported by the Newton Abraham Studentship (Oxford) and the Clarendon fund in association with the Lincoln College Kingsgate award (Oxford).

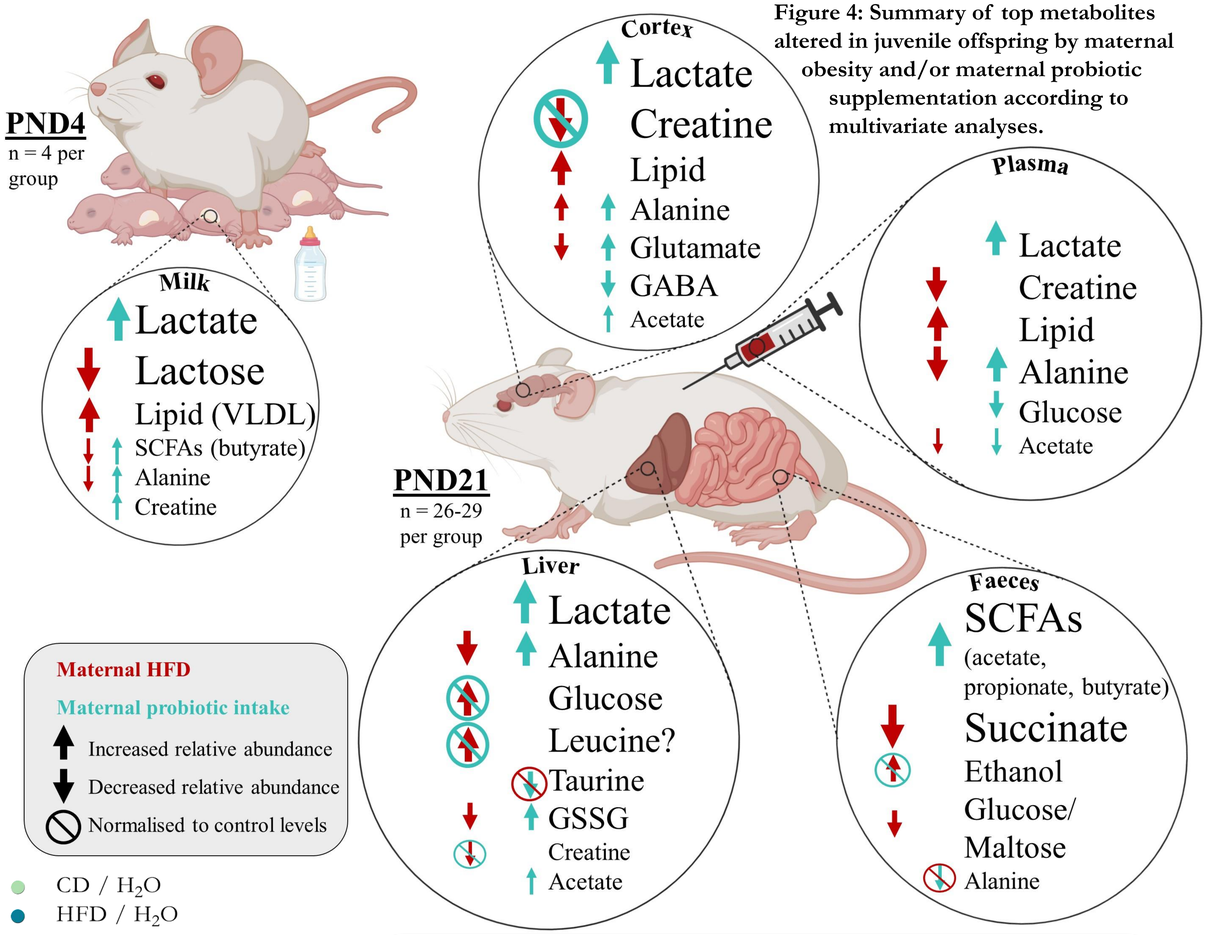


Figure 4: Summary of top metabolites altered in juvenile offspring by maternal obesity and/or maternal probiotic supplementation according to multivariate analyses.

Conclusions:

- We have shown for the first time that maternal probiotic supplementation during gestation and nursing is protective against increased anxiety-like behaviour, which occurred in the male and female offspring of obese dams.
- We observed pervasive metabolic effects of maternal obesity across the gut-liver-brain axis.
- Maternal perinatal probiotic intake increased blood and brain lactate in offspring at weaning age, which may have contributed to the increased resilience to maternal obesity exhibited in these offspring.

References:

1. Malhi GS, Mann JJ. Depression. Lancet. 2018;392(10161):2299-312.
2. Devlieger R, Benhalima K, Damm P, Van Assche A, Mathieu C, Mahmood T, et al. Maternal obesity in Europe: where do we stand and how to move forward?: A scientific paper commissioned by the European Board and College of Obstetrics and Gynaecology (EBCOG). Eur J Obstet Gynecol Reprod Biol. 2016;201:203-8.
3. Yogev Y, Catalano PM. Pregnancy and obesity. Obstet Gynecol Clin North Am. 2009;36(2):285-300, viii.
4. Buffington SA, Di Prisco GV, Auchtung TA, Ajami NJ, Petrosino JF, Costa-Mattioli M. Microbial Reconstitution Reverses Maternal Diet-Induced Social and Synaptic Deficits in Offspring. Cell. 2016;165(7):1762-75.

Power and sample size calculations for microbiome epidemiology

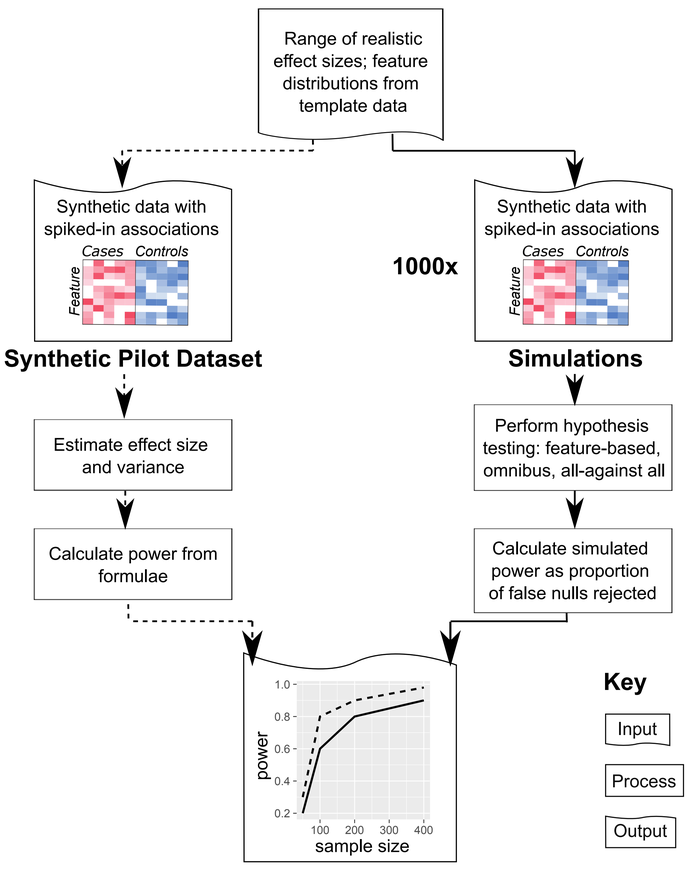
Meghan I. Short[1,2,3], Emma Schwager[1,2,3], Siyuan Ma[1,2,3], Lauren McIver[1,2,3], Jeremy E. Wilkinson[1,2], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3]

[1]Harvard T.H. Chan School of Public Health, [2]Harvard Chan Microbiome in Public Health Center, [3]Broad Institute of MIT and Harvard

Accurately assessing statistical power as a function of sample size and effect size is critical for good study design, particularly with respect to complex human populations and high-dimensional molecular epidemiology. Microbiome data especially pose unique challenges, considering the many biological factors that can influence the microbiome, the multiple types of molecular measurements possible, and their technical and biological variability including compositionality, zero-inflation, and measurement error. Standard methods for calculating power may thus be inadequate for measuring associations between microbial features and biological variables of interest. We demonstrate this using simulated and synthetically spiked microbial profiles containing known relationships of varying types.

Benchmarking and expanding power calculation methods for microbial communities

To compare analytical with actual power, we generated realistic feature tables with SparseDOSSA2, and spiked in associations between features and a binary metadata variable.



Microbiome power can be over- and underestimated by standard methods

Simulated power was defined as the proportion of taxa with spiked-in associations that were differentially abundant (p<0.05) between cases and controls. Means ± SD across 1000 simulations plotted below.

Analytical power for Wilcoxon tests and linear models was obtained using standard formulae, with effect size from syntheic pilot dataset. Values plotted below.

Wilcoxon

Linear model

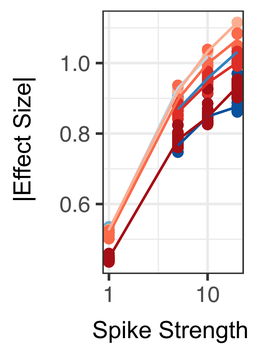
Nonparametric methods overestimate power and can fail to control FPR

Linear methods under- or over-esimated power, depending on the effect size, and controlled FPR

Developing flexible power formulae for diverse microbiome models

Per-feature tests	**Omnibus tests**	**All-against-all tests**
Identifies microbial features associated with various metadata	Associations between broader community composition and metadata	Superimposition of microbiome with other high-dimensional data types

Future work: Calibrate traditional effect size measures to microbiome exposures

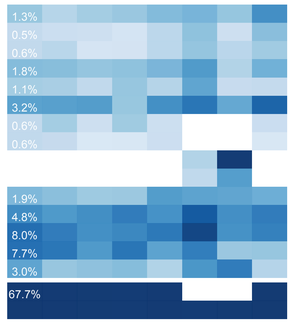


Expected effect from:

Antibiotics?

IBD?

Dietary variation?



Software interface will enable simple study planning

Inputs:

Acknowledgments

This work has been supported by sponsored research from Hill's Pet Nutrition and NIH NIDDK R24DK110499

http://huttenhower.sph.harvard.edu

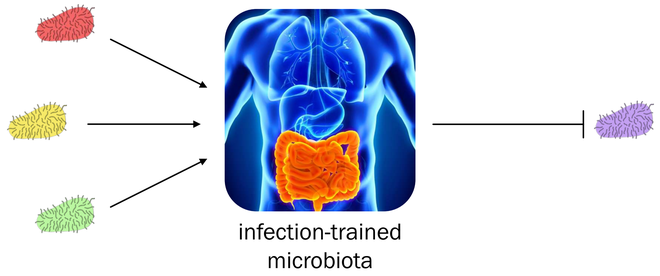
Infection trains the host for microbiota-enhanced resistance to pathogens

Apollo Stacy[1,2,3], Vinicius Andrade-Oliveira[1,4], John A. McCulloch[5], Benedikt Hild[6], Ji Hoon Oh[6], P. Juliana Perez-Chaparro[2], Choon K. Sim[7,8], Ai Ing Lim[1], Verena M. Link[1], Michel Enamorado[1], Giorgio Trinchieri[5], Julia A. Segre[7], Barbara Rehermann[6], Yasmine Belkaid[1,2]

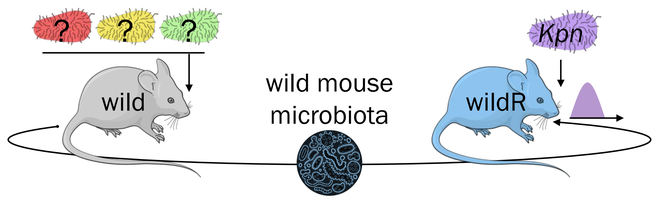
[1]Metaorganism Immunity Section, Laboratory of Host Immunity and Microbiome, NIAID; [2]NIAID Microbiome Program, NIAID; [3]Postdoctoral Research Associate Training Program, NIGMS; [4]Current Affiliation: Center for Natural and Human Sciences, Federal University of ABC, Santo André-SP, Brazil; [5]Laboratory of Integrative Cancer Immunology, Center for Cancer Research, NCI; [6]Immunology Section, Liver Diseases Branch, NIDDK; [7]Microbial Genomics Section, Translational and Functional Genomics Branch, NHGRI; [8]Current Affiliation: Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore

Introduction

- Gut microbiota protects against colonization by pathogens
- Often study microbiota resistance in "clean" laboratory mice
- Can infection-induced microbiota shield against subsequent infection?

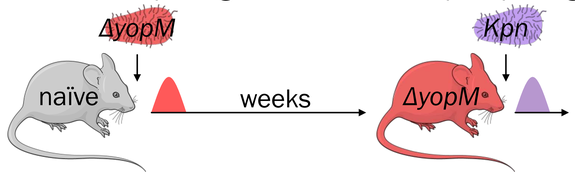


Infection-induced host models



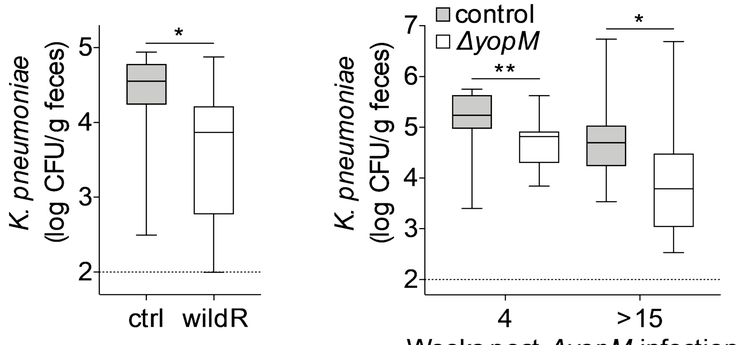
Y. pseudotuberculosis
- food-borne pathogen

Klebsiella pneumoniae
- hospital pathogen



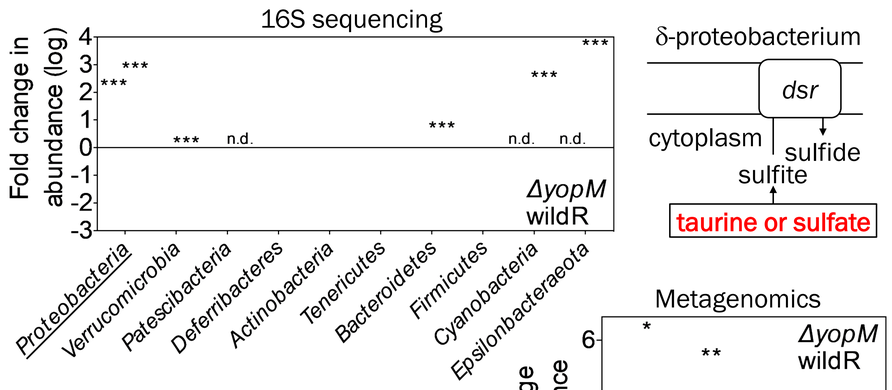
Does prior infection alter resistance to subsequent infection?

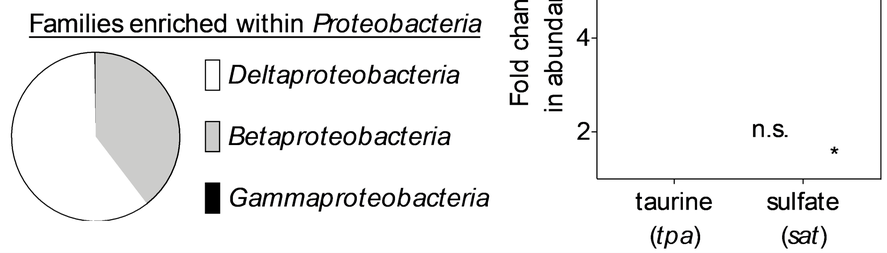
- Prior gut infection enhances resistance to subsequent *K. pneumoniae* gut infection



How does prior infection impact the gut microbiota?

