Schloss Journal Club

Genetic Determinant of the Gut Microbiome in UK Twins

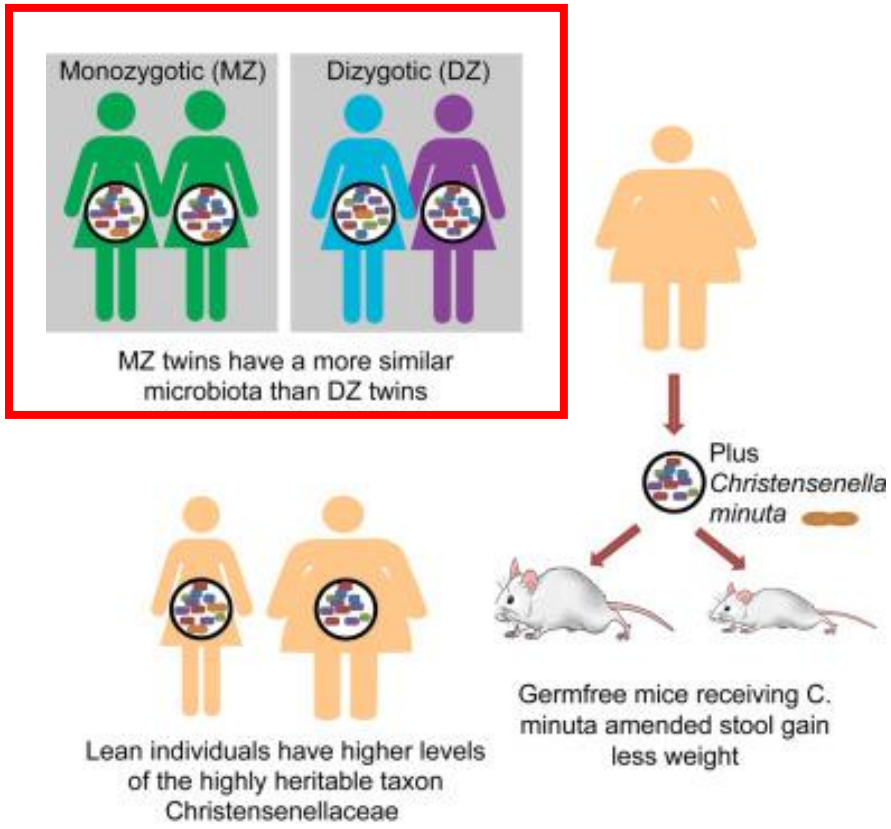
Marc Sze

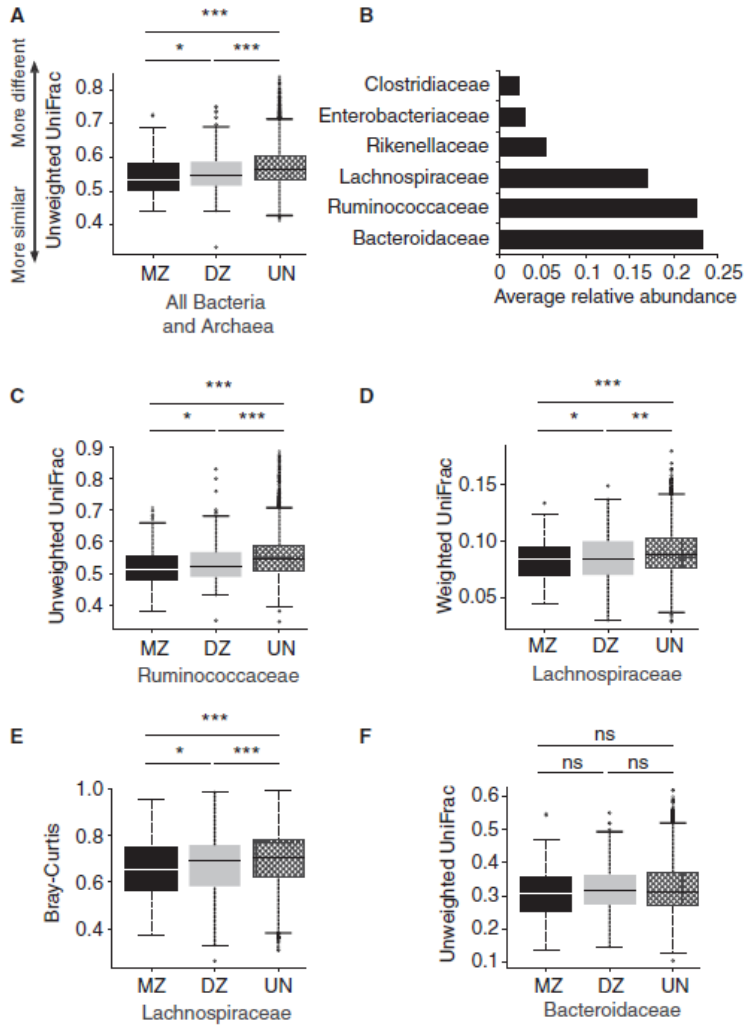
July 14th, 2016

Introduction

Human Genetics Shape the Gut Microbiome

Julia K. Goodrich,[1,2] Jillian L. Waters,[1,2] Angela C. Poole,[1,2] Jessica L. Sutter,[1,2] Omry Koren,[1,2,7] Ran Blekhman,[1,8] Michelle Beaumont,[3] William Van Treuren,[4] Rob Knight,[4,5,6] Jordana T. Bell,[3] Timothy D. Spector,[3] Andrew G. Clark,[1] and Ruth E. Ley[1,2,*]





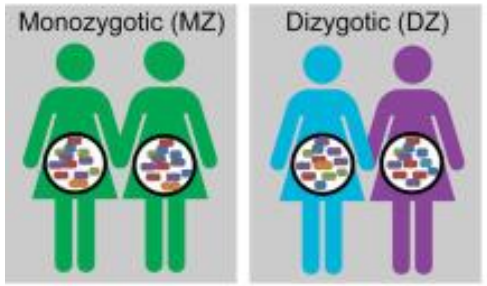
		P-value MZ more similar than DZ
Weighted UniFrac	All Bacteria	0.094
	Ruminococcaceae	0.07
