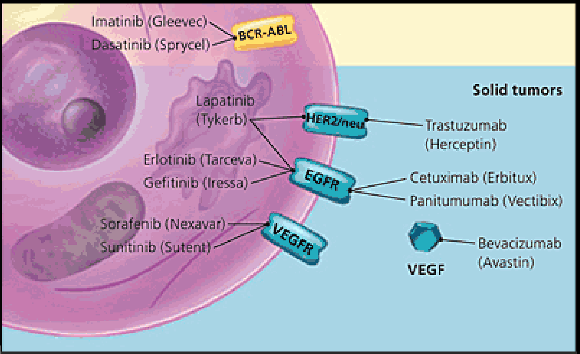
Functional analysis of human microbiome metagenomes, metatranscriptomes, and multi'omics

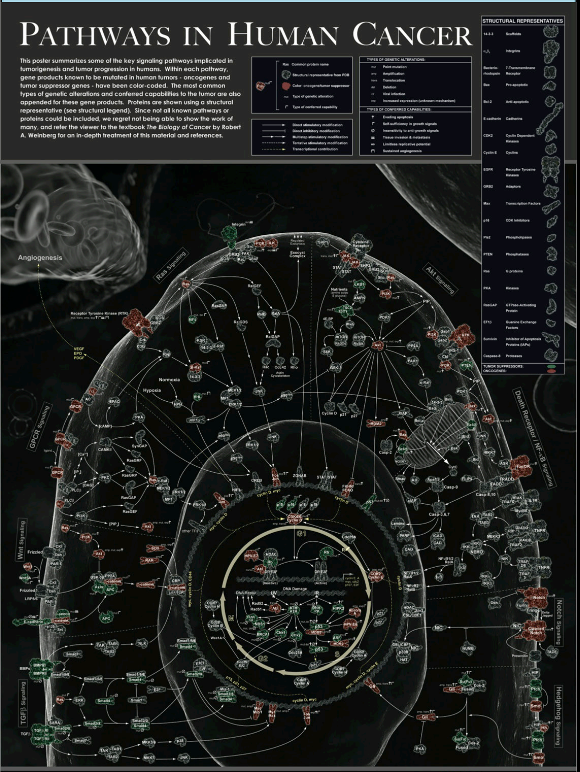
Curtis Huttenhower

Harvard School of Public Health
Department of Biostatistics

07-24-13

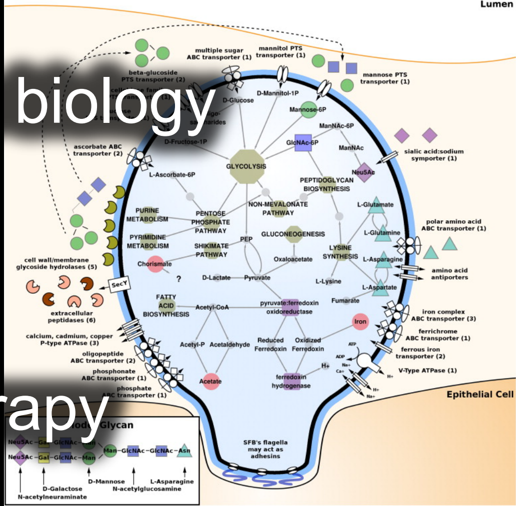
Linking who, what, and how in the human microbiome



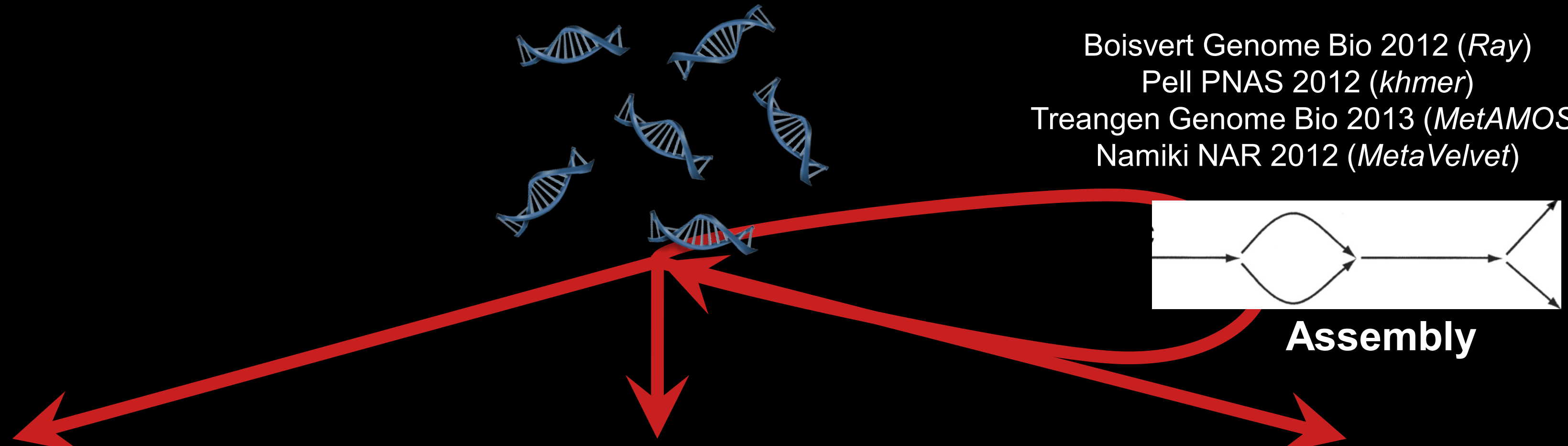


What are the <u>biomolecular networks</u> driving <u>emergent phenotypes</u> in the microbiome and their <u>influences on human health</u>?

- Links translation to basic biology
- Identifies actionable molecular targets for therapy



What does shotgun meta'omics tell us about the human microbiome?



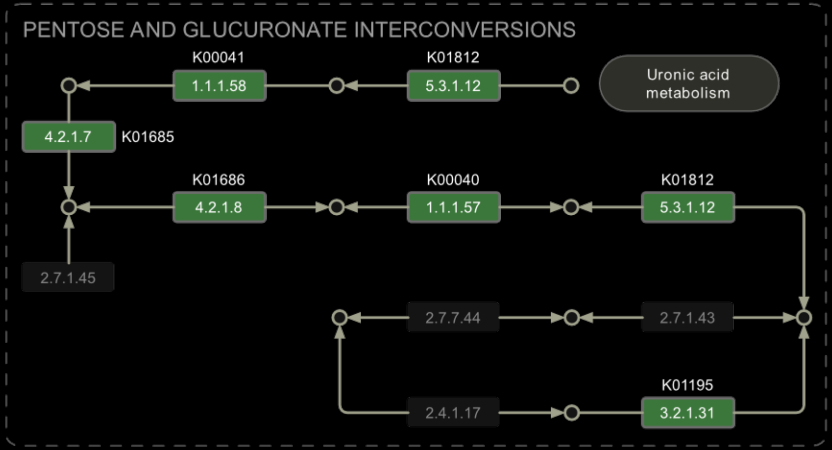
Boisvert Genome Bio 2012 (*Ray*)
Pell PNAS 2012 (*khmer*)
Treangen Genome Bio 2013 (*MetAMOS*)
Namiki NAR 2012 (*MetaVelvet*)

Assembly

Taxonomy/phylogeny

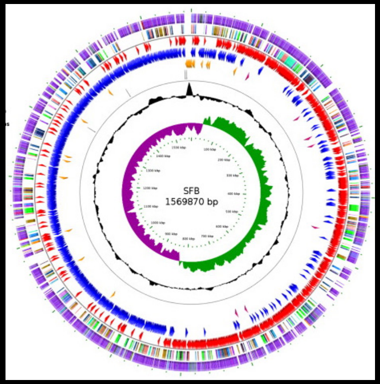
Brady NatMet 2011 (*PhymmBL*)
Patil PLoS ONE 2012 (*PhyloPythia*)
Wu Bioinf 2012 (*AMPHORA*)
Segata NatMet 2012 (*MetaPhlAn*)