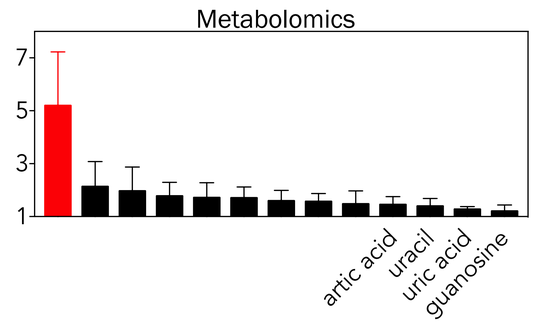
- Prior infection elicits taurine-utilizing *Deltaproteobacteria*

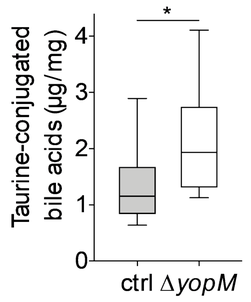




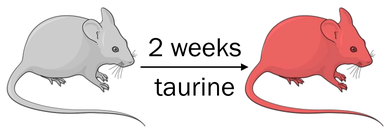
Does prior infection increase taurine levels?

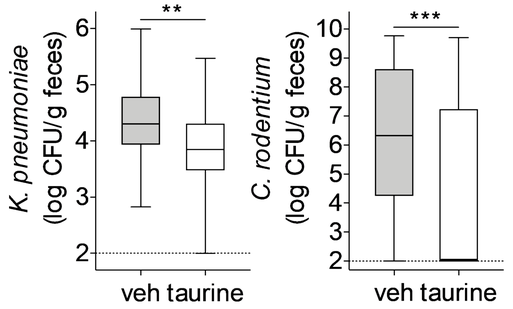
- Infection-induced *Deltaproteobacteria* likely expand in response to bile acid-derived taurine





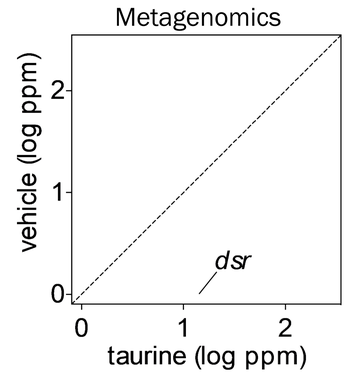
Does taurine alone enhance resistance?

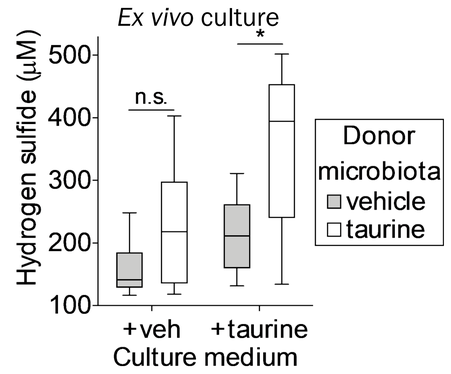




How does taurine impact the gut microbiota?

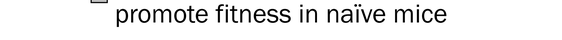
- Taurine amplifies the microbiota's production of sulfide





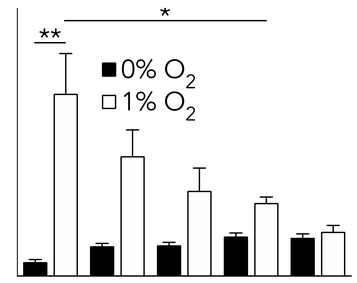
How does taurine enhance resistance?

- Taurine-derived sulfide restricts *K. pneumoniae* from utilizing 1,2-propanediol, a strictly respiratory substrate

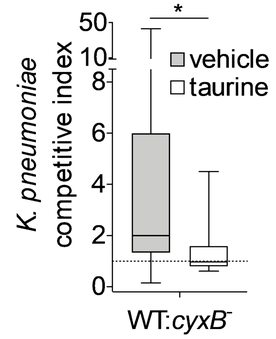


K. pneumoniae genes that promote fitness in naïve mice

Growth on 1,2-propanediol

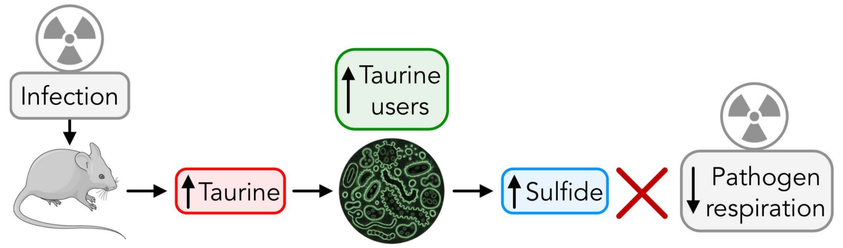


Competition assay



Conclusion

- The gut microbiota can be trained by the host after infection
- Training metabolite can be co-opted into a therapeutic



Probiotics alter the antibiotic resistance gene reservoir along the human GI tract

Emmanuel Montassier[1,2,3,*], Rafael Valdés-Mas[4,*], Eric Batard[1,2], Niv Zmora[4,5,6], Mally Dori-Bachash[4], **Jotham Suez[4,7,9]**, Eran Elinav[4,8,9]

,Microbiota Hosts Antibiotics and bacterial Resistances (MiHAR), Université de Nantes, Nantes, France (2) Department of Emergency Medicine, CHU de Nantes, Nantes, France (3) F-CRIN INI-CRCT, Nancy, France (4) Immunology Department, Weizmann Institute of Science, Rehovot, Israel (5) Sackler Faculty of Medicine, Tel Aviv University (1) Tel Aviv, Israel (6) Internal Medicine Department, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel (7) Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA (8) Division of Cancer-Microbiome Research, DKFZ, Heidelberg, Germany (9) Co-last authors (*) Equal contribution

Can probiotics prevent the spread of antibiotic resistance genes?

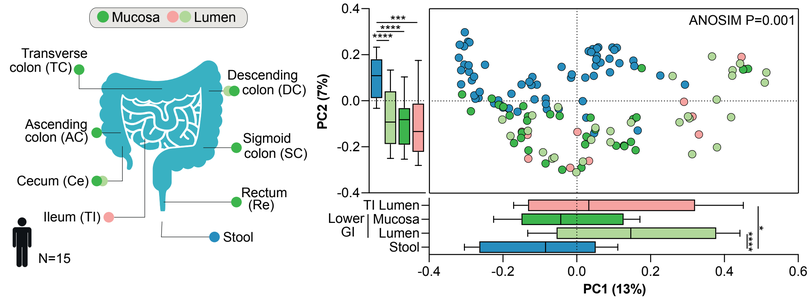
Antimicrobial resistance poses a major global health threat

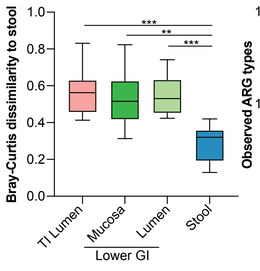
The human gut microbiome serves as a reservoir for antimicrobial resistance genes (ARG, “the **resistome**”)

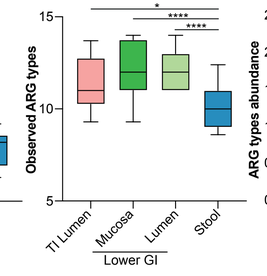
Supplements containing live **probiotic** microorganisms have been suggested to ameliorate resistome expansion – **but evidence are limited and conflicting**

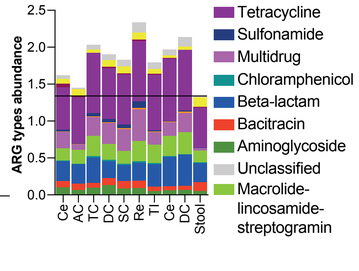
Are we looking in the wrong place?

Stool samples do not accurately reflect the gut resistome

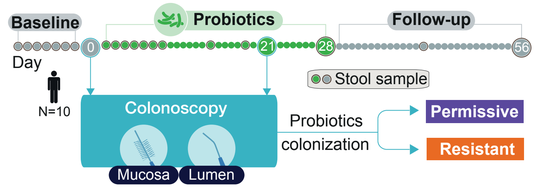


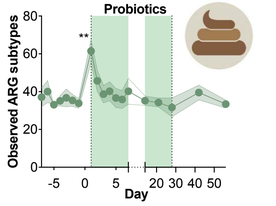


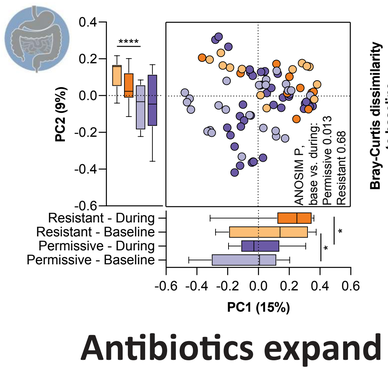


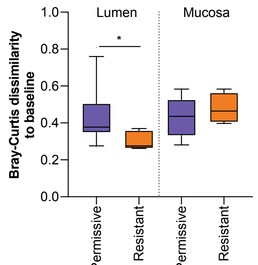


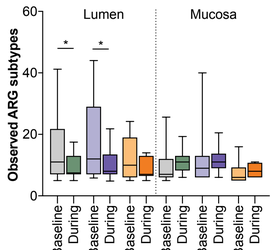
Probiotics colonization is associated with reduced ARG load in endoscopic samples





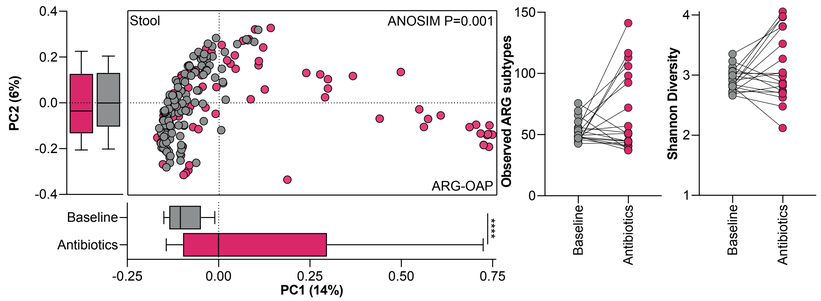


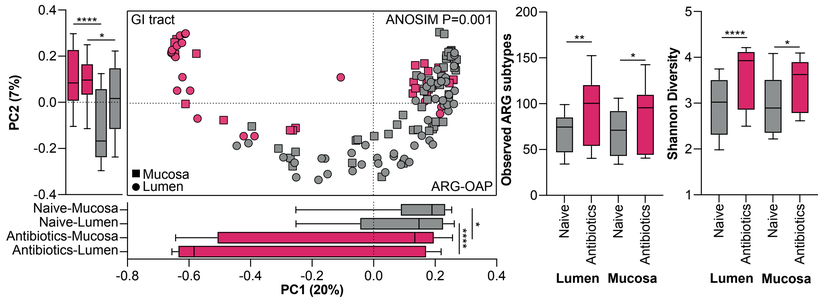




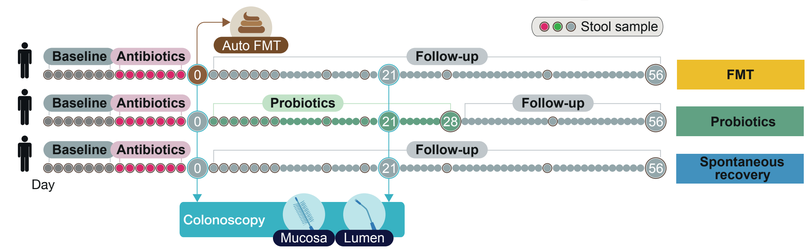
PC1 (15%)

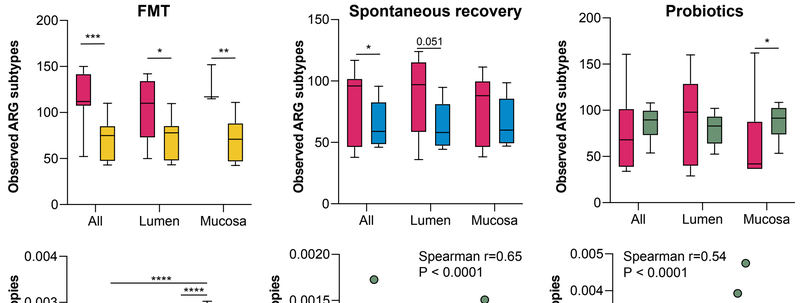
Antibiotics expand the resistome in the lower GI

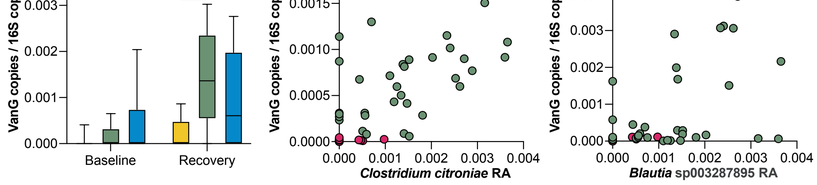


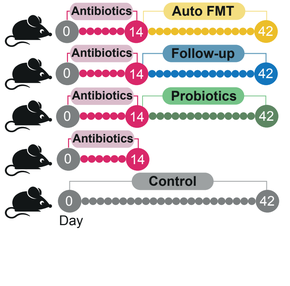


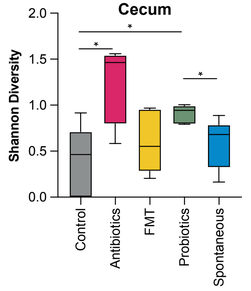
Probiotics are associated with increased ARG content in the lower GI following antibiotics

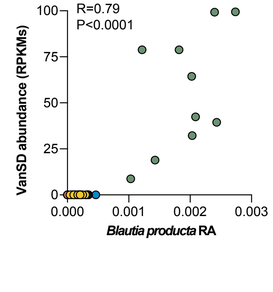












A beneficial or deterimental effect of probiotics on the gut resistome is antibiotics-dependent & person-specific

Probiotic Supplementation Increase Monocyte's Function and Maintair Gut Microbiota Alpha Diversity of Marathon Runners

Edgar Tavares-Silva1, Geovana S F Leite2, Helena A P Batatinha3, Ayane S Resende2, Antônio H Lancha Junior2, José C R Neto3, Ronaldo V Thomatieli-Santos1.