	Lachnospiraceae	0.016
	Bacteroidaceae	0.806
Unweighted UniFrac	All Bacteria	0.032
	Ruminococcaceae	0.024
	Lachnospiraceae	0.39
	Bacteroidaceae	0.196
Bray Curtis	All Bacteria	0.11
	Ruminococcaceae	0.097
	Lachnospiraceae	0.036
	Bacteroidaceae	0.448

*No mention of multiple comparison correction

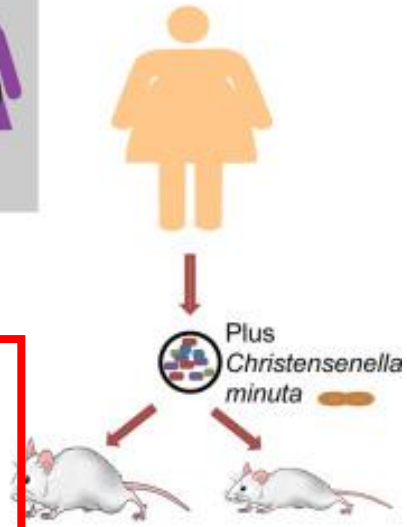
Introduction

Human Genetics Shape the Gut Microbiome

Julia K. Goodrich,[1,2] Jillian L. Waters,[1,2] Angela C. Poole,[1,2] Jessica L. Sutter,[1,2] Omry Koren,[1,2,7] Ran Blekhman,[1,8] Michelle Beaumont,[3] William Van Treuren,[4] Rob Knight,[4,5,6] Jordana T. Bell,[3] Timothy D. Spector,[3] Andrew G. Clark,[1] and Ruth E. Ley[1,2,*]