Functional profiling



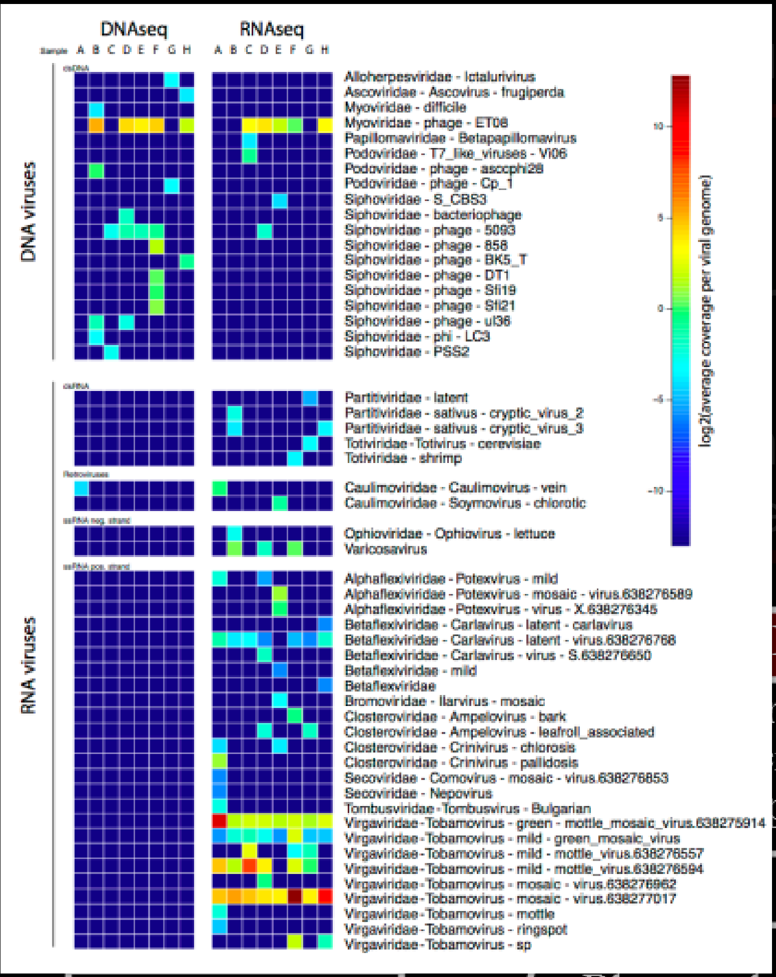
Abubucker PLoS CB 2012 (*HUMAnN*)
Meyer BMC Bioinf 2008 (*MG-RAST*)
Markowitz NAR 2011 (*IMG/M*)
Konwar BMC Bioinf 2013 (*MetaPathways*)

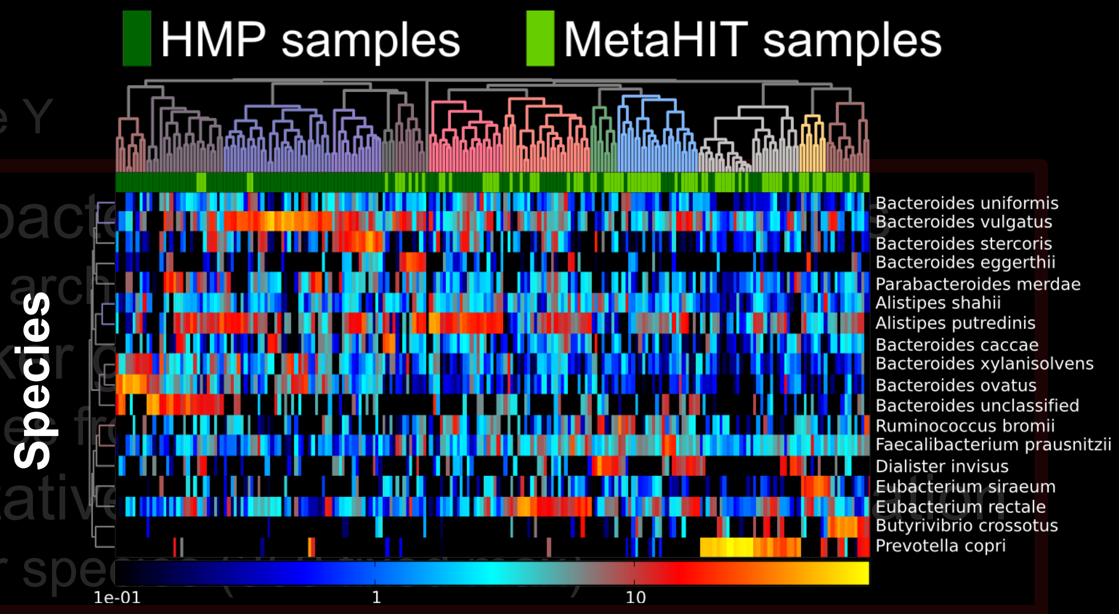
Comparative genomics



Schloissnig Nature 2013
Hehemann Nature 2010
Stern Genome Res 2012
Rho PLoS Gen 2012

Meta'omic taxonomic profiling with **MetaPhlAn**: leveraging 1,000s of microbial genomes





PhlAn (offline pipeline)