1- Department of Psychobiology, Federal University of São Paulo, Brazil.
2- Department of Biodynamics of the Movement of the Human Body, School of Physical Education and Sports, University of São Paulo, Brazil.
3- Department of Cell and Tissue Biology, Institute of Biomedical Science, University of São Paulo, Brazil.
Corresponding author: tavares.silva@unifesp.br

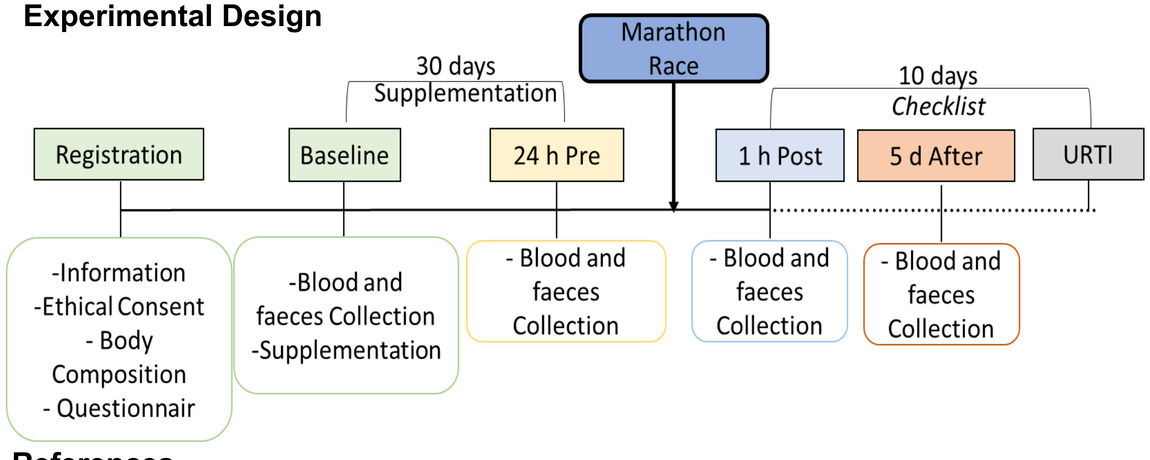
Background

After prolonged physical exercise, the immunological response of athletes can be impaired. Recently, the relationship between the intestinal microbiota and the immunological response has been postulated. We investigated whether probiotic supplementation modulates athlete's intestinal microbiota and the immune response before and after a marathon race.

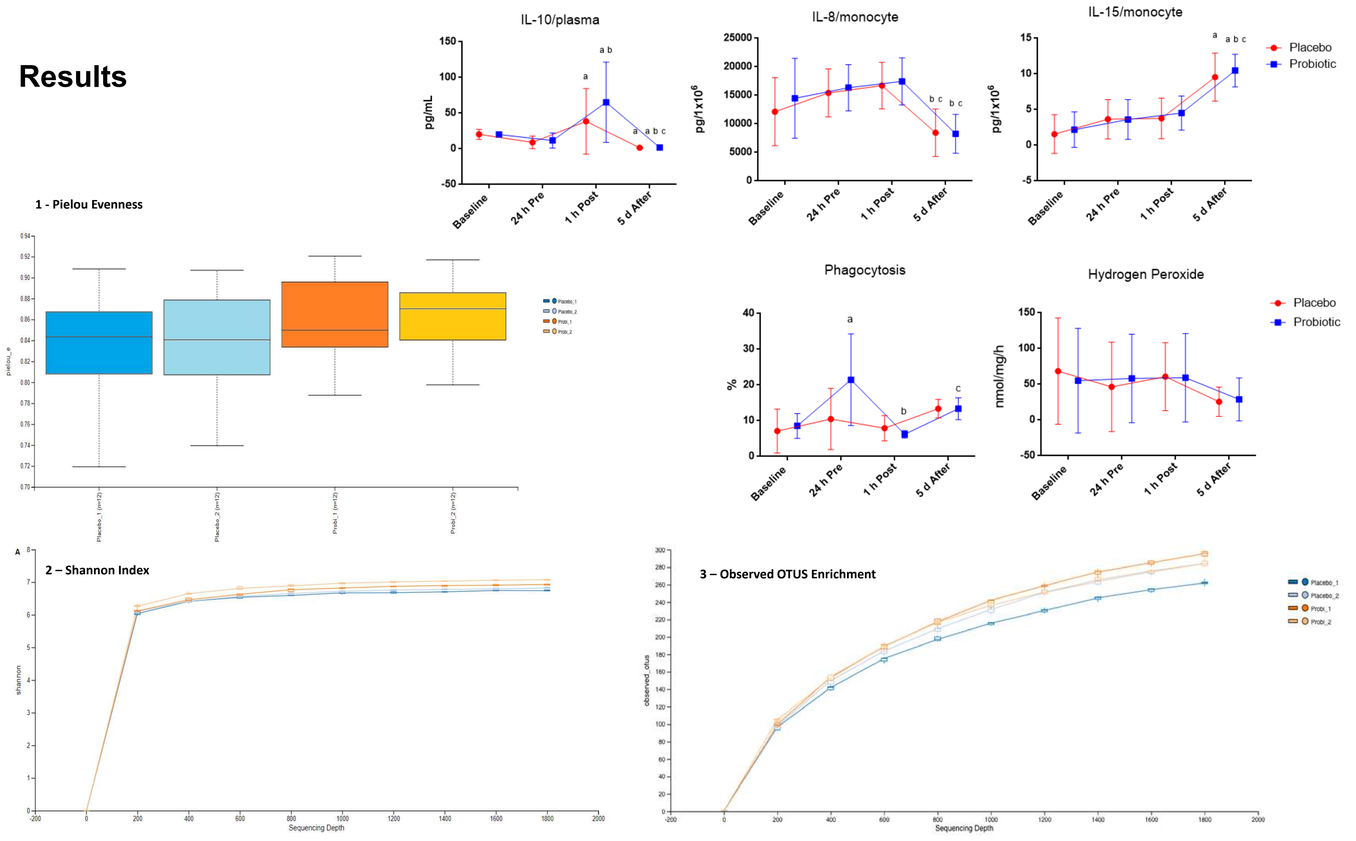
Materials and Methods

Marathon athletes (n=30) were allocated into placebo (maltodextrin 5g) or probiotic ($10x10^9$ CFU of *Lactobacillus acidophilus* LA-G80 and $10x10^9$ CFU of *Bifidobacterium animalis* subsp. *lactis* BL-G101) and received a double-blind supplementation for thirty days. Before the supplementation period (Baseline) and 24 hours before the race (24h-Pre), faeces were collected to analyze microbiota Alpha diversity. Blood was collected at four different times (Baseline, 24h-Pre, 1h-Post and 5 days after the race) to analyze monocyte's function and plasma cytokine. For ten days after the marathon race, athletes answered a checklist about symptoms of URTI. Bacterial genetic sequencing was based on the V3-V4 regions of rRNA 16S following Illumina's MiSeq platform system and visualization by Quantitative Insights Into Microbial Ecology – QIIME. The data normality was verified using the Shapiro-Wilk test, and the Anova Two-Way applied with a significance level of $p \leq 5\%$ for immune response.

Experimental Design



Results



Monocytes functions at Baseline, 24h Pre, 1h Post and 5 d After a marathon race. a.different from baseline. b. different from 24 h Pre. c. different from 1 h Post. $p \leq 5\%$. Alpha diversity 1- Pielou Evenness; 2 – Shannon Index; 3 – Observed OTUS.

Conclusions

Supplementation of $10x10^9$ CFU of *Lactobacillus acidophilus* LA-G80 and $10x10^9$ CFU of *Bifidobacterium animalis* subsp. *lactis* BL-G101 for thirty days was not sufficient to modify the microbiota Alpha diversity of marathon athletes and did not modify symptom parameters of opportunistic infections in the upper respiratory tract. However, probiotic supplementation was able to modulate the cellular response of monocytes, with a significant increase in phagocytosis rate after the supplementation period. Different studies have demonstrated the efficiency of probiotic supplementation on the immunological response and intestinal microbiota modulation in recent years. We believe it is necessary to investigate different doses and supplementation time for this specific population of marathon runners.

Acknowledgements: Financial Support: FAPESP #2016/25821-5 and we declare that there is no conflict of interest in research.

References

- Bermon S, Castell LM, Calder PC, Bishop NC, Blomstrand E, Mooren FC, Kruger K, Kavazis AN, Quindry JC, Senchina DS, Nieman DC, Gleeson M, Pyne DB, Kitic CM, Close GL, Larson-Mayer DE, Marcos A, Meydani SN, Wu D, Walsh NP, Nagatomi R. Consensus Statement Immunonutrition and Exercise. Exer. Immunology Review v. 23 p 24, 2017).
- Cox, A.J., Pyne, D.B., Saunders, P.U., Fricker, P.A. Oral Administration Of The Probiotic Lactobacillus Fermentum VRI-003 And Mucosal Immunity In Endurance Athletes. British Journal Of Sports Medicine V.44, N.4, 2008.
- Schiffrin, E. J. et al. (1995) „Immunomodulation of Human Blood Cells Following the Ingestion of Lactic Acid Bacteria", Journal of Dairy Science, 78(3), pp. 491–497.
- Nieman, D. C., Fagoaga, O. R., et al. (1997b) „Carbohydrate supplementation affects blood granulocyte and monocyte trafficking but not function after 2.5 h of running", American Journal of Clinical Nutrition, 66(1), pp. 153–159.
- Rinninella, E. et al. (2019) „What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases", Microorganisms, 7(1).

The gut microbiome in patients with juvenile idiopathic arthritis

Kelsey N Thompson[1,2,3], Kevin S Bonham[2,3],Elizabeth C Rosser[4,5], Emma Robinson[5,6], Emma Jordan[5,6], Fatema Jeralii[5,6], Hannah Peckham[4,5], Nicholas E Ilott[7], Lilian H Lam[7], Paula Colmenero[7], Sam J Bullers[7], the Inflammatory Arthritis Microbiome Consortium, Coziana Ciurtin[5,6], Lucy R Wedderburn[4,5], Fiona Powrie[7], and Curtis Huttenhower[1,2,3]

1:Harvard Chan Microbiome in Public Health Center, Harvard T. H. Chan School of Public Health 2:Department of Biostatistics, Harvard T. H. Chan School of Public Health 3:Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard 4:Centre for Adolescent Rheumatology Versus Arthritis 5:Division of Medicine, University College London 6;UCL GOS Institute of Child Health 7:Kennedy Institute of Rheumatology, Oxford University

Juvenile idiopathic arthritis (JIA) is among the many chronic systemic inflammatory diseases in which the gut microbiome has been implicated, but careful multi'omic studies of this complex condition have not previously been carried out. Samples were collected from children with juvenile enthesitis-related arthritis, oligoarticular JIA, polyarticular JIA, and psoriatic arthritis for a total of 113 JIA samples, plus 38 samples collected from age matched controls. We found that JIA subtype, current arthritis modifying drugs, and HLA-B27 status all explained a significant amount of taxonomic and functional variability within the gut ecosystem (PERMANOVA R2 2-3%, q<0.25). Limited joint count, which is used to determine clinical inflammation status, also explained nominal amounts of variation. Age of the patient, as expected, explained the largest portion of the taxonomic (but not functional) variation. We observed the loss of key gut commensals (e.g. *Faecalibacterium prausnitzii*) associated with inflammatory markers (e.g. limited joint counts) and increased prevalence and abundance of proinflammatory taxa such as *Streptococcus*. Additionally, sub-species phylogenetic associations were found with age, current drug, and JIA subtype. Functionally, several alterations across sulfur related cycling and metabolism and carrier protein biosynthesis were observed.

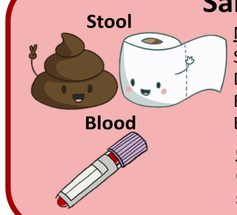
Experimental design for microbiome profiling in Juvenile Idiopathic Arthritis

Participants JIA

Age 1-15
PolyGos (n = **43**)
HC (n = **25**)

Age 16-30
ERA (n = **26**)
PolyGos (n = **14**)
HC (n = **11**)

Sample collection

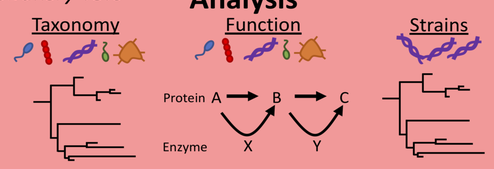


Microbial community analysis
Stored (fresh frozen/ethanol fixed)
DNA extracted (ethanol)
Extraction for NMR metabolomics (ethanol)
Extraction for LC-MS metabolomics (fresh/frozen)

Clinical metadata
C-reactive protein (CRP), eosinophil sedimentation rate (ESR), hemoglobin etc.

Analysis

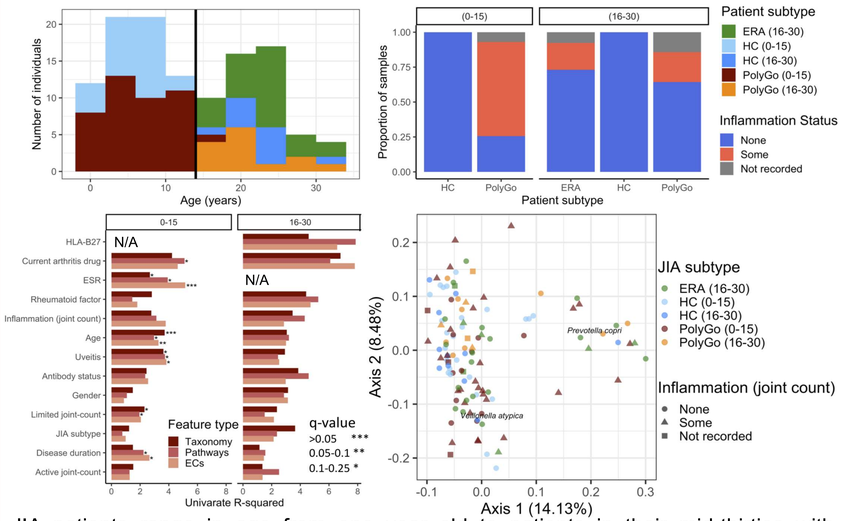
bioBakery v3.0



Statistics: PERMANOVA (community composition), mixed linear model (MaAsLin 2; per-feature)
Visualization: Ordination, bar plots density

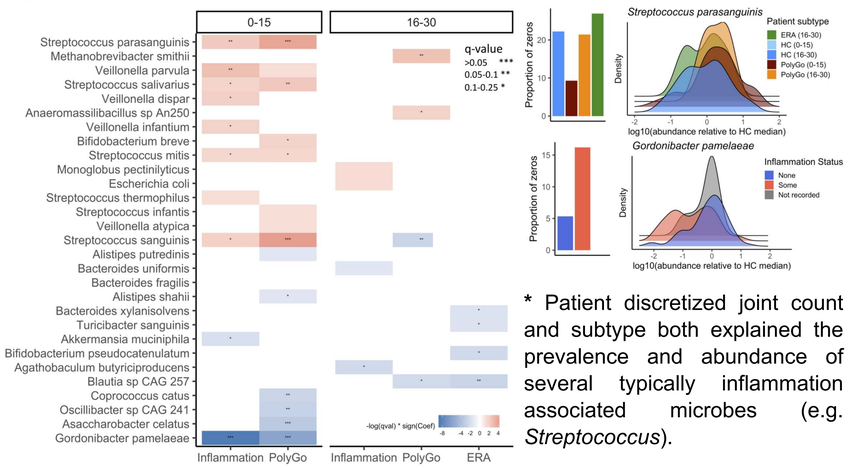
JIA subtype - age split analysis (0-15 and 16-30)
PolyGo: Oligoarticular JIA and Rheumatoid factor negative Polyarticular JIA
ERA: Juvenile enthesitis-related arthritis
HC: Healthy controls

The human gut microbiome is altered in inflammatory JIA



JIA patients range in age from one year old to patients in their mid-thirties with differences in diagnosis splitting out around sixteen years of age (top left). Active joint count was discretized into either no active joints ("none") or current inflamed joints at the time of sample collection ("some"). We used this new categorical variable as a systemic inflammatory marker (top right). Current patient inflammation (e.g. Eosinophil settling rate (ESR) and limited joint count) explained significant amounts of the overall compositional differents in gut (bottom left). JIA subtype did not explain significant compositional differences in the gut microbiome (bottom).