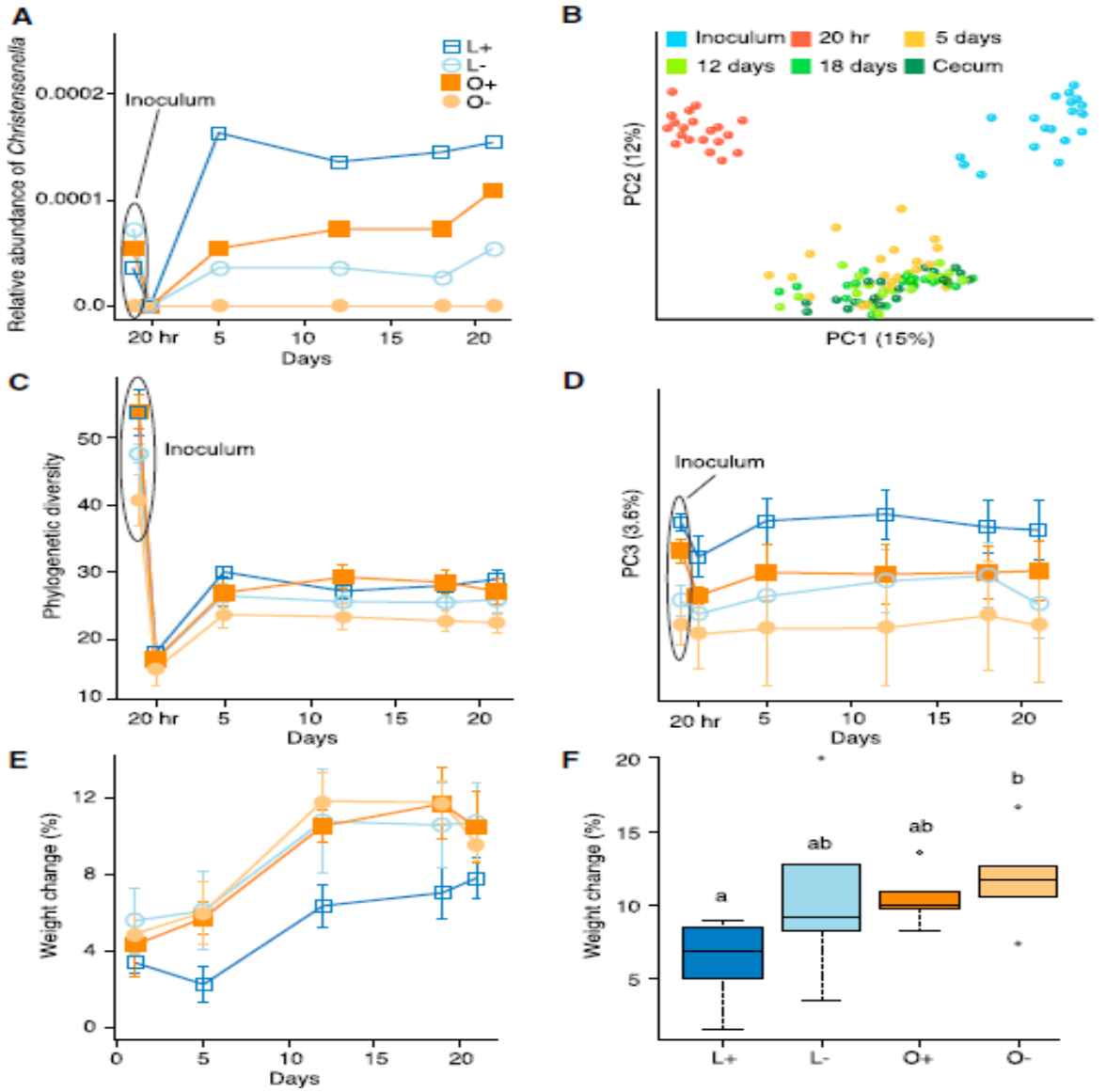


MZ twins have a more similar microbiota than DZ twins



Germfree mice receiving C. minuta amended stool gain less weight

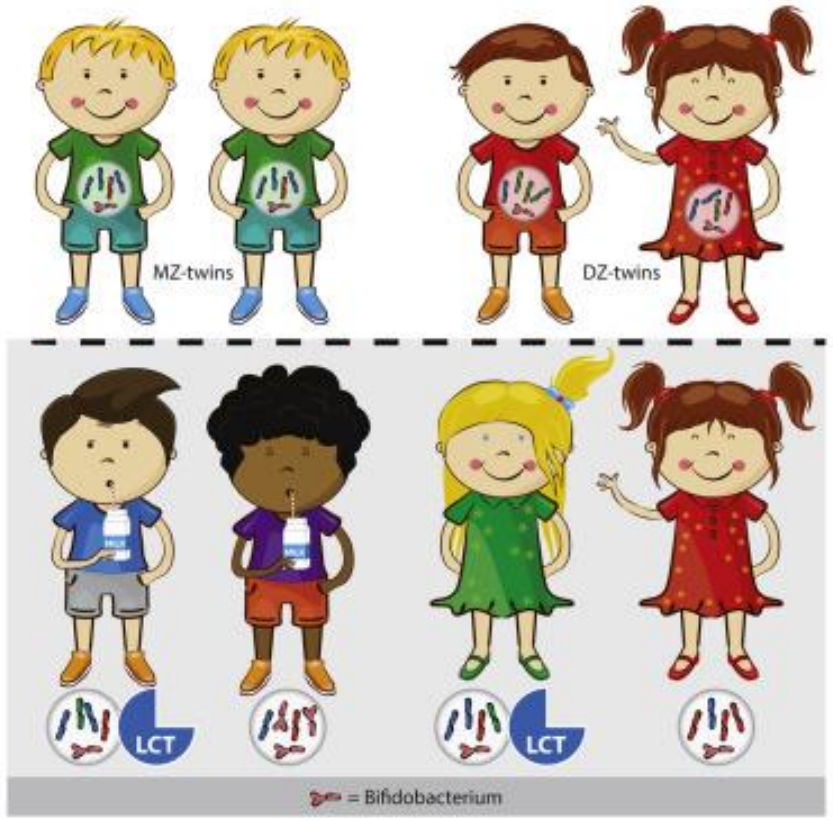
Lean individuals have higher levels of the highly heritable taxon Christensenellaceae



Introduction

Genetic Determinants of the Gut Microbiome in UK Twins