re **genes** for all clades

nes for unique **marker genes**

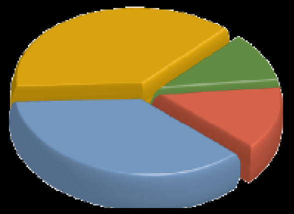
presentative marker genes

Available reference genomes

Metagenome

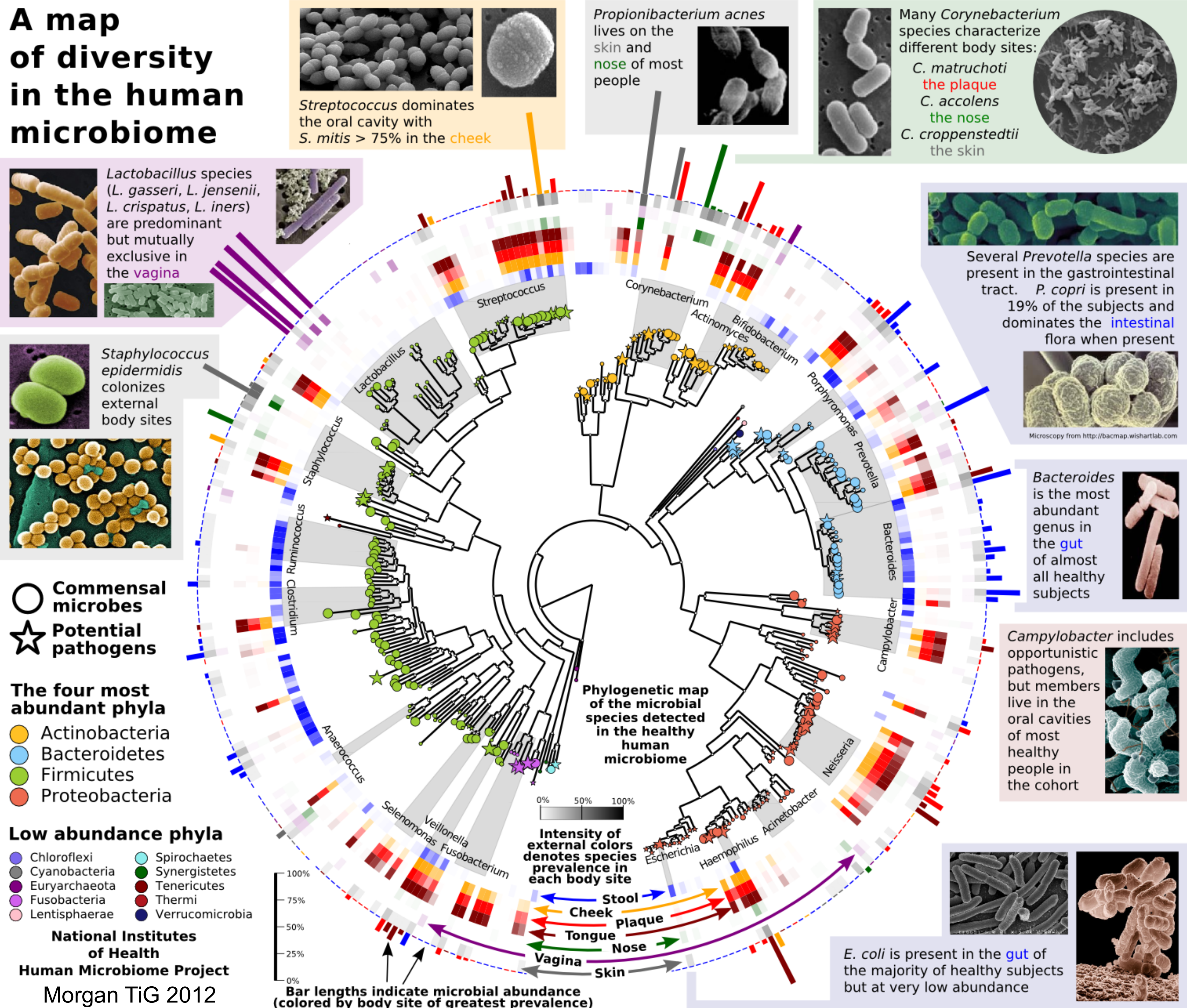
MetaPhlAn

- Blast reads agains the marker genes
- Assign, count, normalize reads



Segata NatMet 2012 (*MetaPhlAn*)

A map of diversity in the human microbiome



Streptococcus dominates the oral cavity with *S. mitis* > 75% in the cheek

Propionibacterium acnes lives on the skin and nose of most people

Many *Corynebacterium* species characterize different body sites:

C. matruchoti
the plaque
C. accolens
the nose
C. croppenstedtii
the skin

Lactobacillus species (*L. gasseri, L. jensenii, L. crispatus, L. iners*) are predominant but mutually exclusive in the vagina

Staphylococcus epidermidis colonizes external body sites

Several *Prevotella* species are present in the gastrointestinal tract. *P. copri* is present in 19% of the subjects and dominates the intestinal flora when present

Bacteroides is the most abundant genus in the gut of almost all healthy subjects

Campylobacter includes opportunistic pathogens, but members live in the oral cavities of most healthy people in the cohort

E. coli is present in the gut of the majority of healthy subjects but at very low abundance

Phylogenetic map of the microbial species detected in the healthy human microbiome

Intensity of external colors denotes species prevalence in each body site

Bar lengths indicate microbial abundance (colored by body site of greatest prevalence)

○ **Commensal microbes**
☆ **Potential pathogens**

The four most abundant phyla

- ● Actinobacteria
- ● Bacteroidetes
- ● Firmicutes
- ● Proteobacteria

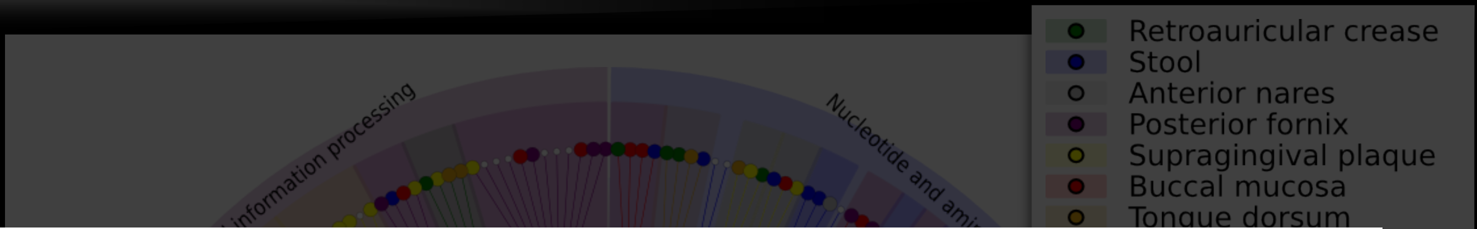
Low abundance phyla

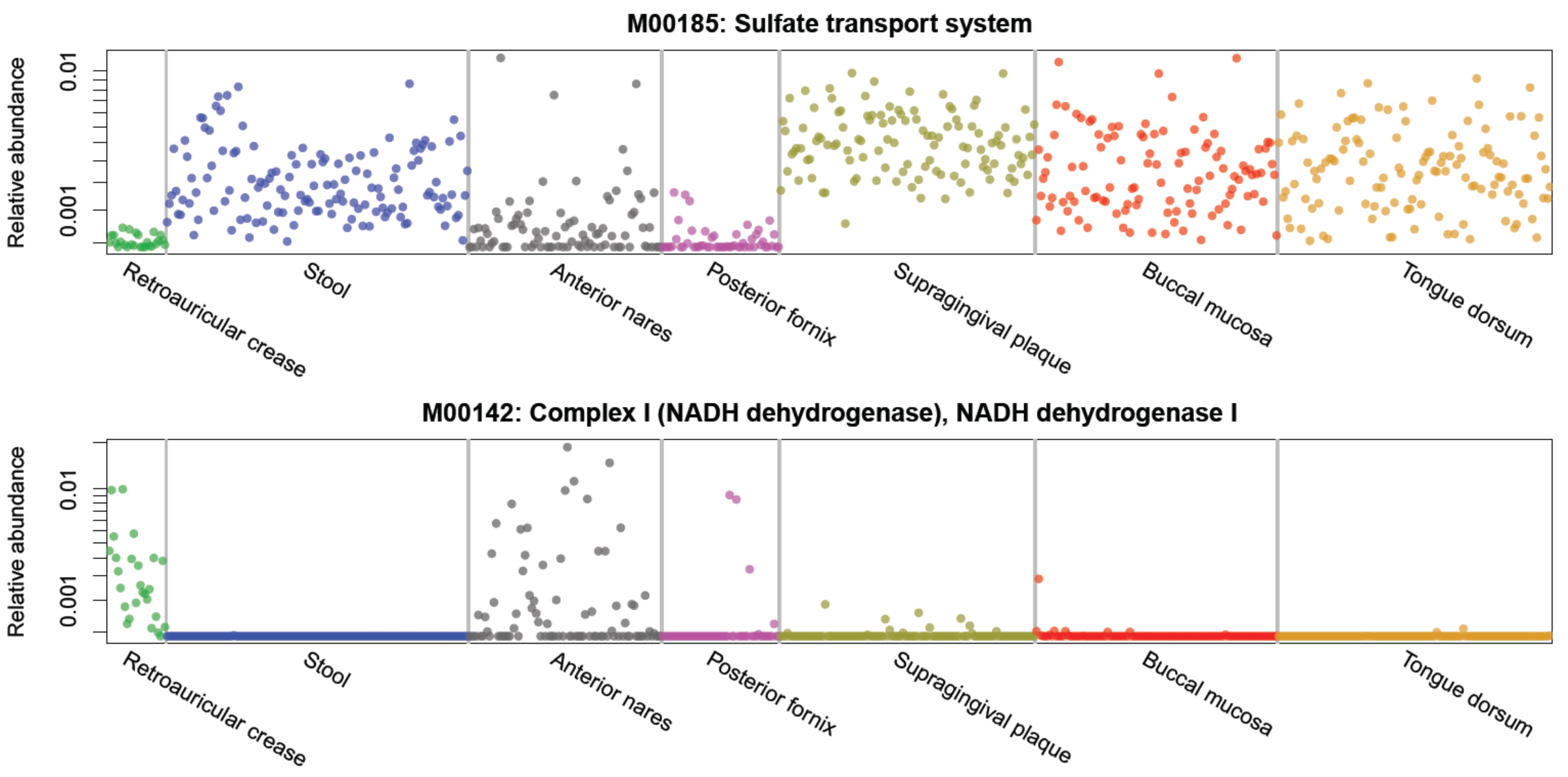
● Chloroflexi	● Spirochaetes
● Cyanobacteria	● Synergistetes
● Euryarchaeota	● Tenericutes
● Fusobacteria	● Thermi
● Lentisphaerae	● Verrucomicrobia

National Institutes of Health Human Microbiome Project

Morgan TiG 2012

Meta'omic functional profiling with **ShortBRED:** the Short Better REad Database creator