Patients with JIA exhibit altered clade prevelence and abundance

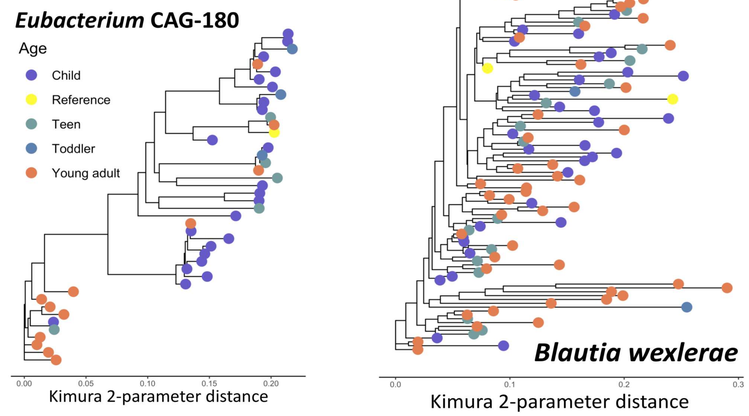


* Patient discretized joint count and subtype both explained the prevalence and abundance of several typically inflammation associated microbes (e.g. *Streptococcus*).

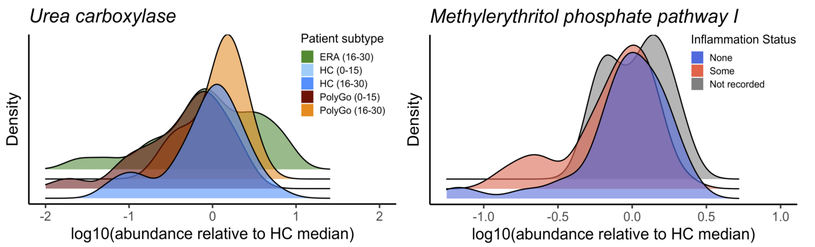
* Meanwhile, extending our results to the expanded databases within bioBakery v3 also allowed the discovery of several newly identified taxa associated with diagnosis (e.g. Gordonibacter pamelaeae).

Many clades exhibited gradient phylogenetic associations with age

* While investigating the subspecies ("strain") level associations with JIA, a stark trend of subspecies gradient structing with age was observed (17/57 species exhibited age associations). Here *Blautia wexlerae* and *Eubacterium* GAG180 are shown, both of which exhibited significant associations with patient age.



Functional alteration by subtype and inflammatory markers



* JIA subtype was associated with several functional classes of biomolecular pathways included amine and polyamine biosynthesis and degradation (left), aromatic molecule processing and vitamin B1 and B2 biosynthesis and salvage.
* While discretized active joint count was associated with isoprenoid biosynthesis (right), cell wall modifications, and glycan and amino acid biosynthesis.

Acknowledgments

We appreciate the valuable scientific contributions of Armaiti Batki and all members of the IAMC on this project. This work has been supported by supported by Arthritis Research UK. Methods used for analysis are available in the bioBakery at:

http://huttenhower.sph.harvard.edu

Evolutionary Implications of Reproductive Tract Microbiota in the Polygynandrous Red Junglefowl (*Gallus gallus*)

Liisa Veerus*, Allison M. Roth, Yunke Wang, Shorok B. Mombrikotb, Emma Ransome, Rosie Eccleston, Thomas Bell, Tommaso Pizzari

*Department of Zoology, Edward Grey Institute, University of Oxford, Oxford, UK & Center for Advanced Biotechnology and Medicine (CABM), Rutgers University, New Jersey, USA

Contact details: lv273@cabm.rutgers.edu, @LiisaVeerus

Abstract

Host social behavior is hypothesized to influence host-host transmission dynamics of gut microbiota. In turn, the gut bacteria may affect host social decision making [1]. A similar, yet under-explored evolutionary interdependence could also occur between host mating behavior and its reproductive microbiota [2]. We therefore characterized the reproductive tract microbiota of female and male red junglefowl (*Gallus gallus*; Figure 1) to investigate the potential role of reproductive bacteria in host evolution.

Figure 1. Female and male red junglefowl support sex-specific reproductive tract microbiota that may be exchanged between hosts during mating.

Background

- Human reproductive tracts harbor both pathogenic and mutualistic bacteria [3,4] that may be sexually-transmitted and affect host health [5]
- Little is known about whether reproductive bacteria could influence mate choice, determine compatibility between sexual partners, and affect socio-sexual host network structure [2]
- Interdependence between host evo-ecology and microbiota may be easier to investigate studying a non-human host, e.g., the red junglefowl—the main wild ancestor of the domestic chicken (*Gallus domesticus*)
- Female and male fowl mate with multiple partners (polygynandry), while practicing mate choice [6], allowing us to explore the possible evolutionary implications of exchanging reproductive bacteria during mating

Objectives

(1) Characterize previously unexplored bacteria found in female and male reproductive tracts, and ejaculates collected from males

(2) Investigate patterns of microbiota variation in sexually-interacting groups of individual hosts

Discuss the results in the context of host evolution and ecology.

Methods

Sterile collection of reproductive tract, ejaculate, and cloacal wipe samples + controls from the digestive tract

Bacterial DNA extraction and qPCR/PCR of the 16S SSU rRNA gene V4 hypervariable region + inclusion of negative and positive control samples

Illumina MiSeq sequencing

Bioinformatic processing with DADA2 and statistical analysis

Results

(1) Reproductive tract microbiota is sex-specific (A), with more pronounced bacterial diversity variation present in the female reproductive tract (B).

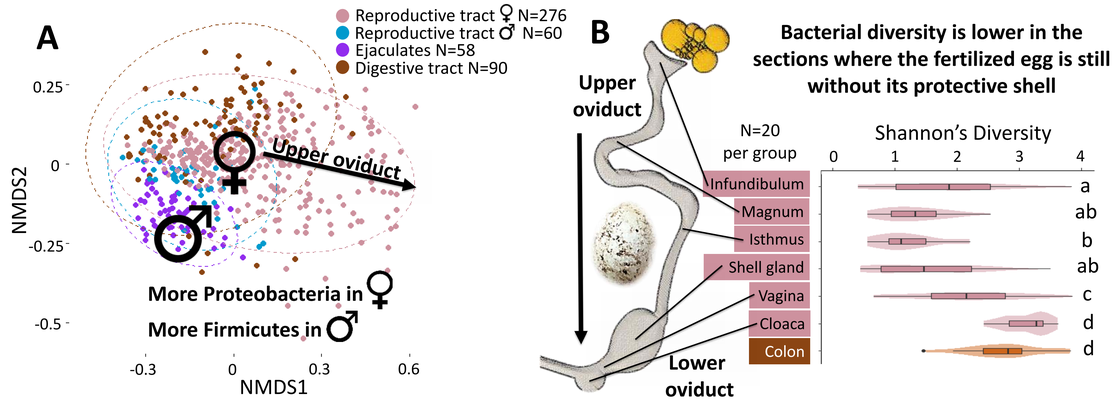


Figure 2. NMDS (stress=0.18) based on ASV-level Jaccard indices (A) and violin plots with indicated medians of Shannon's diversity in tissue samples (B). Significance (q<0.05) was assessed using FDR-corrected least-squares means (B).

Conclusion: Females show patterns suggestive of host-mediated microbiota maintenance in the upper oviduct, which may protect eggs from bacterial infections.

(2) Host cloacal microbiota is affected by the number of secured partners in the mating population (C). Bacterial exchange during mating is asymmetric between the sexes (D).

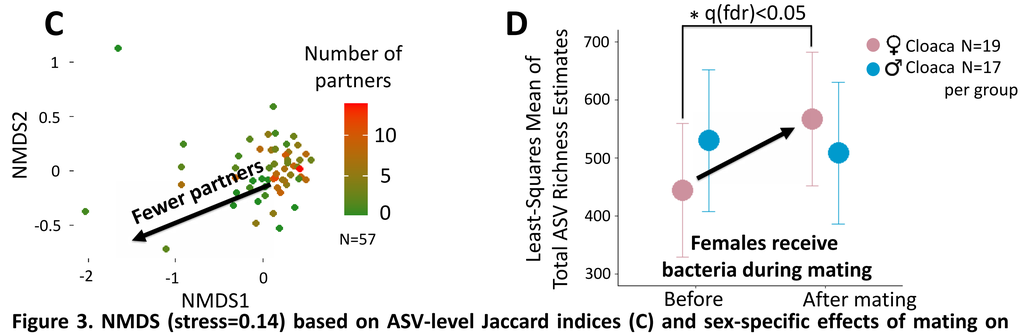


Figure 3. NMDS (stress=0.14) based on ASV-level Jaccard indices (C) and sex-specific effects of mating on cloacal microbiota diversity (D).

Conclusion: Mating challenges the established reproductive tract microbiota and could be particularly costly for females.

Future Directions

- Given that bacteria may come in contact with the unprotected fertilized egg, mother-to-offspring vertical bacterial transmission in egg-laying hosts merits more attention
- Further research is needed to test whether a specific reproductive tract microbiota underlies host attractiveness in a mating context, and elevated offspring number/health

References: [1] Archie EA &Tung J (2015) Social behavior and the microbiome. *Curr Opin Behav Sci*
[2] Rowe M, Veerus L, Trosvik P, Buckling A & Pizzari T (2020) The reproductive microbiome: an emerging driver of sexual selection, sexual conflict, mating systems, and reproductive isolation. *TREE*
[3] Chen X, Song X, Wei W, Zhong H, Dai J, Lan Z et al (2017) The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. *Nat Commun*
[4] Mändar R (2013) Microbiota of male genital tract: Impact on the health of man and his partner. *Pharmacol Res*
[5] Mändar R, Punab M, Borovkova N, Lapp E, Kiiker R, Korrovits P et al (2015) Complementary seminovaginal microbiome in couples. *Res Microbiol*
[6] Pizzari T & McDonald GC (2019) Sexual selection in socially-structured, polyandrous populations: Some insights from the fowl. *Advances in the Study of Behavior*

Acknowledgements: I would like to thank my funding by Oxford-Merton-NaturalMotion Graduate Scholarship and ESFUSA scholarships named after Ene Silla.

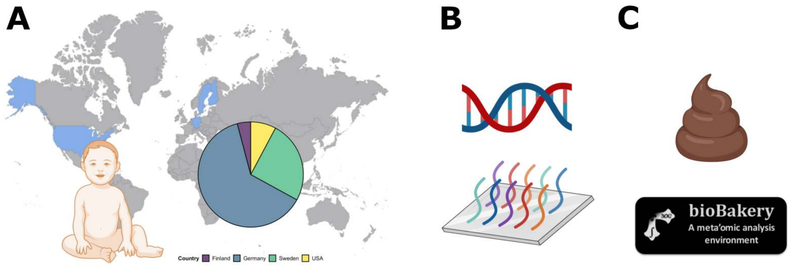
The role of host genetics in the development of the infant gut microbiome

Aaron M. Walsh[1,2], Tommi Vatanen[1,3], George Weingart[2],
Mondher Khidiri[1,2], Kendra Vehik[4], Eric A. Franzosa[1,2], and Curtis Huttenhower[1,2]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA, [2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA, 02115, USA,
[3]Liggins Institute, University of Auckland, Auckland, New Zealand, [4]Health Informatics Institute, University of South Florida, Tampa, FL, USA

The infant gut microbiome is crucial for the maturation of the host immune system, and dysbiosis of the infant gut microbiome has been linked to autoimmunity later in life. The infant gut microbiome, itself, is shaped by a number of factors, including the mode of delivery, breastfeeding, and antibiotics. Although studies in adults have reported that a handful of taxa are associated with mutations in genes that are involved in immunity, it is unclear if host genetics plays a role in shaping the infant gut microbiome. The aim of our study is to determine if host genetics interacts with the infant gut microbiome.

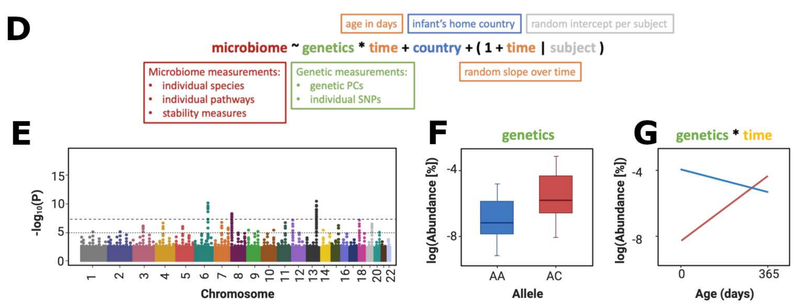
Cohort and study design



902 infants from the TEDDY cohort (Fig. 1 A).
- Infants at risk of developing type 1 diabetes (T1D).
- Infants recruited from 4 countries.

Infants were genotyped using the Illumina Infinium ImmunoArray (Fig. 1 B).
- 253,702 immune specific markers.
- Genotype data was analysed using PLINK 1.90.

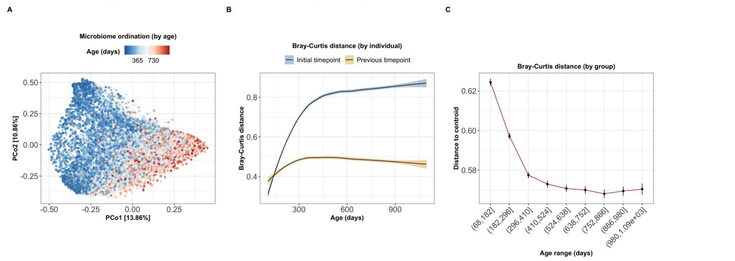
Shotgun metagenomics was used to profile the infant gut microbiome from 0-3 years (Fig. 1 C).
- 12,159 samples were analysed using bioBakery tools.



A random slopes model was used to identify the additive and indirect effects of genetics. on the microbiome (Fig. 1 D-G).

Figure 1. Schematic overview of the study.

Age dominates over genetics in shaping the early gut microbiome



The microbiome was characterized by convergence over time (Fig. 2 A-C).

The microbiome did not stratify by country (Fig. 2 D), whereas principal component analysis of genotype data revealed that Finnish infants separated along PC1 (Fig. 2 E). The Mantel test indicated that microbiome structure was not correlated with genetic structure (Fig. 2 F).

Figure 2. The structures of the data within the TEDDY cohort.

A subset of individual species is associated with genetic structure

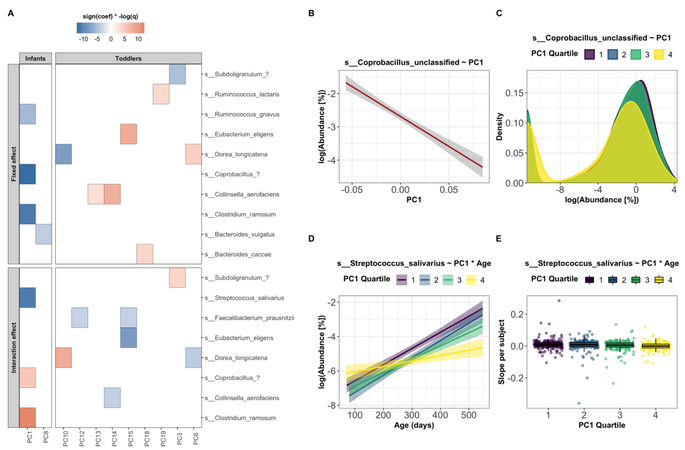


Figure 3. The species associated with genetic principal components.