Julia K. Goodrich,[1] Emily R. Davenport,[1] Michelle Beaumont,[2] Matthew A. Jackson,[2] Rob Knight,[3] Carole Ober,[4] Tim D. Spector,[2] Jordana T. Bell,[2] Andrew G. Clark,[1] and Ruth E. Ley[1,5,*]



Mentioned by

- **34** news outlets
- **4** blogs
- **156** tweeters
- **3** Facebook pages
- **3** Google+ users
- **2** Redditors

Readers on

- **68** Mendeley
- **3** CiteULike

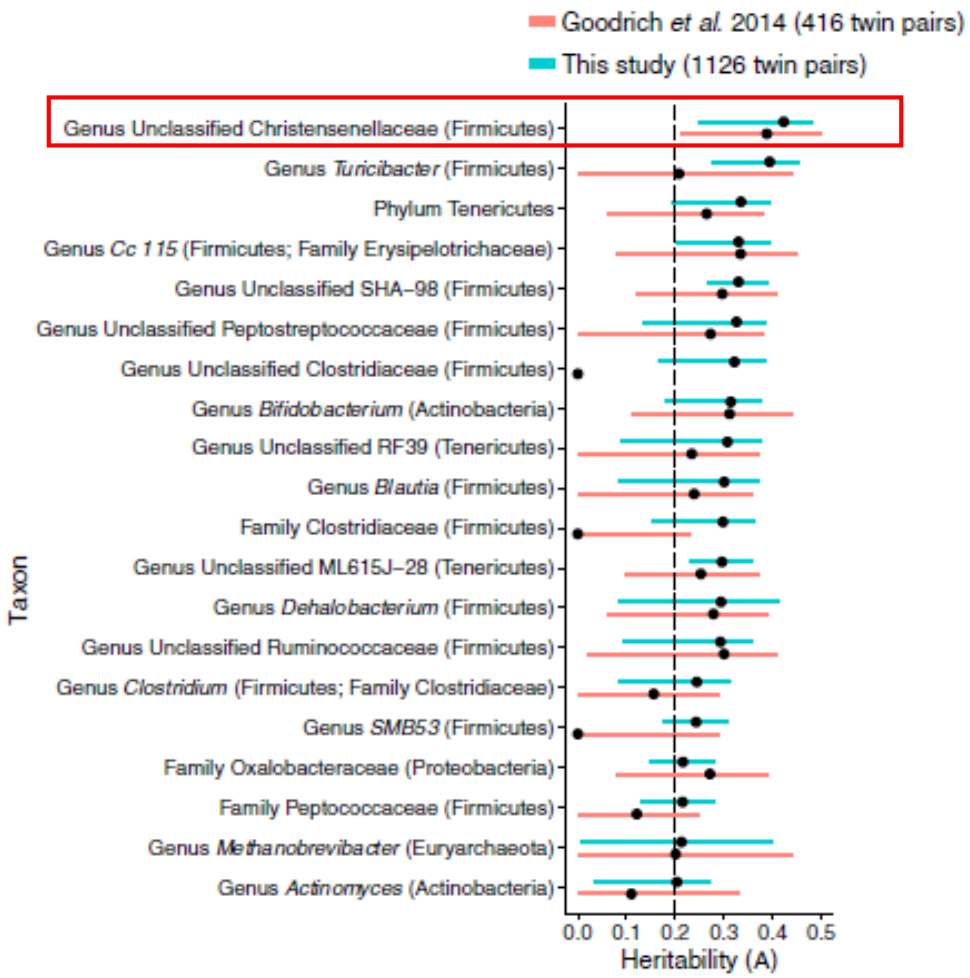
In the top 5% of all research outputs scored by Altmetric