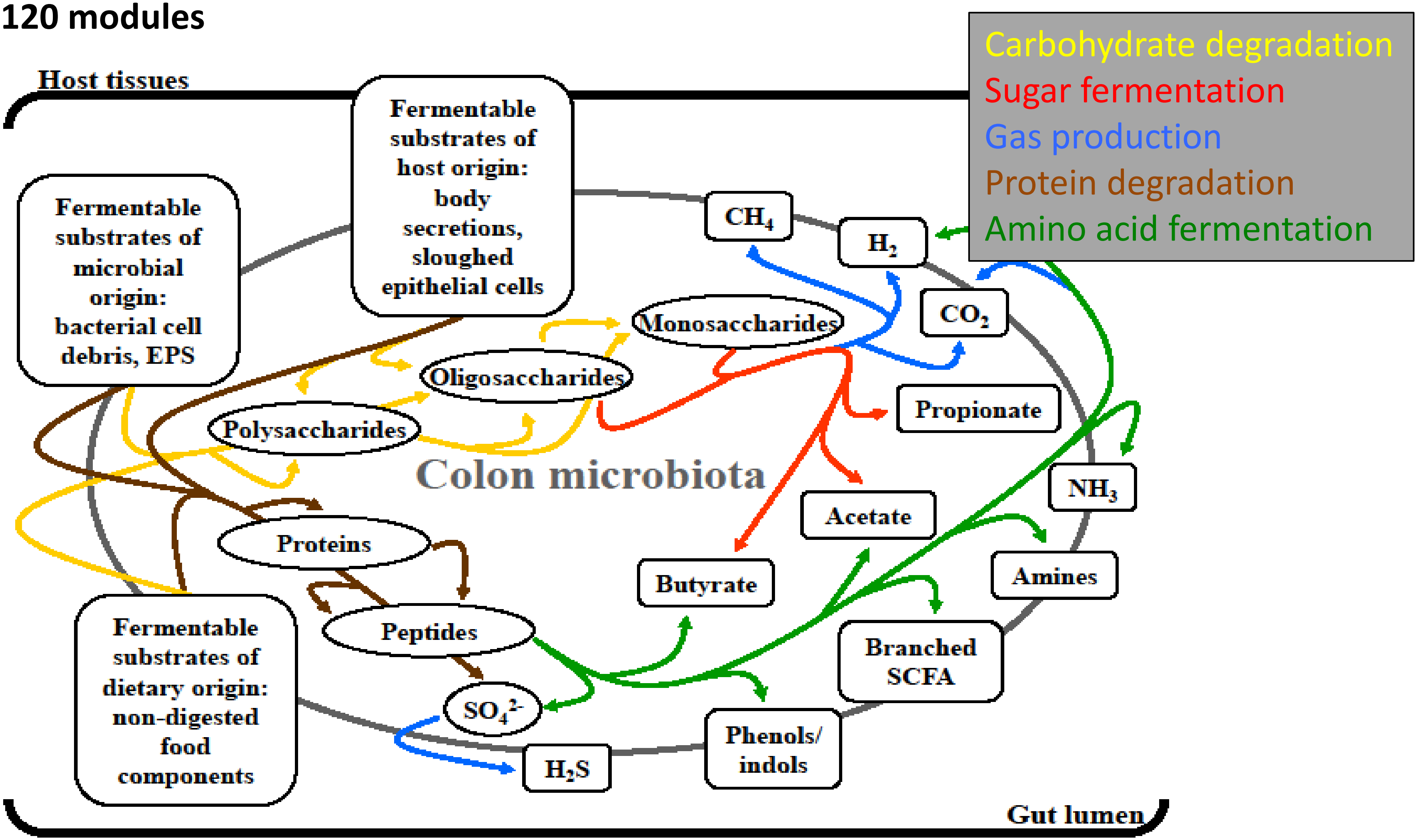


- Most *processes are "core"*: <10% are *differentially present/absent* even by body site
  - Contrast *zero* microbes meeting this threshold!
- Most *processes are habitat-adapted*: >66% are *differentially abundant* by body site

Reconstructing the gut metabolism from metagenomics data

Manually compiled set of pathway modules that captures microbial ‘food chain’

120 modules



Raes lab, VUB-VIB-KULeuven

Meta'omic ecological profiling with **CCREPE**: identifying co-occurring microbial consortia

With Jeroen Raes, Karoline Faust

Co- **Anti-**

Anterior nares

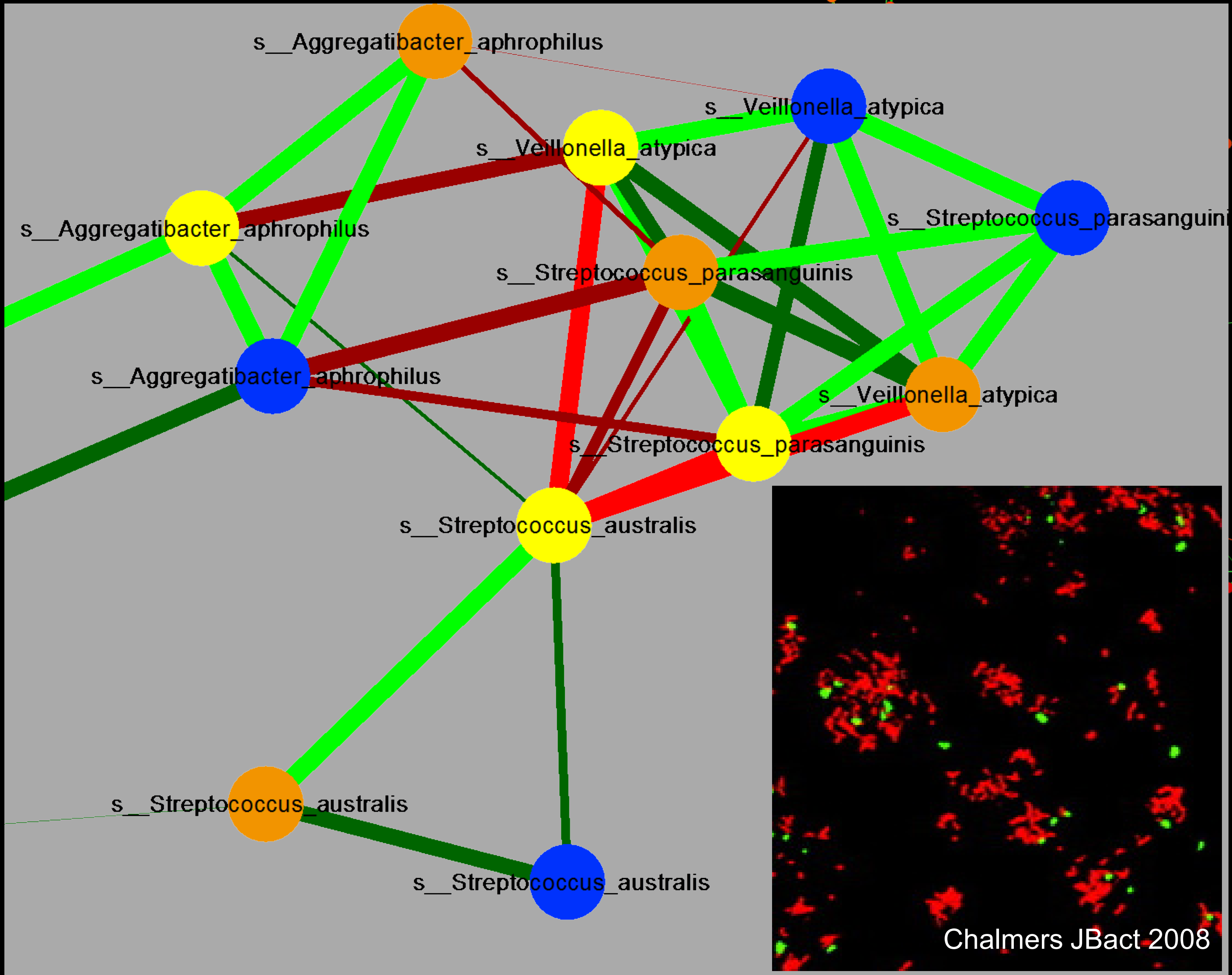
Buccal mucosa
Hard palate
Keratinized gingiva
Palatine tonsils
Saliva
Subgingival plaque
Supragingival plaque
Throat
Tongue dorsum