10 species displayed significant additive effects (e.g. Fig. 3 B-C), while 8 species displayed significant interaction effects (e.g. Fig. 3 E-F).

Variation in the MHC region influences the microbiome

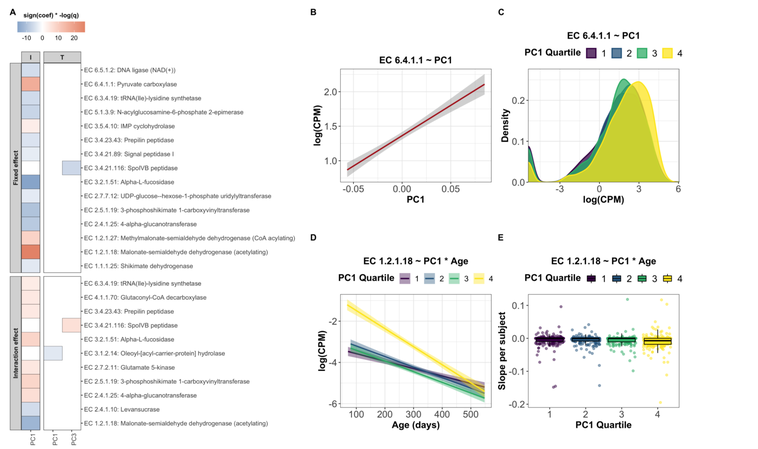


Figure 4. The abundances of 19 level 4 Enzyme Commission (EC) categories associated with genetic principal components.

15 ECs displayed significant additive effects (e.g. Fig. 3 G-H), while 11 ECs displayed significant interaction effects.

Gene set enrichment analysis of SNP loadings revealed that variants in the major histocompatibility complex (MHC) drove PC1, which was had the greatest number of associations with the microbiome.

A subset of microbiome features associates with individual SNPs

10 species were significantly associated with ≧1 SNP.

42 ECs were significantly associated with ≧1 SNP.

Bacteroides dorei is associated with the MHC region and PFKB3 (Fig. 5 A).

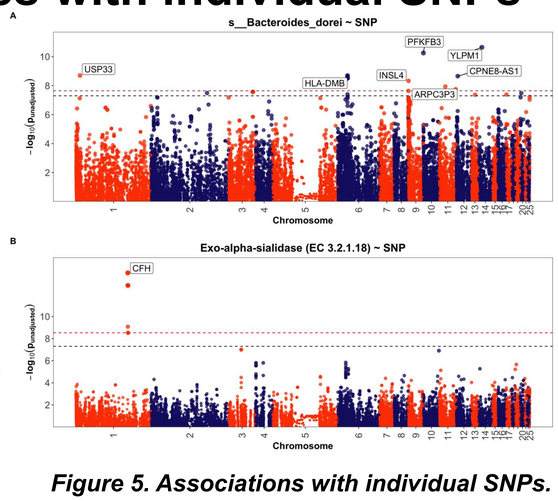


Figure 5. Associations with individual SNPs.

Exo-alpha-sialidase is associated with CFH (Fig. 5 B).

http://huttenhower.sph.harvard.edu

The modulating role of the gut microbiome in body weight responses to physical activity

K. Wang[1], W. Ma[2,3], LH. Nguyen[2,3,4], D. Wang[5], RS. Mehta[2,3], D. Hang[5,6,7], L. Al-Shaar[5], CH. Pernar[1], CH. Lo[2,3], B. Fu[1], S. Ogino[1,8,9,10], EB. Rimm[1,5,11], FB. Hu[1,5,11], WS. Garrett[12,13], Q. Sun[5,11], AT. Chan[2,3,10,11,12], C. Huttenhower[4,10,12], M. Song[1,2,3,5]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA [2]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA [3]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA [4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA [5]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA [6]Department of Epidemiology and Biostatistics, International Joint Research Center on Environment and Human Health, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China [7]Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China [8]Department of Oncologic Pathology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA [9]Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA [10]Broad Institute of MIT and Harvard, Cambridge, MA [11]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA [12]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA [13]Department of Medicine, Harvard Medical School, Boston, MA

INTRODUCTION

Although physical activity (PA) is a major approach to weight control, PA does not always result in expected weight loss and shows high individual variability in body weight responses.

The gut microbiome plays an important role in host energy balance, but how the gut microbiome modulates the body weight response to PA remains unknown.

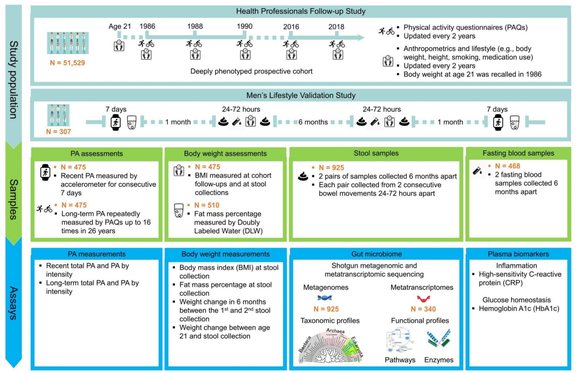
AIM

The current study analyzed the gut microbiomes profiled by shotgun metagenomics and metatranscriptomics sequencing in modifying the association of recent and long-term PA with body weight measures and relevant plasma biomarkers.

METHOD

- We collected data on PA type and intensity and body weight using the validated biennial questionnaires since 1986 from 51,529 men enrolled in the Health Professionals Follow-up Study.
- In a subcohort of 307 healthy men, we collected up to 2 pairs of stool samples and 2 blood samples, 6 months apart in 2012-2013.
- We profiled 925 stool metagenomes, 340 stool metatranscriptomes, and 468 blood samples.
- One month before and after the stool collections, participants were asked to wear accelerometer for consecutive 7 days to monitor their PA and received Doubly Labeled Water (DLW) test for body weight and fat mass assessment.
- We assessed the overall gut microbiome configurations, microbial species abundances, microbial functional pathways and enzymes in relation to
  - recent PA level measured by accelerometer,
  - long-term PA level from questionnaires,
  - body mass index and fat mass percentage measured by DLW,
  - short-term body weight change in 6 months between the 1st and 2nd stool collection,
  - long-term body weight change from age 21 to stool collection, and
  - plasma high-sensitivity C-reactive protein (CRP) and hemoglobin A1c (HbA1c) levels at stool collection.
- We then examined how the microbial species might modify the associations of PA with the body weight measures and biomarkers.

RESULTS



b

Variable	Quartiles of total PA measured by accelerometer (MET-hours/week)			
	Q1	Q2	Q3	Q4
	0-15.8	15.9-22.4	22.5-29.9	30.0-73.4
N of individuals (%)	86 (28)	79 (26)	75 (24)	67 (22)
N of stool samples (%)	230 (25)	231 (25)	231 (25)	233 (25)
Age at stool collection, years	72.6 (5)	70.9 (4.1)	70.8 (4.1)	69.8 (3.2)
PA by intensity measured by accelerometer, MET-hours/week				
Vigorous	0.1 (0.9)	0.8 (2.4)	1.9 (4.4)	5.8 (11.1)
Moderate	3.7 (5.4)	6 (6)	10.4 (7.7)	18.4 (12.2)
Light	10.2 (8.1)	11.1 (6.9)	13.8 (10.5)	14.6 (9.4)
Long-term average PA from 1986 to stool collection, MET-hours/week				
Total	33.2 (20.6)	46.2 (22.1)	47.4 (19.9)	61.4 (27.5)
Vigorous	10.2 (14.4)	16.5 (17.5)	18.1 (13.8)	24.2 (23)
Moderate	15.8 (12.6)	23 (14.6)	22.4 (15.9)	30.4 (20)
Light	3.1 (2.8)	2.9 (2.4)	3.3 (2.7)	3.5 (3)
BMI at stool collection, kg/m^2	28.2 (6)	26.3 (3.6)	25.2 (3.1)	24.8 (2.9)
Fat mass percentage at stool collection, %	32.2 (7.6)	28.4 (6.4)	27 (5.4)	26 (6.6)
Weight change in 6 months between the 1st and 2nd stool collection, kg	1.1 (0)	1.4 (1.6)	3.6 (0)	0.5 (0)
Weight change between age 21 and stool collection, kg	13.3 (13.2)	8 (10.6)	5.9 (9.4)	5.3 (8.7)
Current smoker at stool collection, %	2.2	2.6	1.7	1.7
Total energy intake, kcal/d	1941.2 (465.5)	2129.2 (481.1)	2089 (448.7)	2257.5 (498)
Used probiotics in 2 months, %	5.2	7.8	6.9	7.3
Used antibiotic in 12 month, %	26.5	28.6	29.9	19.3
Bristol score, %				
1-2, hard stool	15.2	10.2	16.9	10.3
3-5, normal stool	70.9	71.5	65.8	73.4
6-7, loose stool	13.9	16.6	17.3	16.3
CRP, mg/dL	2.8 (3.8)	2 (3.2)	2.1 (3.4)	1.4 (2.2)
HbA1c, %	5.8 (0.4)	5.8 (0.4)	5.7 (0.4)	5.7 (0.3)

Figure 1. Study design for linking physical activity (PA), body weight measures, plasma biomarkers, and the gut microbiome in the Men's Lifestyle Validation Study (MLVS) and characteristics of participants in the MLVS. a, To associate the gut microbiome with PA and body weight measures, we profiled stool metagenomes, metatranscriptomes and blood biomarkers from the MLVS. The MLVS is a sub-study of the HPFS, an ongoing prospective cohort totaling 51,529 men. The HPFS has repeatedly collected PA and body weight information using validated questionnaires and health-related information since 1986. In the period 2011-2013, the MLVS collected stool samples at up to four time points per individual, blood samples at up to two time points, and additional PA information using accelerometer and body weight information using Doubly Labeled Water (DLW) from 307 participants. We applied MetaPhlAn 2 and HUMAnN 2 to perform taxonomic and functional profiling from stool shotgun metagenomes and metatranscriptomes. Plasma biomarkers of inflammation (high-sensitivity C-reactive protein, CRP) and glucose homeostasis (hemoglobin A1c, HbA1c) were measured using standard methods. We employed generalized linear mixed-effects regression models to account for within-subject correlation due to repeated sampling and occasional missing data. **b**, Characteristics of participants in the MLVS according to quartiles of PA levels measured by accelerometer. Values are means for continuous variables and percentages for categorical variables. The variable of weight change in 6 months between the 1st and 2nd stool collection was based on the data collected at the 1st stool collection.

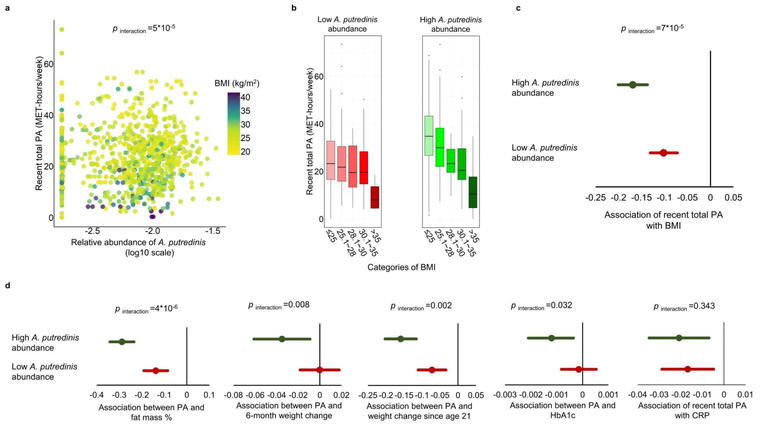


Figure 3. *Alistipes putredinis* abundance modulates the associations of physical activity (PA) with body weight measures and plasma biomarkers. The interactions between PA and *A. putredinis* abundance (with median level as cutoff for low and high abundance) are significant in relation to all of body mass index (BMI), fat mass percentage, short-term and long-term body weight changes, and plasma hemoglobin A1c (HbA1c) level. **a**, The interaction between recent total PA and *A. putredinis* abundance in relation to BMI. **b**, Recent total PA levels in relation to BMI among participants with low and high *A. putredinis* abundance separately. Box plot centers show medians of the PA measures with boxes indicating their inter-quartile ranges (IQRs), upper and lower whiskers indicating 1.5 times the IQR from above the upper quartile and below the lower quartile, respectively. **c,** Association between recent total PA and BMI according to *A. putredinis* abundance. The dots in the plot indicate beta coefficients in the multivariable-adjusted generalized linear mixed-effects regression models, with error bars indicating upper and lower limits of their 95% confidence intervals. Beta coefficients and $P_{interaction}$ were calculated from multivariable-adjusted generalized linear mixed-effects regression models while adjusting for age, smoking, total energy intake, probiotic use, antibiotic use, and Bristol stool scale. **d**, Associations between PA levels with other body weight measures, including fat mass percentage, short-term body weight change (6-month weight change), long-term body weight change (weight change between age 21 and stool collection), plasma HbA1c and high-sensitivity C-reactive protein (CRP) levels.

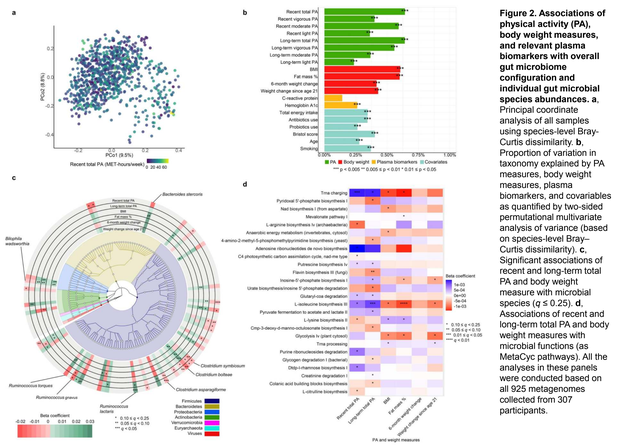


Figure 2. Associations of physical activity (PA), body weight measures, and relevant plasma biomarkers with overall gut microbiome configuration and individual gut microbial species abundances. a, Principal coordinate analysis of all samples using species-level Bray-Curtis dissimilarity. **b**, Proportion of variation in taxonomy explained by PA measures, body weight measures, plasma biomarkers, and covariables as quantified by two-sided permutational multivariate analysis of variance (based on species-level Bray–Curtis dissimilarity). **c**, Significant associations of recent and long-term total PA and body weight measure with microbial species ($q \leq 0.25$). **d**, Associations of recent and long-term total PA and body weight measures with microbial functions (as MetaCyc pathways). All the analyses in these panels were conducted based on all 925 metagenomes collected from 307 participants.