- Replicate previous findings using larger n
- Identify additional heritable taxa
- Identify gene-microbiome interactions (co-evolution)

Heritability Calculations

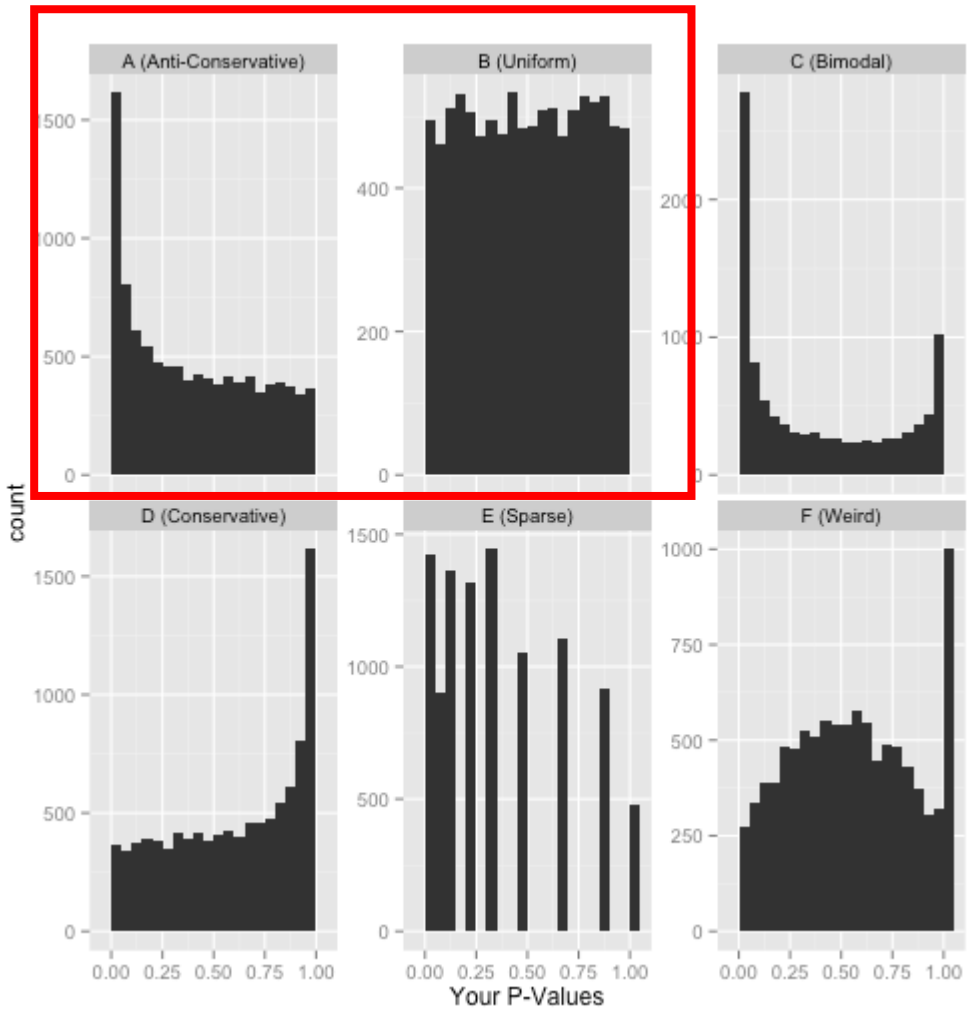
- Used ACE model
  - A (additive genetics)
    - $A = 2 * (r_{mz} – r_{dz})$
      - r_{mz} = correlation of trait in identical twins
      - r_{dz} = correlation of trait in fraternal twins
  - C (common environment)
    - $C = r_{mz} - A$
  - E (unique environment)
    - $E = 1 – r_{mz}$
- Falconer’s formula to calculate Heritability
  - $H2 = A = 2 * (r_{mz} – r_{dz})$

