Meta-analysis of microbiome studies to identify disease-specific and shared responses

Claire Duvallet, Sean Gibbons, Thomas Gurry, Rafael Irizarry, Eric Alm

center for microbiome informatics & therapeutics

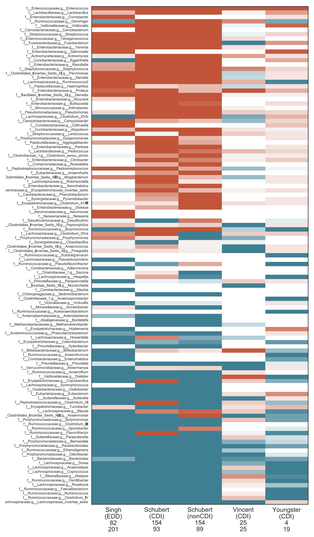
MIT BE BIOLOGICAL ENGINEERING

Many studies report associations between the human gut microbiome and disease.
However, results from individual studies can be inconsistent and are difficult to compare because of a lack of standard methods.

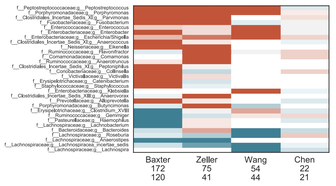
We reprocessed 28 case-control studies from 10 diseases with standard methods, compared results across studies, and developed a method to correct for batch effects. Data is available in **MicrobiomeHD database** on Zenodo.org.

Consistent direction and extent of microbiome shifts can inform treatment strategies

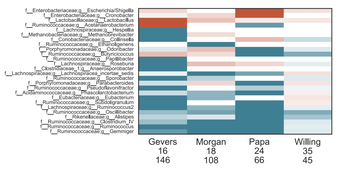
Diarrhea comes with a broad community re-structuring. Fecal transplants restore community.

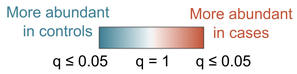


Colorectal cancer is characterized by **enrichment of disease-associated bacteria**: treat with antibiotics or develop diagnostic?

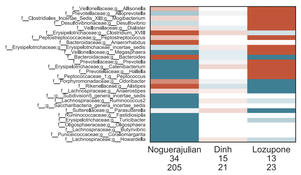


Inflammatory bowel disease is characterized by **depletion of health-associated bacteria**: treat with probiotics?

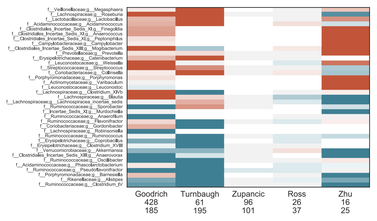




HIV studies are confounded with behavior.



Obesity has few clear biomarkers or patterns of disease-associated shifts.

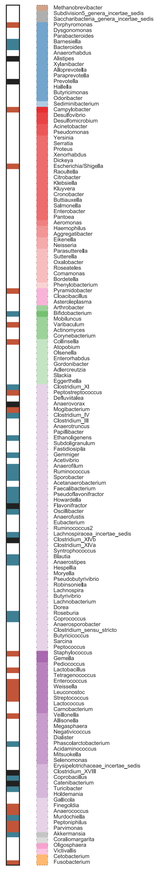


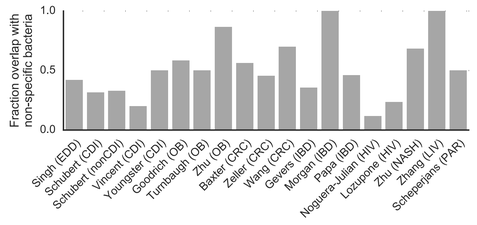
Many bacteria are non-specifically associated with health and disease, suggesting shared response to disease

Non-specific bacteria	Health	Disease
Bacteroidetes	2	1
Firmicutes		
Clostridia	17	6
Bacilli	0	6
Other	3	1
Proteobacteria	0	2
Fusobacteria	0	1
Other	2	2

Lachnospiraceae and *Ruminococcaceae* are known short-chain fatty acid producers.

Lactobacillales are upper-gut bacteria which may indicate shorter stool transit times.