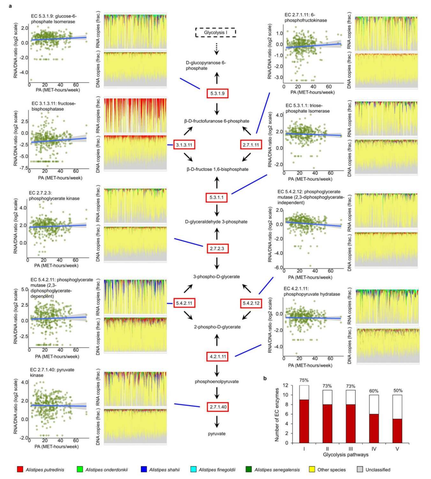


Figure 4. *Alistipes putredinis* is highly involved in Glycolysis pathways and enzymes. a, *A. putredinis* contributes to nine microbial enzymes within the pathway of Glycolysis I, using 340 metatranscriptome and metagenome pairs from 96 participants. The plots are schematic representations of the pathway of Glycolysis I containing nine key enzymes for Glycolysis I. We used EC numbers in the rectangles to represent these enzymes; the rectangles are in red indicates that all the nine enzymes are contributed by *A. putredinis*. The scatter plots show the associations of recent total PA with relative abundance or transcription levels of microbial enzymes. The bar plots show the *Alistipes* species with the greatest contributions to each microbial enzyme, with metagenomic or metatranscriptomic samples along the x axes ordered by recent total PA level (from the lowest to the highest). All statistical tests were two-sided. **b,** The number and percentage of key EC enzymes contributed by *A. putredinis* (in red) among all key enzymes (the whole bar) within each of the five Glycolysis pathways.

CONCLUSIONS

- Individuals with a higher *A. putredinis* abundance may have a better body weight response to PA.
- The modulating role of *A. putredinis* may be partly attributed to its roles in Glycolysis.
- More studies are needed to elucidate the potential of *A. putredinis* as a probiotic in improving body weight response to PA.

ACKNOWLEDGEMENTS

We thank the participants and staff of the MLVS and funding supports from NIH (P01 CA87969, U01 CA186107, P01 CA55075, UM1 CA167552, U01 CA167552, K24 DK098311, R01 CA137178, R01 CA202704, R01 CA176726, K99 CA215314, R00 CA215314) and American Cancer Society (MRSG-17-220-01 - NEC).

CONTACT INFORMATION

Mingyang Song (mingyangsong@mail.harvard.edu)

Kai Wang (kaiwang@hsph.harvard.edu)

The authors have no conflict of interest to disclose.

Characterizing microbial community viability with RNA-based high-throughput sequencing

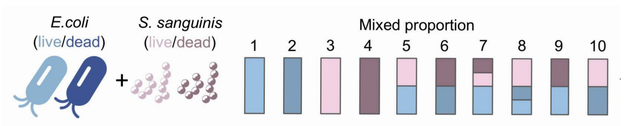
Ya Wang[1,2],Yan Yan[1,2], Kelsey N. Thompson[1], Sena Bae[1,2], Jiaxian Shen[3],
Hera Vlamakis[2], Erica M. Hartman[3], Curtis Huttenhower[1,2]

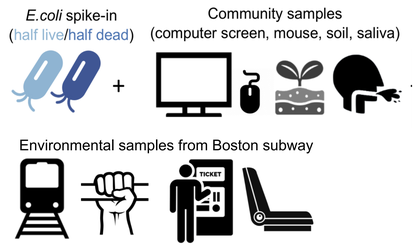
[1]Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard [3]Northwestern University

Characterizating microbial community viability is of great importance: essentially all sequence-based technologies do not differentiate living from dead microbes, whereas the functions of microbial communities are defined by biochemically active ("viable") organisms. As a result, our understanding of microbial community structures and their transmission mechanisms between humans and our surroundings remains incomplete. As a potential solution to this, RNA-based amplicon sequencing has been proposed as a method to quantify the viable fraction of a microbial community, but its reliability has not been evaluated systematically. Here, we present our work benchmarking 16S-RNA-seq (targeting 16S rRNA transcripts and genes for parallel RNA and DNA sequencing) for viability assessment in synthetic and realistic microbial communities, as well as exploring the potential application of protein-coding genes as viability markers in these settings. In synthetic communities, 16S-RNA-seq successfully reconstructed the viable fraction of *Escherichia coli* and *Streptococcus sanguinis*. However, no significant compositional differences were observed in human and environmental microbial communities spiked with known *E.coli* controls. Results were slightly different in environmental samples of similar origins (i.e. from Boston subway systems), where samples were differentiated both by sources as well as by library type. Finally, we explored the use of protein-coding genes as viability markers in synthetic communities, where the chaperonin-encoding gene *cpn60* showed promising qPCR results. Overall, these results show that 16S rRNA amplicons do not reflect microbial viability outside of very simple, synthetic "communities"; alternatively, mRNA amplicon from protein-coding genes may be promising viability markers in natural microbial consortium that worth further exploration.

16S-RNA-seq for viability assessment

We first evaluated 16S-RNA-seq in ten synthetic communities comprising live and/or heat-killed *E.coli* and *S.sanguinis* mixed at different ratios.
Total DNA and RNA were extracted simultaneously, with RNA samples immediately reverse transcribed into cDNA followed by amplicon-sequencing targeting the V4 region of 16S rRNA for all cDNA and their respective DNA samples in parallel.



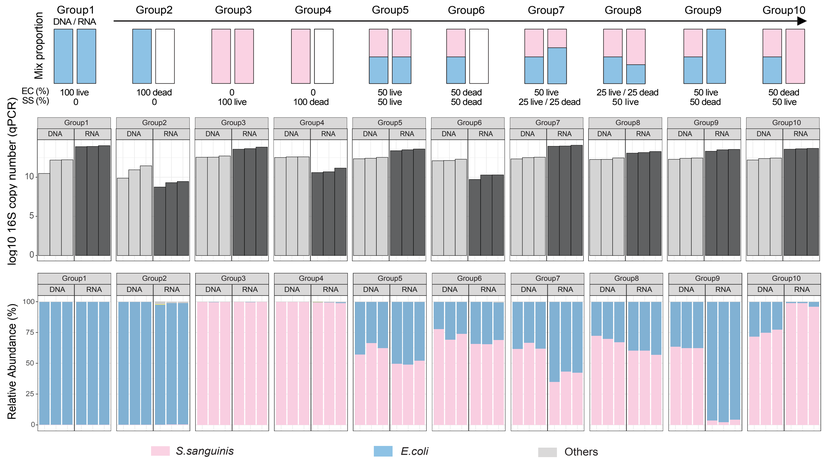


To evaluate its performance under variable biomass, biochemical background and diversity conditions, we next applied 16S-RNA-seq to swabbings of natural microbial communities from computer screens and mice, soil, and human saliva spiked with known concentrations of living and heat-killed *E. coli* controls.

We further validated this technique in built environment microbiome samples collected from Boston subway systems. Surfaces were swabbed from four seats, four walls and four grips in the train cart of the Green Line E branch, and three touchscreens of the ticket machines at Park Street Station.

For 16S-RNA-seq, we had a total of 170 samples comprising of those from synthetic cultures, *E. coli* spike-in experiments, subway microbial communities and several experimental and technical controls. qPCR was performed targeting 16S rRNA gene V4 region and protein-coding gene *cpn60* and *rpoB* on the samples of synthetic cultures to determine the viable (RNA/cDNA) and overall (DNA) bacterial mass.

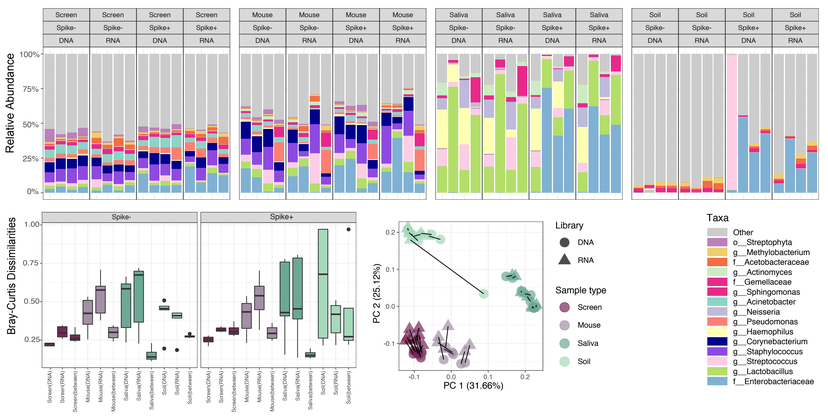
16S-RNA-seq successfully reconstructs the viable fraction in simple synthetic communities



16S-RNA-seq was able to qualitatively differentiate viable from nonviable microbes simple synthetic "communities".

- Viability accurately assessed in groups containing monoculture of viable cells (Groups 1, 2, 9 and 10);
- In mixed culture groups (Groups 5, 7 and 8), the composition of the two microorganisms agreed with the known proportion, though the abundances differed slightly;
- qPCR signals from RNA (cDNA) samples were counterintuitively higher compared to the DNA ones except for the groups containing mostly "dead" cells (Groups 2, 4 and 6), suggesting possible inaccuracy existed in the quantitation using 16S rRNA as marker.

16S-RNA-seq is not able to differentiate the viable vs. whole microbiome in spiked realistic communities



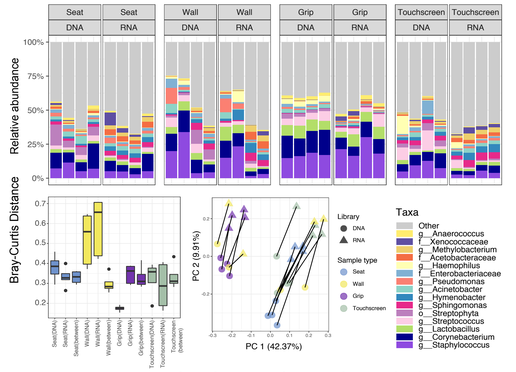
In preliminary spiked natural community samples, 16S-RNA-seq produced almost no differentiation between DNA vs. RNA libraries, i.e. total vs. "viable" microbes.

- Bray-Curtis dissimilarities between DNA and RNA libraries are small;
- Samples were grouped by source while not by library types in principle coordinate analysis.

16S-RNA-seq is able to differentiate viable vs. whole microbial communities in samples of similar origins

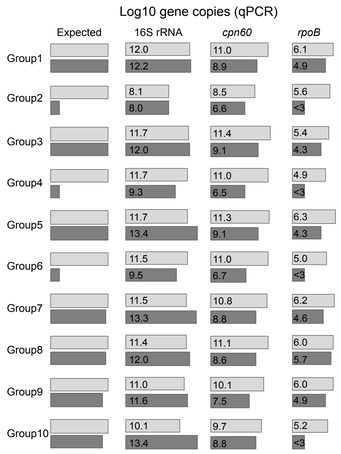
Results were slightly different in samples of similar origins (i.e., from subway systems), where the viable communities could be somewhat distinguished from the whole community.

- Library type made a significant contribution to the overall dissimilarities (q value = 0.014);
- Sample type remained as the major effect driving the compositional differences;



- The abundances of some human commensals changed consistently between DNA and RNA libraries, such as *Staphylococcus*, *Corynebacterium* and *Lactobacillus*, suggestting that some microbes are more or less visible in 16S rRNA vs. DNA amplicons, regardless of actual viability.

Protein-coding transcripts provide potential targets for viability assessment



Near-universal DNA amplicons such as *rpoB* and *cpn60* have been developed as phylogenetic sequence markers amenable to PCR amplification, raising the possibility that "standard" protein coding genes may represent promising targets for viability assessment using RNA amplification. These targets were tested in the ten synthetic cultures by qPCR.

- 16S-rRNA amplicons are not directly, quantitatively enriched for viable microbes.
- The chaperonin-encoding housekeeping gene, *cpn60*, presented qPCR results closer to the expectations in simple synthetic communities, suggesting that it is promising as an amplicon marker for viability assessment in natural microbial communities.

Conclusions

Overall, our results reveal that 16S-RNA-seq is able to profile viable microbial communities in simple "communities", but may be premature for application in realistically complex community samples. The immediate next step is to evaluate protein-coding gene *cpn60* as a target for parallel RNA and DNA sequencing (cpn60-RNA-seq) for viability assessment in natural microbial communities.

Acknowledgments

We are grateful to the Boston MBTA and Transit Police for their assistance with this research, particularly for ensuring that study personnel and subway passengers were safe and informed. Methods for the analysis are available from the bioBakery workflows at:

http://huttenhower.sph.harvard.edu

The Harvard T.H. Chan School of Public Health Microbiome Analysis Core

Jeremy E. Wilkinson[1,2], Lauren J. McIver[1,2], Chengchen Li[1,2], Thomas M. Kuntz[1,2], Curtis Huttenhower[1,2,3]

[1]Harvard Chan Microbiome in Public Health Center [2]Department of Biostatistics, Harvard T.H. Chan School of Public Health

[3]Broad Institute of MIT and Harvard

HCMPH MICROBIOME CENTER

The Microbiome Analysis Core at the Harvard T.H. Chan School of Public Health was established in response to the rapidly emerging field of microbiome research and its potential to affect studies across the biomedical sciences. The Core's goal is to aid researchers with microbiome study design and interpretation, reducing the gap between primary data and translatable biology. The Microbiome Analysis Core provides end-to-end support for microbial community and human microbiome research, from experimental design through data generation, bioinformatics, and statistics. This includes general consulting, power calculations, selection of data generation options, and analysis of data from amplicon (16S/18S/ITS), shotgun metagenomic sequencing, metatranscriptomics, metabolomics, and other molecular assays. The Microbiome Analysis Core has extensive experience with microbiome profiles in diverse populations, including taxonomic and functional profiles from large cohorts, qualitative ecology, multi'omics and meta-analysis, and microbial systems and human epidemiological analysis. By integrating microbial community profiles with host clinical and environmental information, we enable researchers to interpret molecular activities of the microbiota and assess its impact on human health.

Services

Study Design
- Consultation
- Grant assistance
- Power analysis
- Collection methods
- Wet lab
- Dry lab

Analysis
- Raw data processing
- Taxonomic and functional profiling
- Downstream analysis and statistics

Interpretation
- Results
- Discussion
- Manuscript writing/editing
- Response to reviewers

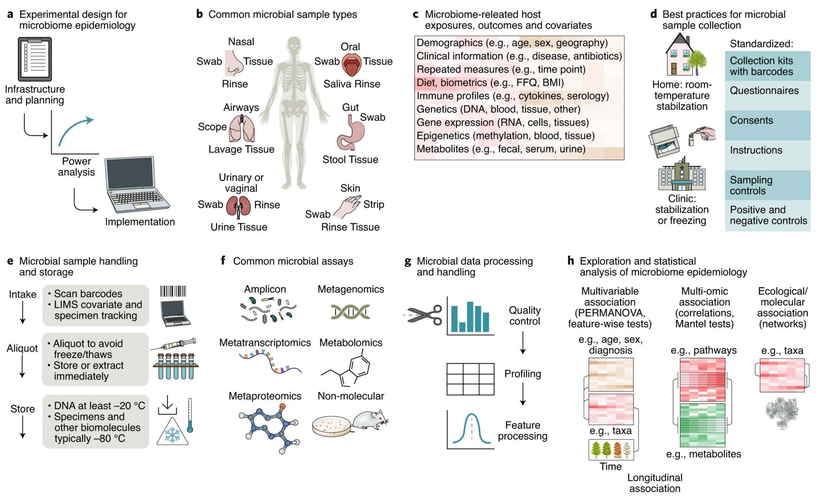
Consultation for microbiome project development
This includes consultation on experimental design, sample collection and sequencing, grant proposal development, study power estimation, bioinformatics, and statistical data analysis.

Validated end-to-end meta'omic analysis of microbial community data
Using open-source analytical methods developed in the Huttenhower laboratory and by other leaders in the field, we provide cutting-edge microbiome informatics and analysis.

Support fully-collaborative grant-funded investigations and others
Includes preliminary data development, hypothesis formulation, grant narrative development, data analysis and inference, custom software development, and co-authored dissemination of findings. We can work on projects in which we are included on the grant as well as projects with existing funding.