Left retroauricular crease
Right retroauricular crease

Right antecubital fossa
Left antecubital fossa

Stool

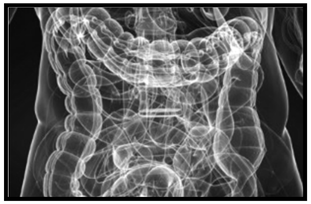
Mid vagina
Posterior fornix
Vaginal introitus



Chalmers JBact 2008

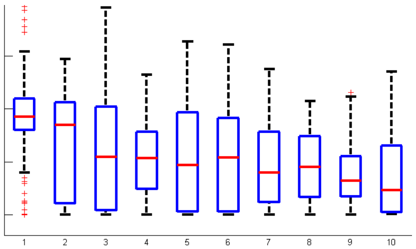
Faust PLoS CB 2012

Comparing Co-Occurrence and Predicted Interactions in the Gut Microbiome

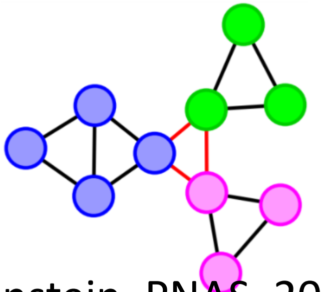


Metagenomics
(124 hosts)

Species abundance

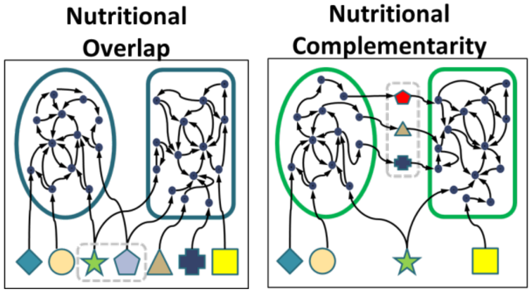


Species co-occurrence

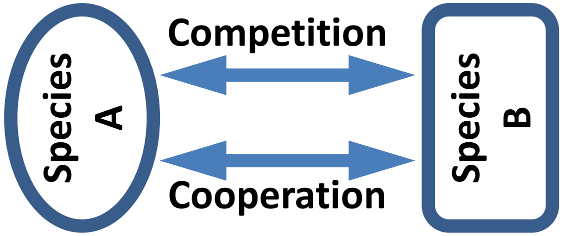


Full genomes
(154 species)

Reverse ecology

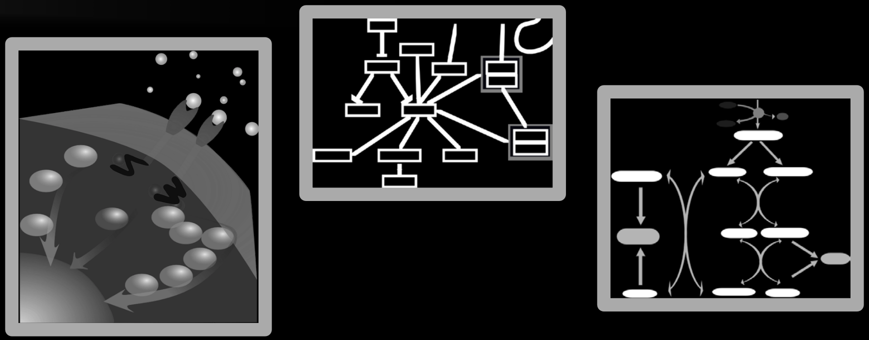


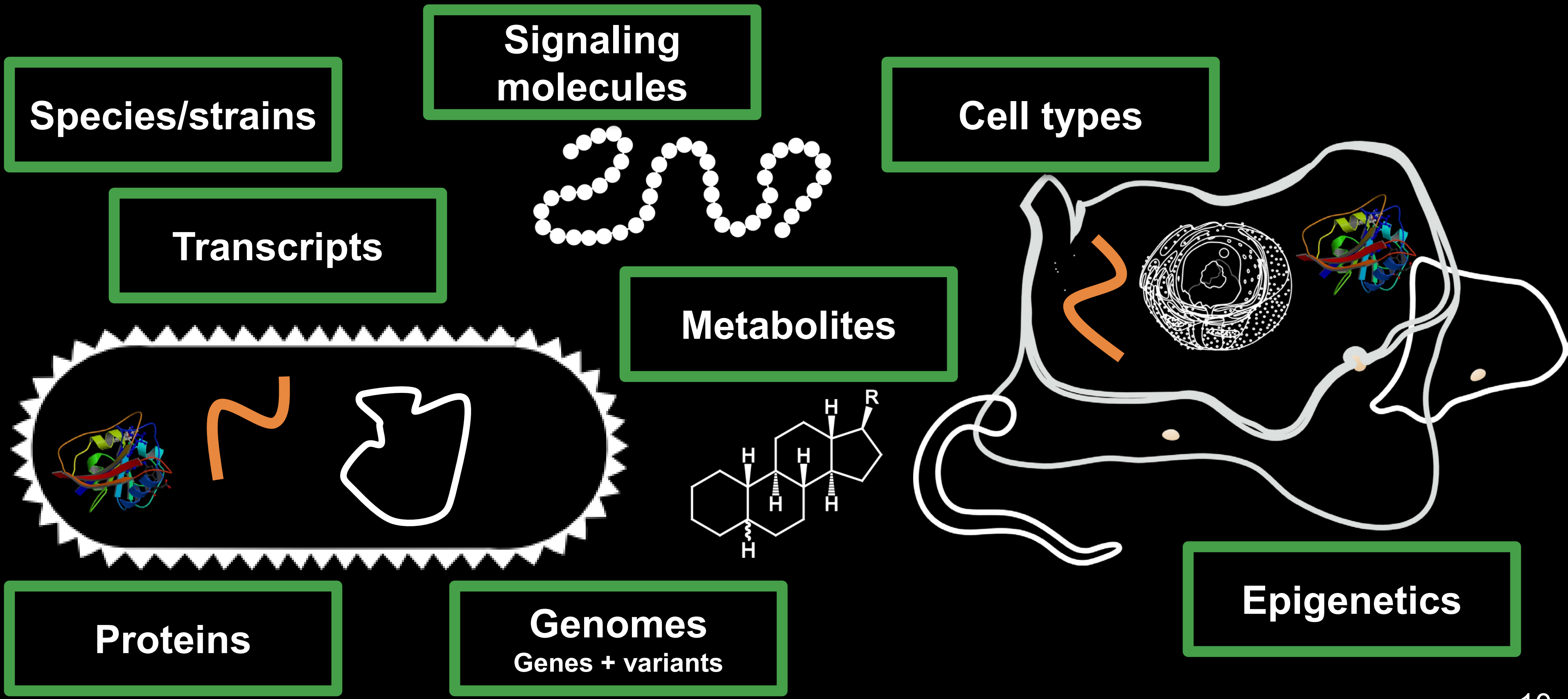
Predicted Interactions



Levy and Borenstein, PNAS, 2013

Multi'omic data integration is necessary to understand biomolecular function in the microbiome