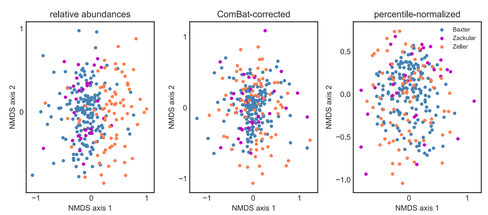




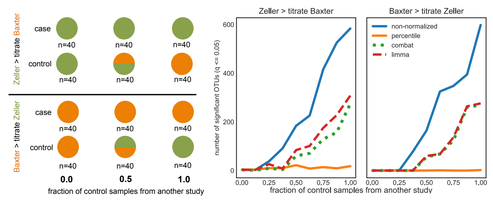
The majority of studies' associations overlap with the non-specific response.

Percentile-normalization corrects for batch effects

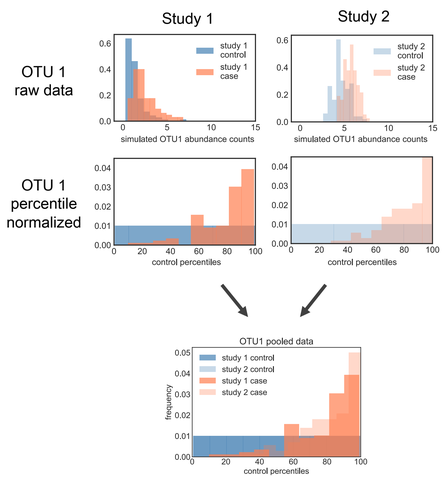
Traditional batch correction methods are not suitable for microbiome data.



Combining raw data across studies leads to spurious hits.



Percentile normalization: OTU abundances in case samples are converted to percentiles of the equivalent OTU in control samples.



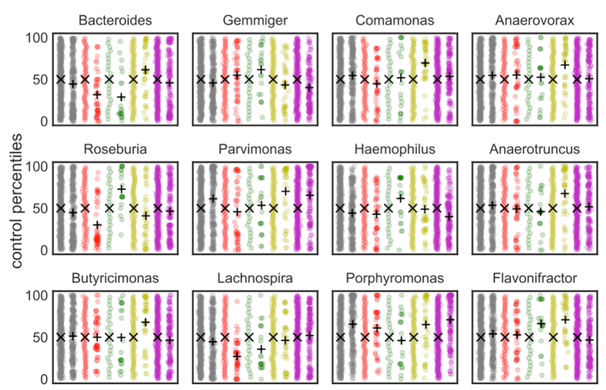
Combining data increases power and reduces false positives

Percentile normalization increases power over other methods.

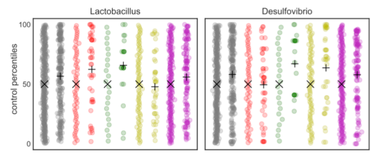
disease	method	pooled	within
CDI (N=2)	percentile	37	U = 38 / 2N = 6
	Fisher	12	
	Stouffer	12	
	ComBat	36	
	limma	36	
CD (N = 4)	percentile	19	U = 13 / I = 0 / 2N = 2
	Fisher	6	
	Stouffer	6	
	ComBat	2	
	limma	1	
UC (N = 3)	percentile	10	U = 17 / I = 0 / 2N = 1
	Fisher	4	
	Stouffer	4	
	ComBat	5	
	limma	5	
CRC (N = 4)	percentile	12	U = 20 / I = 0 / 2N = 3
	Fisher	9	
	Stouffer	7	
	ComBat	5	
	limma	5	
OB (N = 11)	percentile	18	U = 36 / I = 0 / 2N = 6
	Fisher	4	
	Stouffer	6	
	ComBat	13	
	limma	15	

U: total associations
2N: significant in 2 studies
I: significant in all studies

Pooling data after percentile normalization reduces false positives and increases true positives.



12 genera are significant within at least one colorectal cancer dataset, but not after pooling data



2 genera are significant after pooling, but not before

Duvallet et al., *Nature Communications* (2017); Gibbons et al., *PLoS Comp Bio* (in review)