Fee-for-service, cost-recovery core laboratory
We support researchers from a variety of institutions including academic and industry. Our costs are based on an hourly rate of person time and the amount of hours vary depending on the study. Please contact us for more information and a quote.

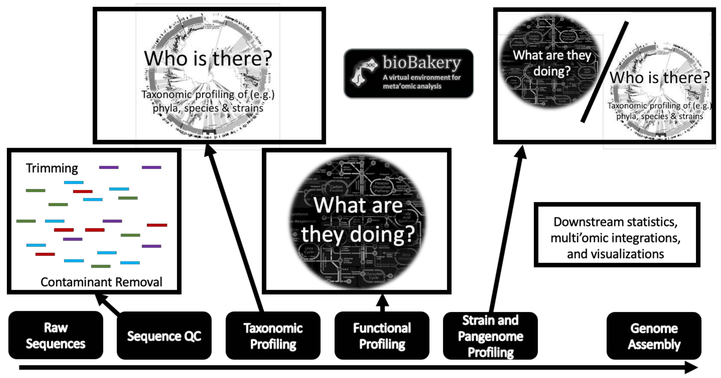
Microbial epidemiology multi'omics



The Microbiome Analysis Core supports microbiome analysis for a variety of molecular data types. A summary of stages to plan for and considerations at each step of microbiome study design and execution. **a**, Microbiome studies require appropriate physical infrastructure and planning. Power calculations for microbiome studies can be challenging, given the diverse types of measurement possible, their technical variability, and biological factors that can simultaneously influence the microbiome. **b**, Biopecimens can be collected from diverse sites with varying processing considerations. **c**, Microbiome analyses are reliant on extensive exposures and phenotypes, as many can affect microbial community state. **d**, It is critical to have streamlined, hygienic kits, questionnaires, and instructions that are also compatible with downstream sample management and assays. **e**, Often in combination with controls, specimens can be aliquoted and/or stored long term, or immediately reduced to molecular components. **f–h**, A wide range of culture-based and culture-independent assays can be applied to population-derived biospecimens (f), after which the resulting molecular data are handled bioinformatically (g) and statistically (h).

Wilkinson, J.E., Franzosa, E.A. et al. A framework for microbiome science in public health. Nature Medicine (2021).

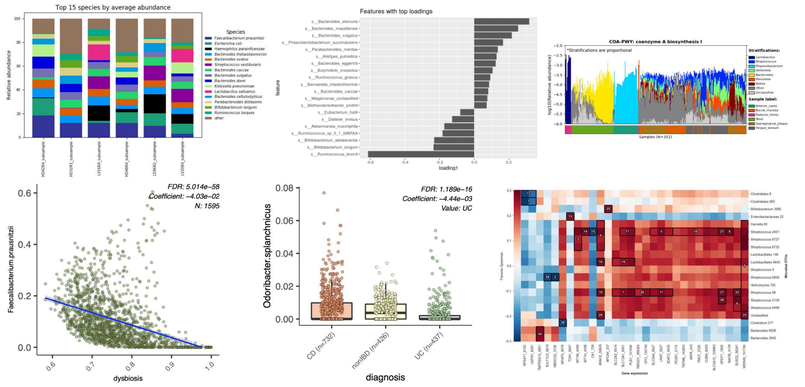
Microbial community profiling



The first step in microbiome molecular data analysis is quality control (KneadData) and profiling to transform raw data into biologically interpretable features using a reproducible workflow (bioBakery). This includes identifying microbial species (MetaPhlAn) and strains (PanPhlAn/StrainPhlAn), characterizing their functional potential or activity (HUMAnN), and integrating metagenomics with other data.

McIver, L. J. et al. bioBakery: a meta'omic analysis environment. Bioinformatics, 34:7, 1235-1237 (2018).

Downstream analysis and statistics



Once profiled, microbial communities are amenable to downstream statistics and visualization much like other molecular epidemiology data such as human genetic or transcriptional profiles. Like these other data types, microbial communities often require tailored statistics for environmental, exposure, or phenotype association (LEfSe, MaAsLin, HAllA) or for ecological interaction discovery (BAnOCC). The Harvard Chan Microbiome Analysis Core provides a variety of analyses for researchers working in the microbiome space.

The HCMPH BIOM-Mass platform

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH
Microbiome Analysis Core

- Harvard Chan Microbiome Collection Core
- Harvard Chan BiOS Freezer Core
- Harvard Chan Microbiome Analysis Core
- Harvard Chan Gnotobiotic Facility
- BIOM-Mass Data Portal

The Harvard Chan Microbiome Analysis Core is a part of the Harvard Chan Microbiome in Public Health Center (HCMPH) and the BIOM-Mass platform for microbiome research.

Want to learn more? Visit **https://hcmph.sph.harvard.edu**

Director: Jeremy E. Wilkinson
Senior Software Developer: Lauren J. McIver
Research Project Manager and Data Analyst: Chengchen (Cherry) Li
Postdoctoral Research Fellow: Thomas M. Kuntz
Scientific Director: Curtis Huttenhower

https://hcmph.sph.harvard.edu/hcmac
https://huttenhower.sph.harvard.edu

Identifying strain-specific functional genes in colorectal cancer

Yan Yan[1,2], Andrew M. Thomas[3], Kelsey N. Thompson[1,2], Paolo Manghi[3], Lauren J. Mciver[1,2], Eric A. Franzosa[1,2]
Nicola Segata[3], Andrew T. Chan[4,5], Wendy S. Garrett[2,6,7,8,9], Curtis Huttenhower[1,2,6*]

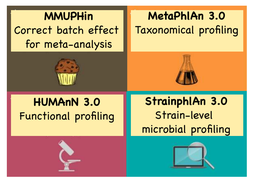
[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, [2]Broad Institute of MIT and Harvard, [3]CIBIO Department, University of Trento, Trento, Italy
[4]Clinical and Translational Epidemiology Unit, [5]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School,
[6]Department of Immunology & Infectious Diseases, Harvard T.H. Chan School of Public Health

Changes in the gut microbiota have been associated with colorectal cancer (CRC), but neither the causal mechanisms nor corresponding microbial strains and small molecule products have been elucidated for CRC. We have developed a new strain-level meta-analysis using stool metagenomic profiles of 600 CRC patients, 143 with pre-cancerous adenomas, and 662 healthy controls from nine recently published CRC microbiome studies. We created the MMUPHin framework to jointly normalise these datasets and identify potential consistently significant links between CRC neoplasia, severity, and microbial species and strains. We identified several species as novel CRC biomarkers including several typical oral species. We observed that CRC cases were depleted in geographically-specific *Prevotella copri* subtype carriers. A group of functional genes unique to subsets of *Escherichia coli* strains was associated with CRC phenotypes, comprising annotations to transporters, type II secretion systems, flagellar and sulfur metabolism. This study adds further evidence to the hypothesis that strain-level genomic variation in gut microbes may be a major driver in the initiation or development of CRC.

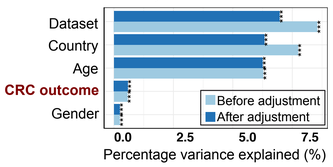
Expanded metagenomes and methods for meta-analyzing the CRC microbiome

A. Comparison of size of previous and expanded CRC datasets

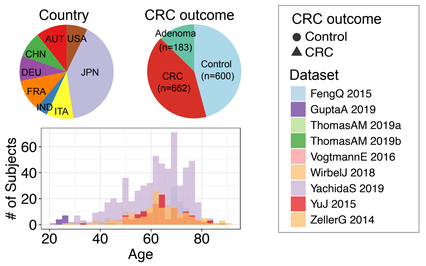
C. Microbiome profiling (bioBakery 3.0)



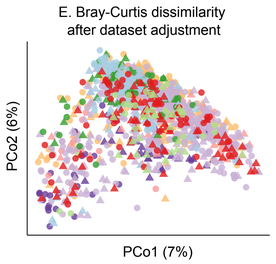
D. Bray-Curtis PERMANOVA



B. Size and characteristics of the expanded CRC dataset

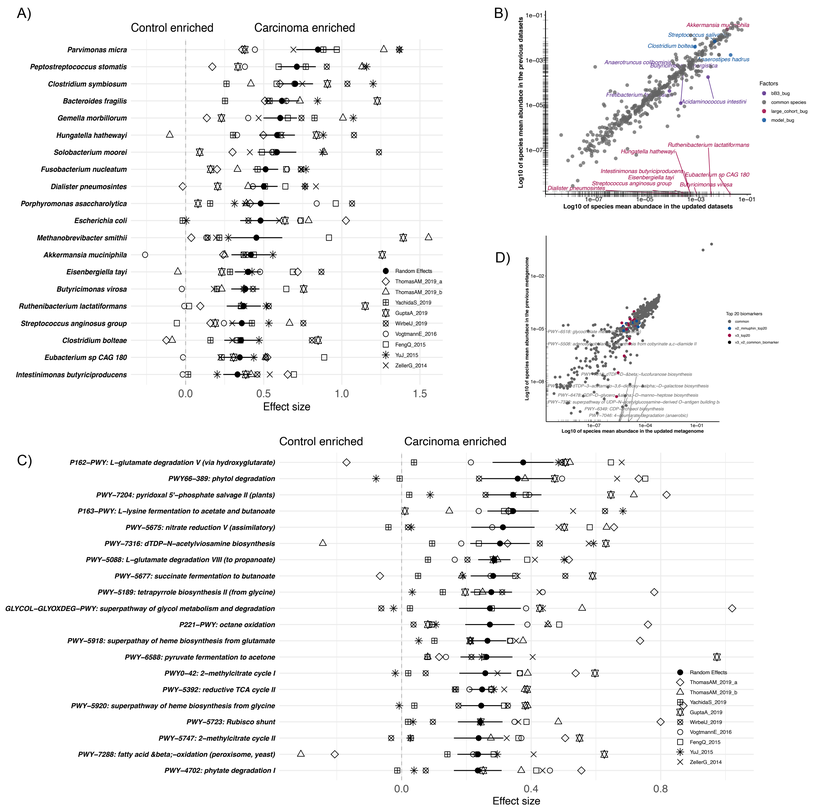


E. Bray-Curtis dissimilarity after dataset adjustment

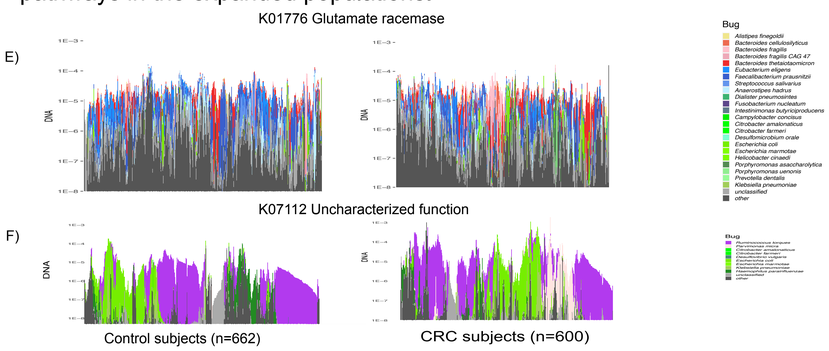


Metagenomics of the stool microbiome in CRC populations. a) size and characteristics of the large scale CRC metagenomic datasets. b) Performing batch (study) effect adjustment in CRC microbial features. c) Principal corrdinate analysis (PCoA) of stool metagenomic species. d) Typical of western populations, gradients of Bacteriodetes and Firmicutes dominance are seen across populations.

Known and novel taxa and pathway enriched in CRC, including typically oral microbes

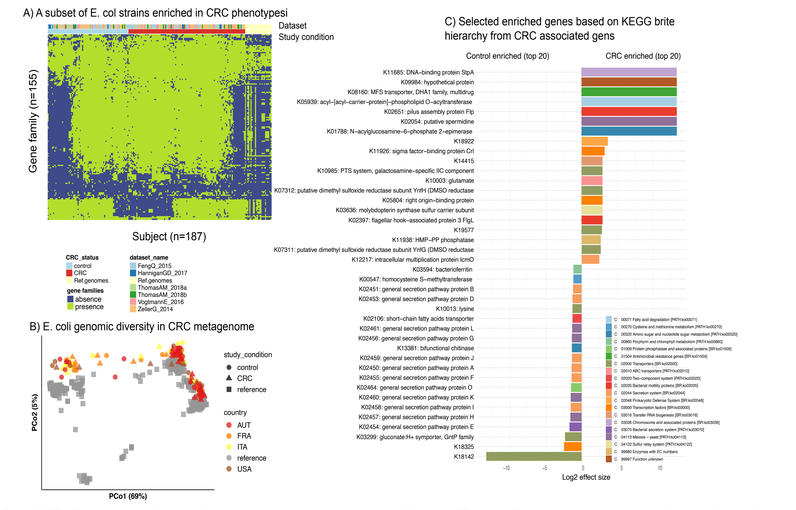


(A) The top 20 biomarker species and functional pathways (B) collectively enriched in CRC subjects, calculated using MMUPHin framework on MetaPhlAn 3.0 species abundance and HUMANn 3.0 pathway abundances. Newly associated (C) biomarker species and (D) pathways in the expanded populations.



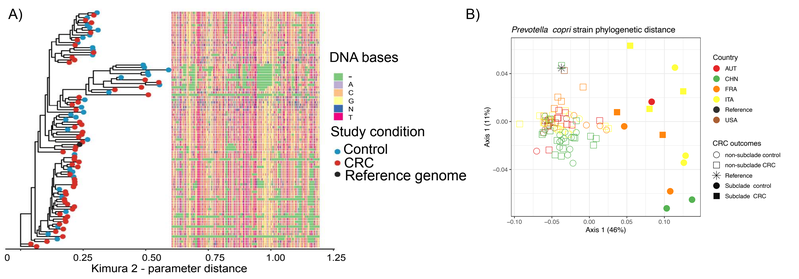
The composition of contributing species represented in metagenomes of Glutamate racemase (E) and unknown function (F) for control and CRC subjects. The relative metagenomic contribution in Glutamate racemase of CRC enriched species including *B. fragilis*, *P. asaccharolytica*, and *F. nucleatum* were different in control and CRC subjects.

Strain-specific *E. coli* gene families associated with CRC



A) CRC associated genes (n=155) within *E. coli* (rows) associated with CRC phenotypes. Each column indicates a strain from CRC subjects (n=105), controls (n=73) and reference genomes (n=24). B) E. coli genomic diversity in the CRC populations as seen by principal coordinates analysis (PCoA) of gene families presence and absence profiles. C) We found a group of functional genes unique to subsets of E. coli strains associated with CRC phenotypes, comprising annotations to transporters, type II secretion system, flagellar and sulfur metabolism.

CRC cases are depleted in geographically-specific *P. copri* subtype carriers



A) The clade with the greatest phylogenetic contribution to CRC was *P. copri* (3.8% of variation, FDR q<0.01), and since the species' subclades are also partialy geographically associated (B), this may suggest that there are regional contributions by *P. copri* subtypes to carcinogenesis.