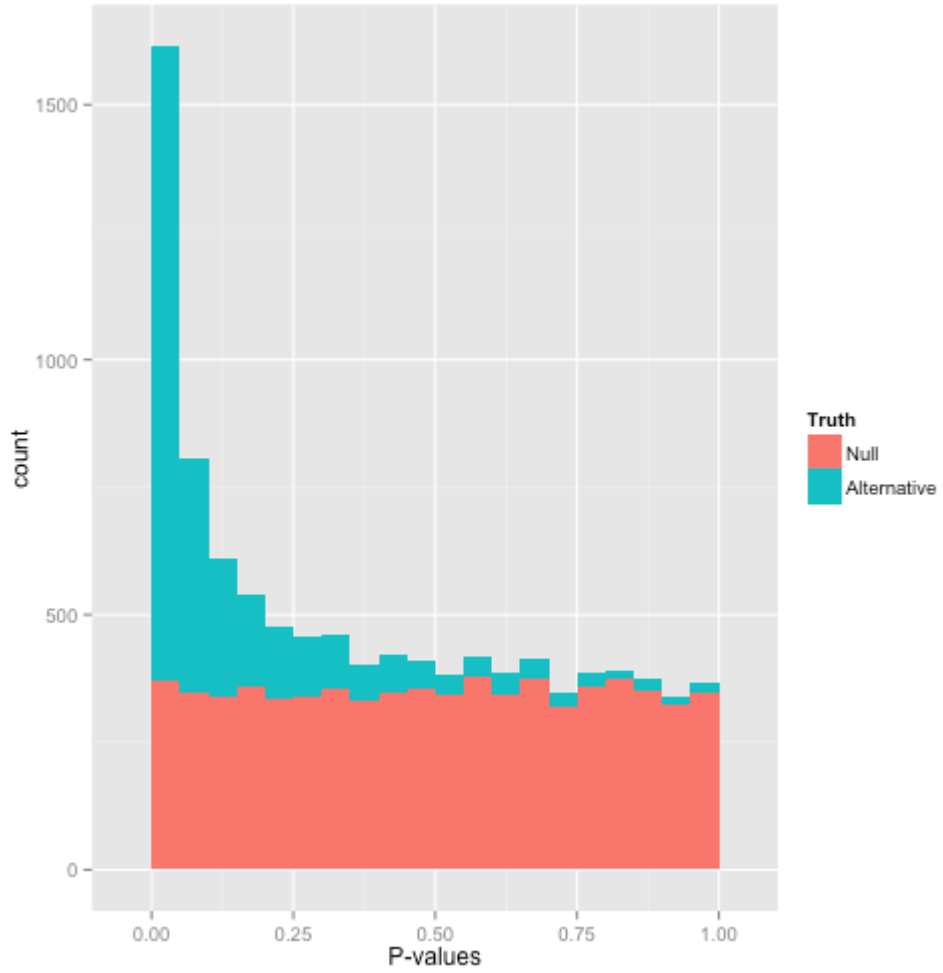
OTU Heritability and Replication



- A few discrepancies:
  - Performed two FDR tests one with all data and one with no OTUs.
  - Original P-values distribution is not one that is amenable to FDR correction (more on this on the next two slide)
  - Multiple significant findings from the same value used (non-independence of samples -> more on the next two slide)