Species/strains

Signaling molecules

Cell types

Transcripts

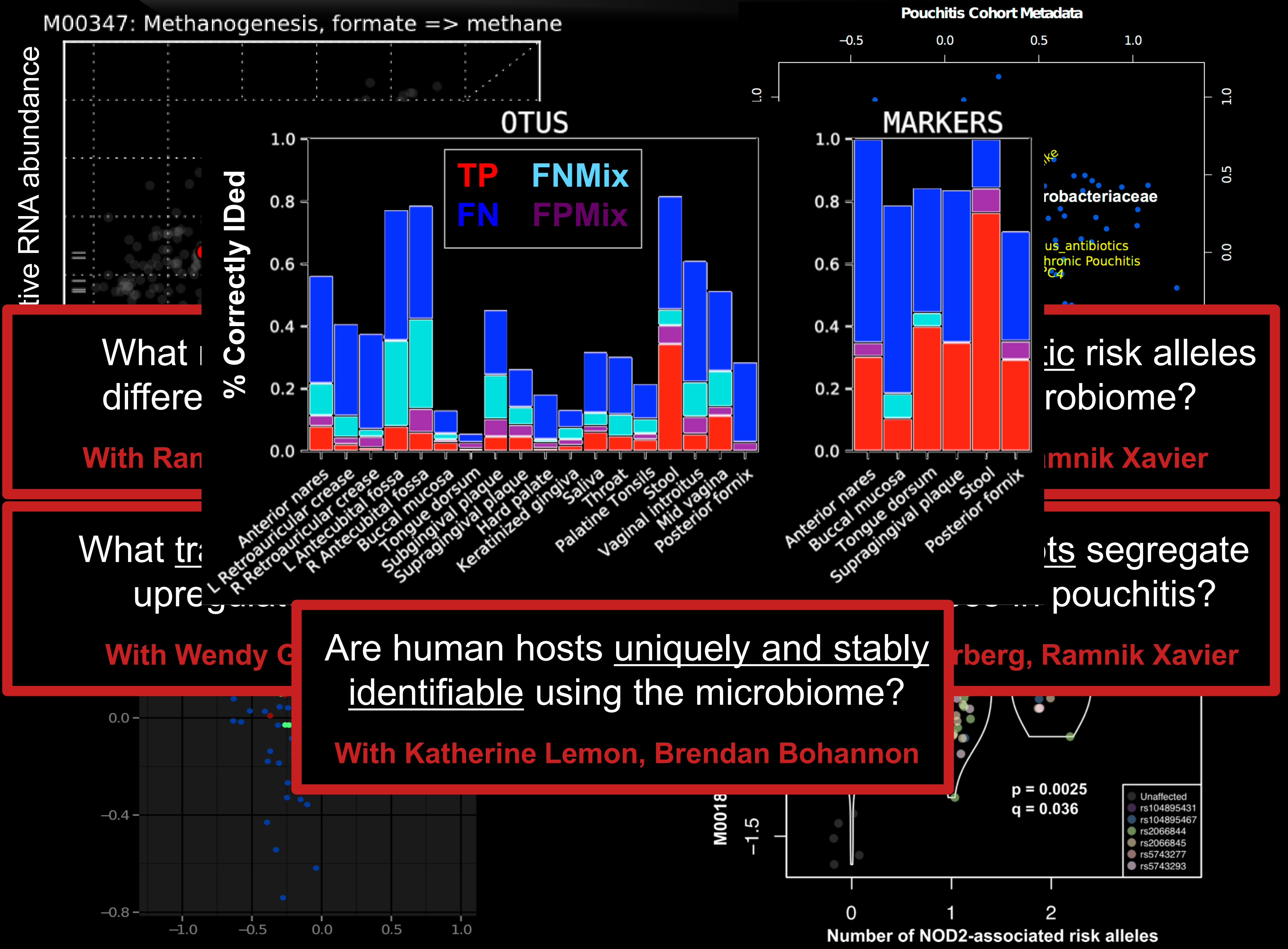
Metabolites

Proteins

Genomes
Genes + variants

Epigenetics

Multi'omic data integration is necessary to understand biomolecular function in the microbiome



What r... differe...

With Ran...

...tic risk alleles ...robiome?

...mnik Xavier

What tr... upre...

With Wendy G...

...ts segregate ... pouchitis?

...rberg, Ramnik Xavier

Are human hosts <u>uniquely and stably identifiable</u> using the microbiome?

With Katherine Lemon, Brendan Bohannon

The microbiome in IBD: a group of complex microbial diseases

- The gut microbiota varies in IBD
  - Diversity is almost certainly reduced
    (Manichanh 2006, Ott 2006, Frank 2007, Sokol 2008, Nishikawa 2009, Willing 2010; contrast Lepage 2009)
  - Specific clades are often over/under enriched
    (Baumgart 2007, Frank 2007, Willing 2010, Joossens 2011, Frank 2011, Lepage 2011)
  - IBD subsets – colitis, ileal CD, etc. – are differentially affected
    (Sokol 2008, Willing 2010, Joosens 2011, Lepage 2011)

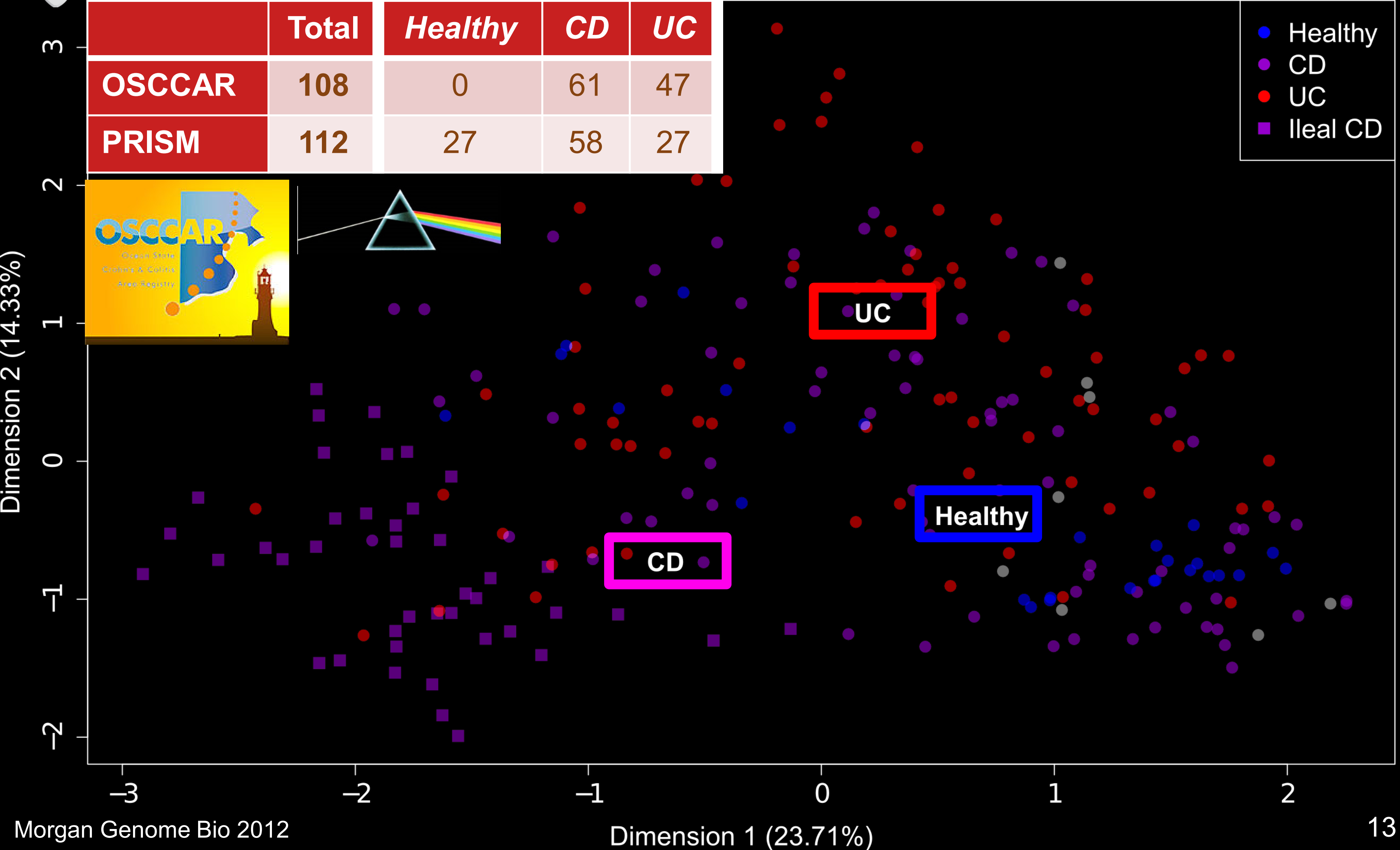
Like disease alleles,
infectious disease ←→ one microbe,
complex disease ←→ many microbes

- Which structural changes might be functional?
  - (If any)
  - (In each subset)
  - And which are instead associated with treatment/environment?
- And <u>why</u>: which specific microbial functions are involved in these changes?