Acknowledgements

This work was funded in part by Charles A. King Trust Postdoctoral Research Fellowship Program (YY), by Cancer Research UK Grand ChallengeInitiative C10674/A27140 (WG), and by NIH NIDDK R24DK110499 (CH). Methods used for analysis are available in the bioBakery at:

http://huttenhower.sph.harvard.edu

A Cross -link Between Dietary Sodium, Gut Microbiome, and Heart Failure

Oliver Duah[1], Haidong Zhu[2], Li Chen[2], Yanbin Dong[2], Lufei Young[1]

[1] Augusta University, College of Nursing

[2] Augusta University, Medical College of Georgia, Georgia Prevention Institute

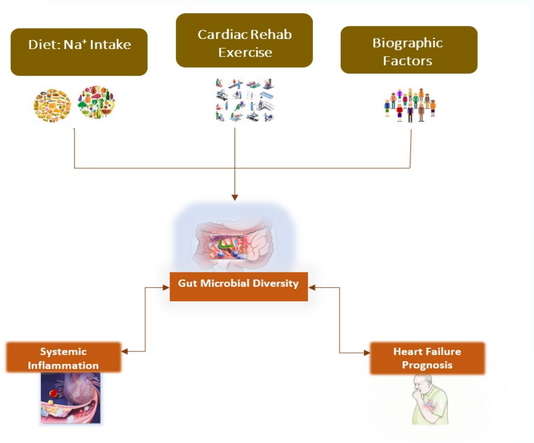
Background & Significance

- Inflammation plays a central role in the development of heart failure (HF).
- Inflammation directly and indirectly modulates the bacterial composition of the human microbiome.
- High sodium intake contributes to progression of HF through inflammation.
- The mechanism through which high sodium diet influences vascular inflammation in HF has not been fully understood.

Purpose

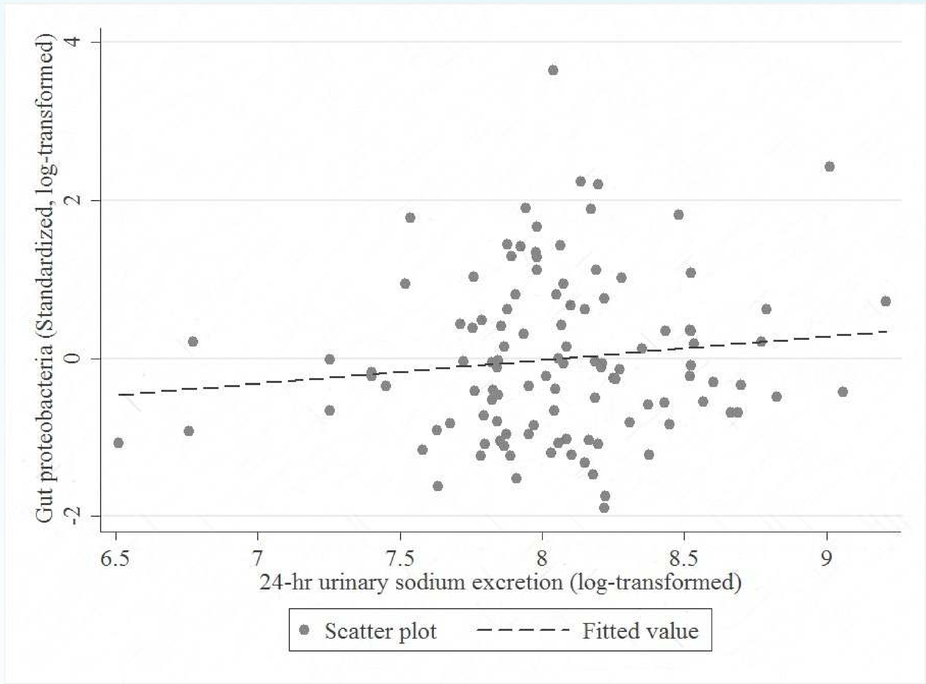
To examine the relationships between sodium intake and gut bacterial composition in a group of Heart Failure patients.

Conceptual Map



Methods and Results

This study was retrospective. Data was collected from 90 HF patients who were the participants of 3 -months cardiac rehabilitation. Daily sodium intake was measured by 24 -hour urine excretion. Fecal samples were collected, and gut microbiota was assessed through 16sRNA sequencing. Gut microbiota data and 24-hour urinary sodium excretion were log -transformed and gut microbiota data were further standardized before statistical analysis. Mixed -effects models were used to assess the association of 24 -hour urinary sodium excretion with gut microbiota with adjustment of age, race, sex, body mass index, group assignment, visit, and diet.



Gut microbiota were measured from 80 HF patients, who were aged 64.2 ± 10.4 years old, 56% male, and 58% Caucasian. The average ejection fraction (EF) was 46.7 ± 15.4 %. We found a significant correlation between sodium intake and the composition of gut proteobacteria ($\beta = 0.57$; $P = 0.026$) among HF patients. HF patients who had higher daily sodium intake had increased proportions of proteobacteria in their gastrointestinal (GI) system.

Discussion

- The significant correlation between sodium intake and the increased proportion of proteobacteria may be explained by
  - the direct impact of high sodium intake on the growth of proteobacteria
  - the indirect effect through the inflammation process.
- The inflammatory cells release reactive nitrogen species which are used by proteobacteria for anaerobic respiration and growth.
- The growth of proteobacteria is a biomarker of systemic inflammation caused by high sodium intake in HF patients..

Conclusions

- The role of gut microbiota in heart failure prognosis may help explain the link between dietary sodium intake and heart failure prognosis.
- Increased proportions of proteobacteria may be a sensitive indicator of worsening HF caused by high sodium intake.
- Additional research is needed to support our finding.

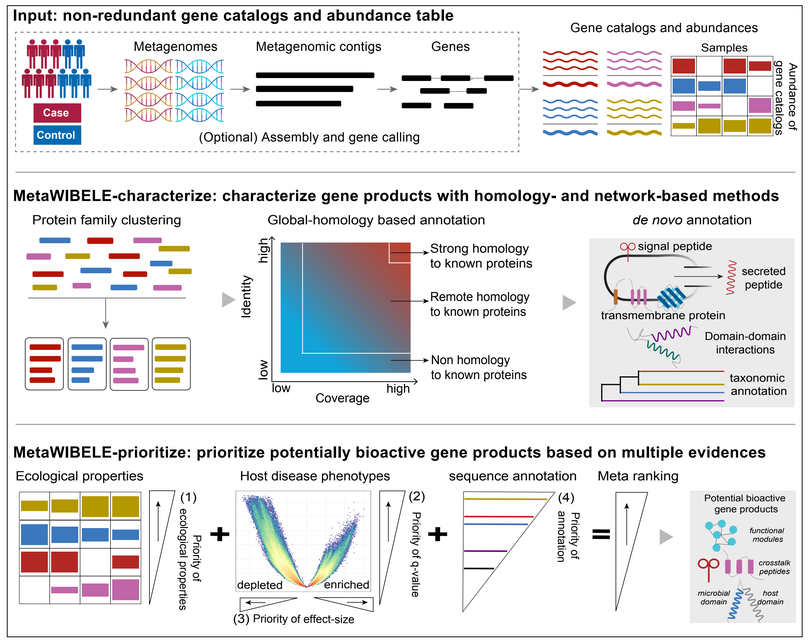
Identifying novel bioactive microbial gene products in inflammatory bowel disease

Yancong Zhang[1,2,3], Amrisha Bhosle[1,2,3], Sena Bae[2,3], Lauren J. McIver[2,3], Emma Accorsi[2,3], Kelsey N. Thompson[1,2,3], Cesar Arze[2], Ya Wang[1,2,3], Ayshwarya Subramanian[1,2], Damian R. Plichita[1], Ali Rahnavard[1,2], Afrah Shafquat[1,2], Ramnik J. Xavier[1,4], Hera Vlamakis[1], Wendy S. Garrett[1,2,3], Andy Krueger[5], Curtis Huttenhower[1,2,3*], Eric A. Franzosa[1,2,3*]

[1]Broad Institute of Harvard and MIT, [2]Harvard T. H. Chan School of Public Health, [3]Harvard Chan Microbiome in Public Health Center
[4]Massachusetts General Hospital, [5]Takeda Pharmaceutical Company Ltd

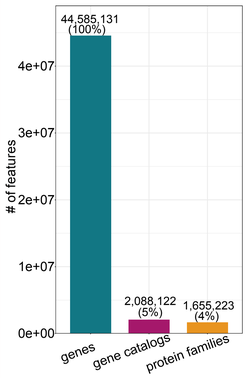
The gut microbiome and associated bioactive compounds are often disrupted in gastrointestinal conditions such as the inflammatory bowel diseases (IBD). Since more than one-third of all proteins in the gut microbiome are uncharacterized, we prioritized potentially bioactive novel proteins from the metagenomes in the Integrative Human Microbiome Project (HMP2). Remarkably, >340,000 protein families are specifically prioritized as potentially bioactive by integrating criteria based on ecological properties and host disease phenotypes. Strikingly, ~23% of them were novel proteins, 36% of which expanded the pangenomes of common gut taxa and >90% of the remainder were assigned at least one putative biochemical annotation. Our analysis methods are generalizable to other microbial communities and human disease phenotypes, and we provide an open source implementation as MetaWIBELE (**W**orkflow to **I**dentify novel **B**ioactive **Ele**ments in the microbiome). The prioritized results provide thousands of new microbial genes likely to interact with host immunity in IBD and gut inflammation, expanding our understanding of bioactive gene products in chronic disease states.

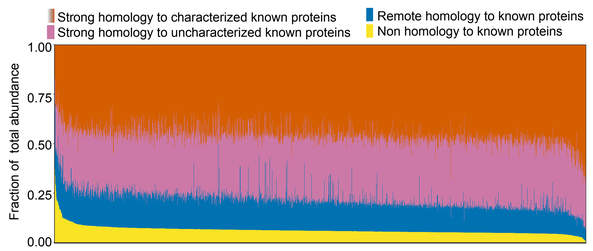
Computational prioritization workflow



MetaWIBELE: assembly-based workflow to prioritize potentially bioactive microbial gene products
- identify characterized and uncharacterized protein families
- assign functional and taxonomic annotations to protein families
- prioritize protein families by combining evidences from abundance-based and sequence-based annotations

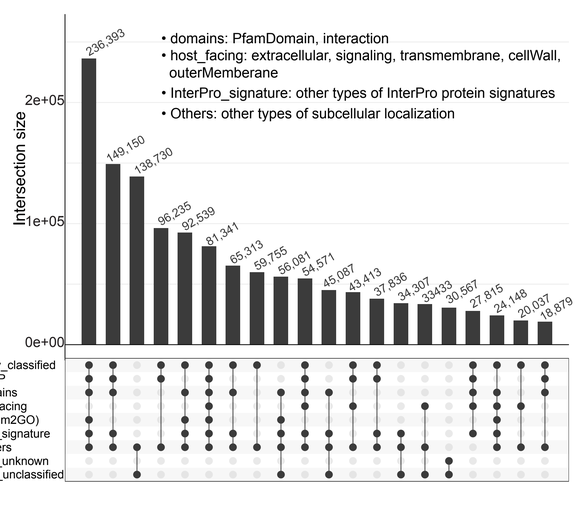
Many protein families are uncharacterized, but can be assigned new annotations





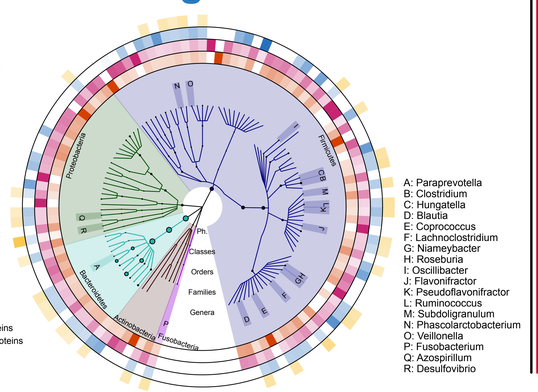
~70% of protein families are uncharacterized
- a catalog of 1.6M protein families was assembled from 1,595 metagenomes in HMP2
- uncharacterized proteins were identified with global homolog search

Uncharacterized proteins are assigned new annotations
- biocheminal annotations were predicted by MetaWIBELE
- only ~3% were still functionally and taxonomically unknown

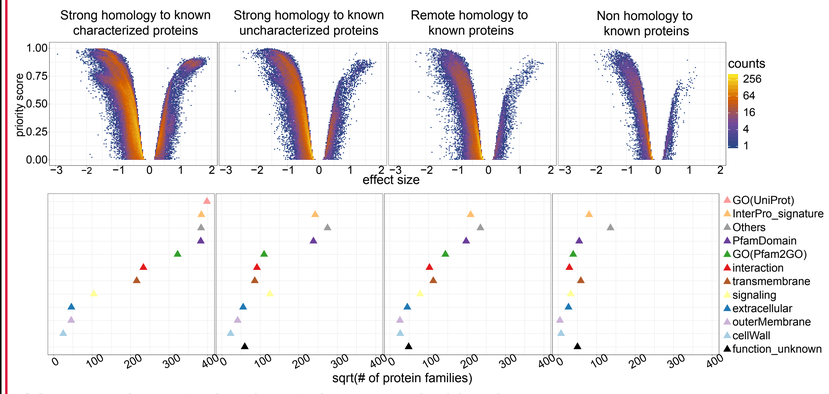


Uncharacterized proteins expanded the pangenomes of common gut taxa

Uncharacterized proteins are classified into phylogenetic clades
- uncharacterized proteins covered common gut taxa
- some novel proteins dominated in some clades

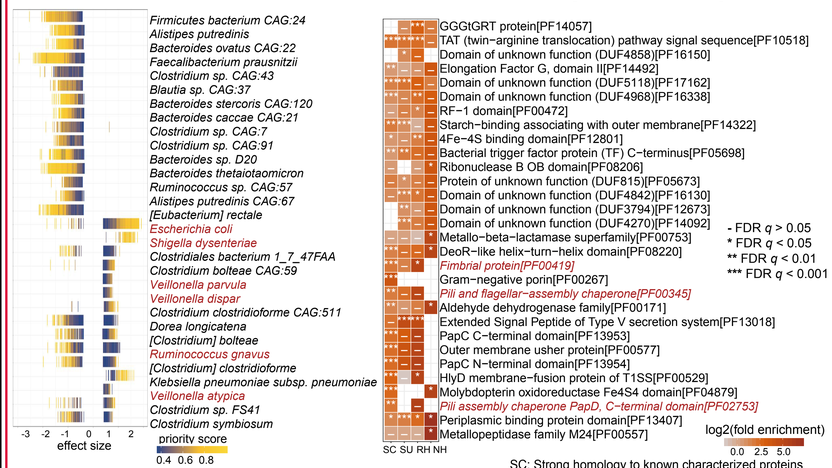


Uncharacterized proteins implicated in bioactivity are prioritized



Many uncharacterized proteins are prioritized
- 90% of prioritized proteins were depleted in disease-active state
- half of prioritized proteins were uncharacterized with 23% novel proteins
  - 39% of them expanded the pangenomes of common gut taxa
  - 90% of the remainder were assigned at least one putative annotation

Taxonomic and functional enrichment of prioritized proteins



Highly prioritized proteins are potentially bioactive related to gut inflammation during IBD
- prioritized protein families showed taxonomic agreement with reported IBD-associated species
- enriched pilin proteins could interact with the human epithelial surface

SC: Strong homology to known characterized proteins
SU: Strong homology to known uncharacterized proteins
RH: Remote homology to known proteins
NH: Non homology to known proteins

Acknowledgements

This work has been supported in part by a grant from Takeda Pharmaceuticals (CH) and NIH NIDDK grant R24DK110499 (CH).

http://huttenhower.sph.harvard.edu
https://huttenhower.sph.harvard.edu/metawibele