How is the gut microbiome disrupted during IBD and its treatment?

With Ramnik Xavier, Bruce Sands

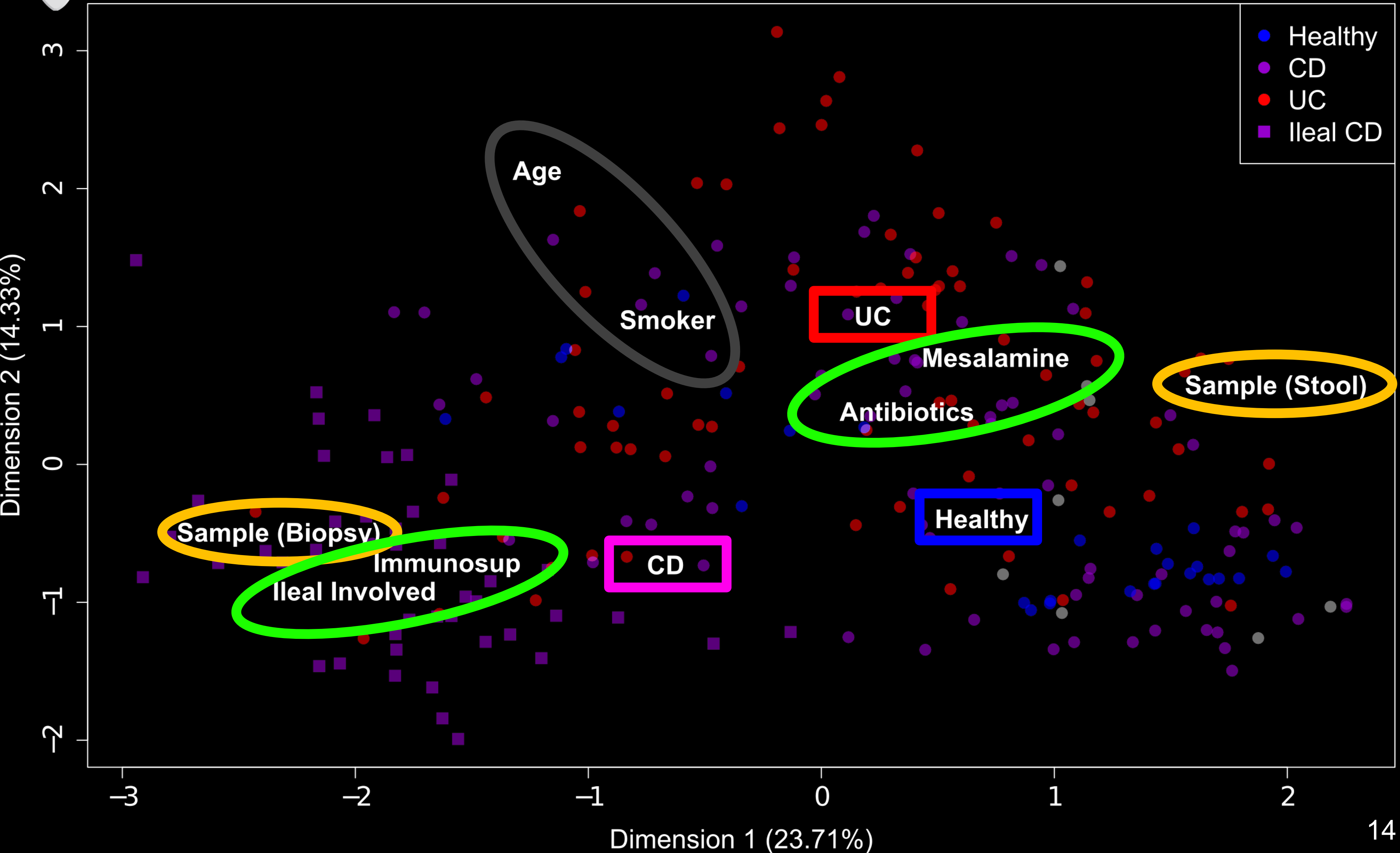
	Total	*Healthy*	*CD*	*UC*
OSCCAR	**108**	0	61	47
PRISM	**112**	27	58	27



Morgan Genome Bio 2012

How is the gut microbiome disrupted during IBD and its treatment?

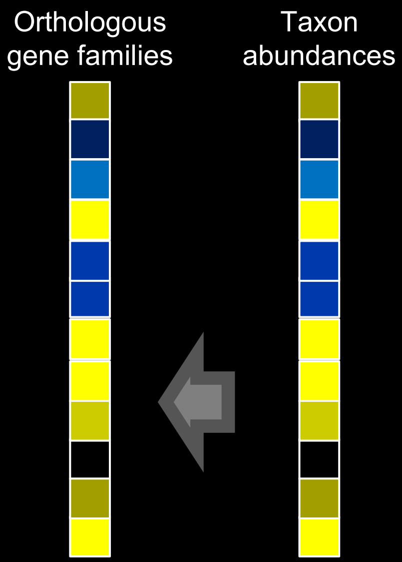
With Ramnik Xavier, Bruce Sands

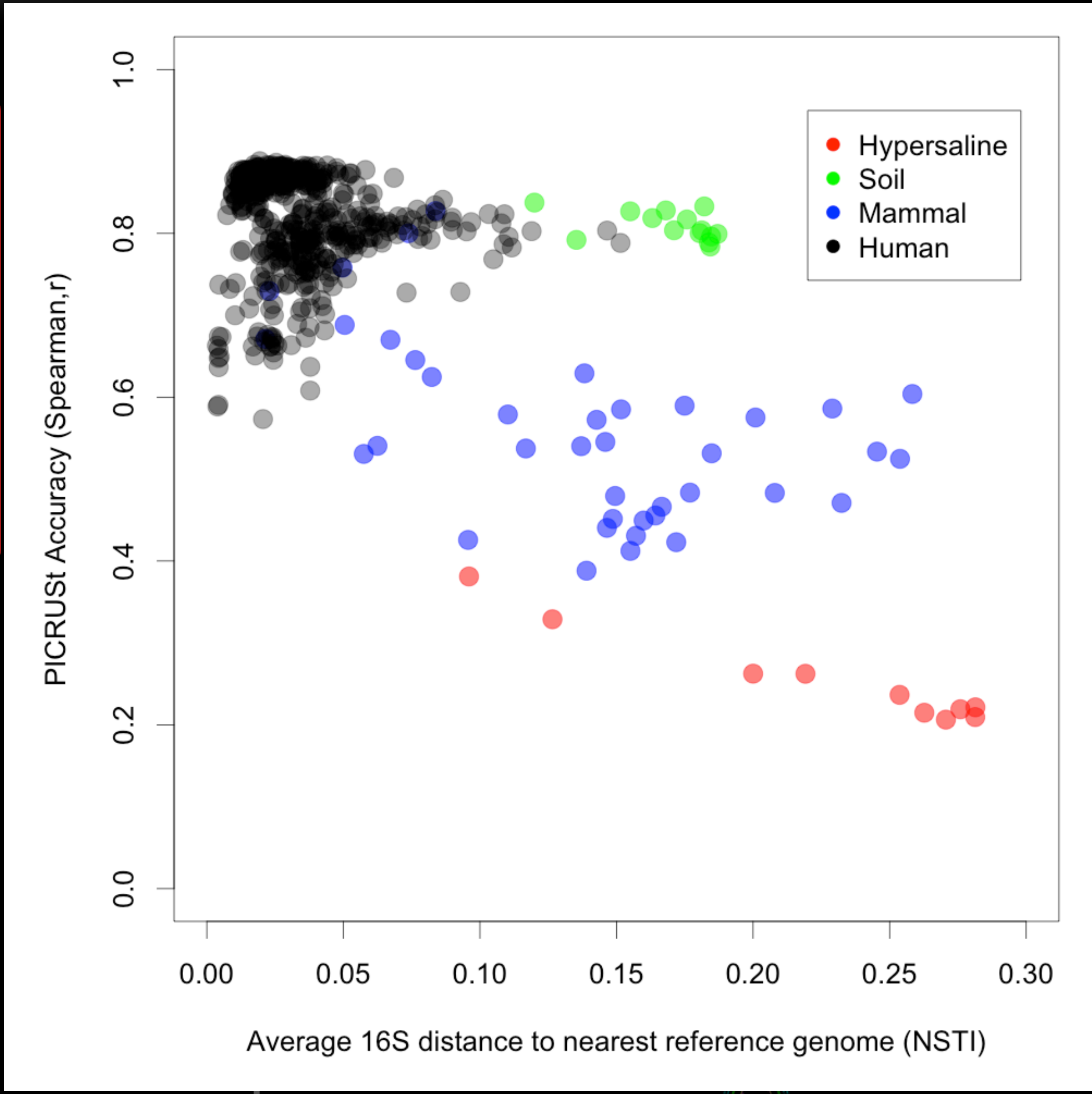


PICRUSt: Inferring community metagenomic potential from marker gene sequencing

With Rob Knight, Rob Beiko

One can recover ***general*** community function with reasonable accuracy from 16S profiles.



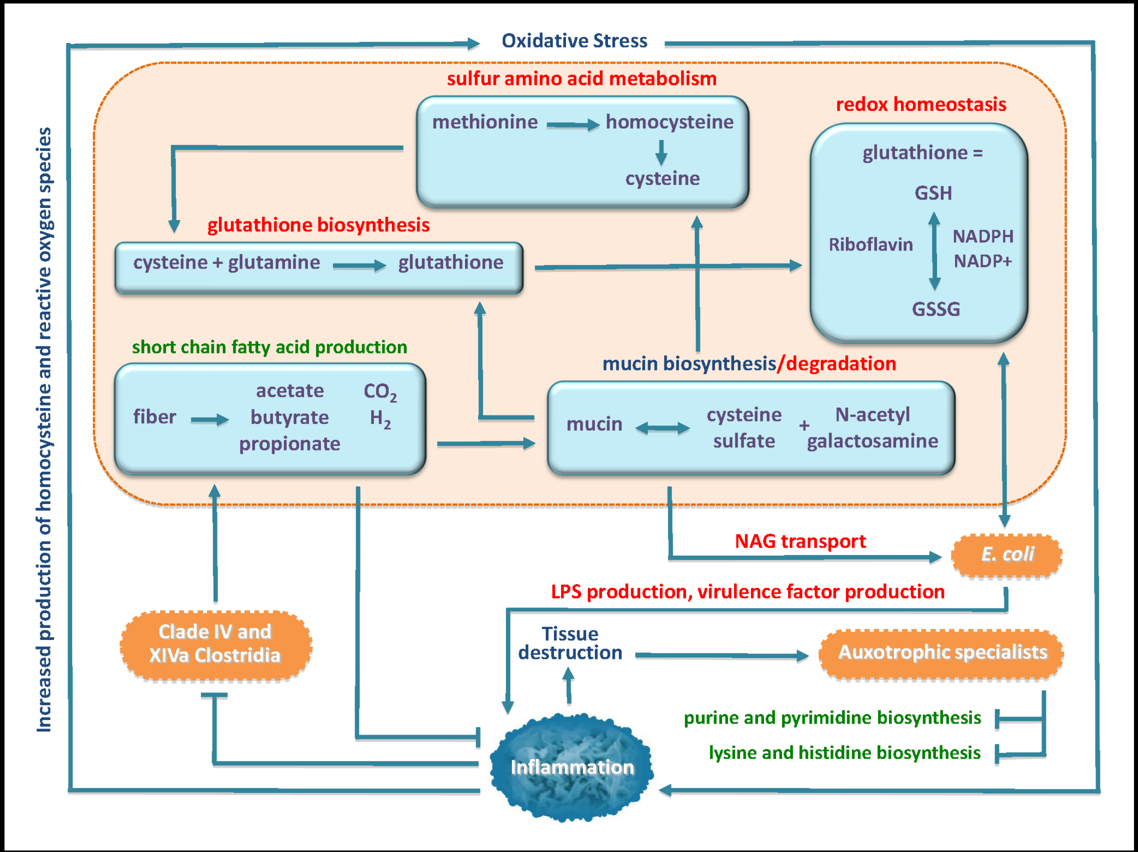


Langille NatBt in press

Relative abundance

Which *functions* of the gut microbiome are disrupted by IBD?

- Over *six times* as many microbial metabolic processes disrupted in IBD as microbes
  - If there’s a transit strike, everyone working for the MBTA is disrupted, not everyone named Smith or Jones



Gaps in knowledge and methods for microbiome functional ‘omics:

- **Tools to make meta’omics as easy as microarray analysis**
  - Web for data organization and acquisition, desktop for visualization and manipulation, cloud for democratized scalability
- **Systematic, cross-species microbial protein function cataloging**
- **Quantitative models of community metabolic and regulatory networks**
- **Exhaustive identification of microbe-microbe and host-microbe interaction mechanisms**
  - Small molecule signals, bioactive metabolites, secreted and cell surface peptides...
- **Detailed, temporally-resolved "microbiogeography”**
- **In vitro models of human-associated microbial communities for controlled gene and microbe “knock out” and “knock in” experiments**
- **Standards for reproducibility of all aspects human microbiome experiments and analysis to ensure translation-quality results**

Thanks!


Nicola Segata

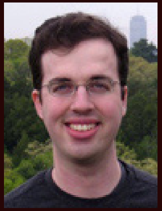

Levi Waldron


Xochi Morgan

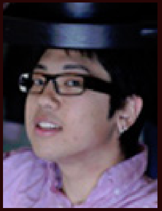

Tim Tickle

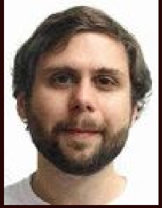

Felix Wong

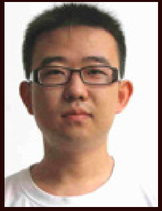

Emma Schwager


Eric Franzosa


Daniela Boernigen


Joseph Moon


Jim Kaminski


Boyu Ren


Brian Palmer


Ren Lu


Koji Yasuda


Dirk Gevers
Kat Huang


Ramnik Xavier
Harry Sokol
Dan Knights
Moran Yassour

Human Microbiome Project

Owen White
Joe Petrosino
George Weinstock
Karen Nelson
Lita Proctor
Erica Sodergren
Anthony Fodor
Marty Blaser
Jacques Ravel
Pat Schloss

Sahar Abubucker
Brandi Cantarel
Alyx Schubert
Mathangi Thiagarajan
Beltran Rodriguez-Mueller
Makedonka Mitreva
Yuzhen Ye
Mihai Pop
Larry Forney
Barbara Methe

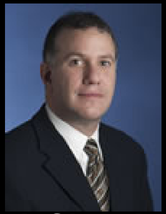
Bruce Birren
Doyle Ward

Mark Daly
Ashlee Earl


Wendy Garrett
Michelle Rooks


Rob Beiko
Morgan Langille


Rob Knight
Jesse Zaneveld
Greg Caporaso


Mark Silverberg
Boyko Kabakchiev
Andrea Tyler

Ruth Ley
Omry Koren

Jacques Izard
Katherine Lemon

Bruce Sands

Jeroen Raes
Karoline Faust

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

Why shotgun metagenomics and metatranscriptomics?

- Currently the simplest and most cost-effective way to assess microbiome biomolecular function
  - ~3-8x per-sample cost of 16S amplicon sequencing
  - *Strain* level identification of microbes
  - Readily accesses bacteria, archaea, viruses, and eukaryotes
  - Exposes not just who's there, but genetic potential, synteny, regulation, and variation
  - Leverages analysis methods from single-organism DNA/RNA-seq

- What's the bad news?
  - ~3-8x per-sample cost of 16S amplicon sequencing
  - Requires samples with greater biomass
  - Sensitive to samples with greater host contamination
  - Can require more complex informatics