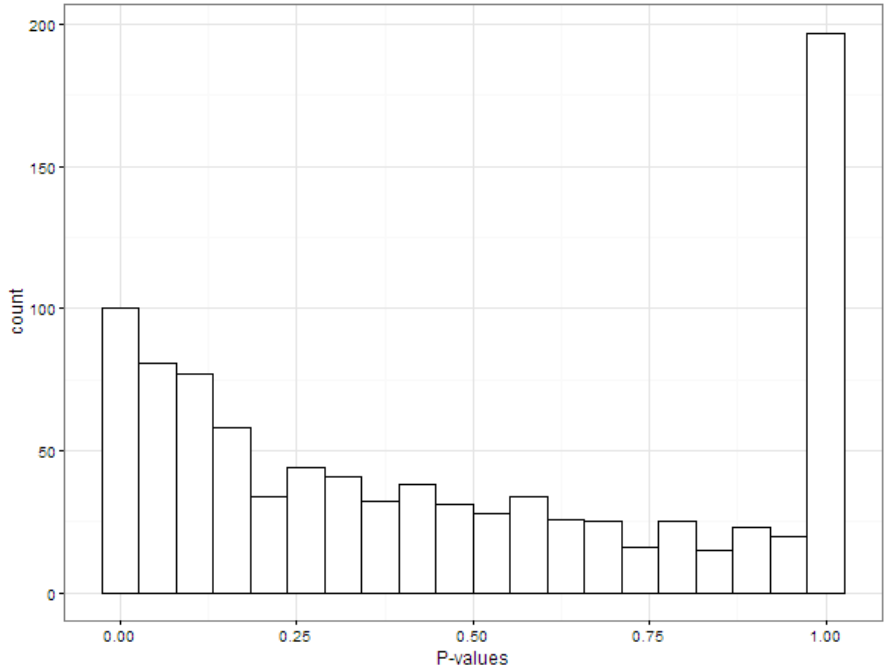
OTU Heritability and Replication



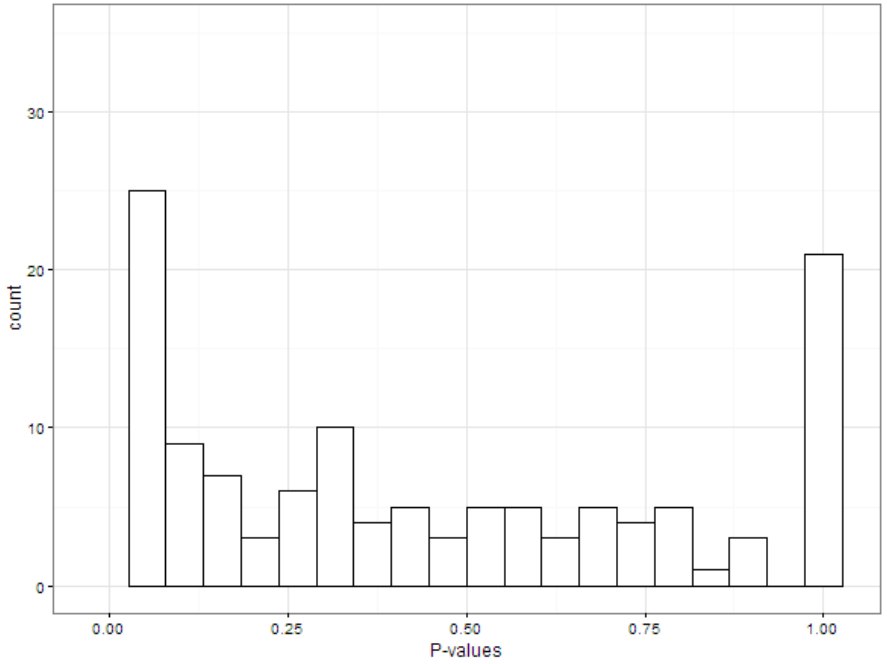


Images from David Robinson: http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

OTU Heritability and Replication



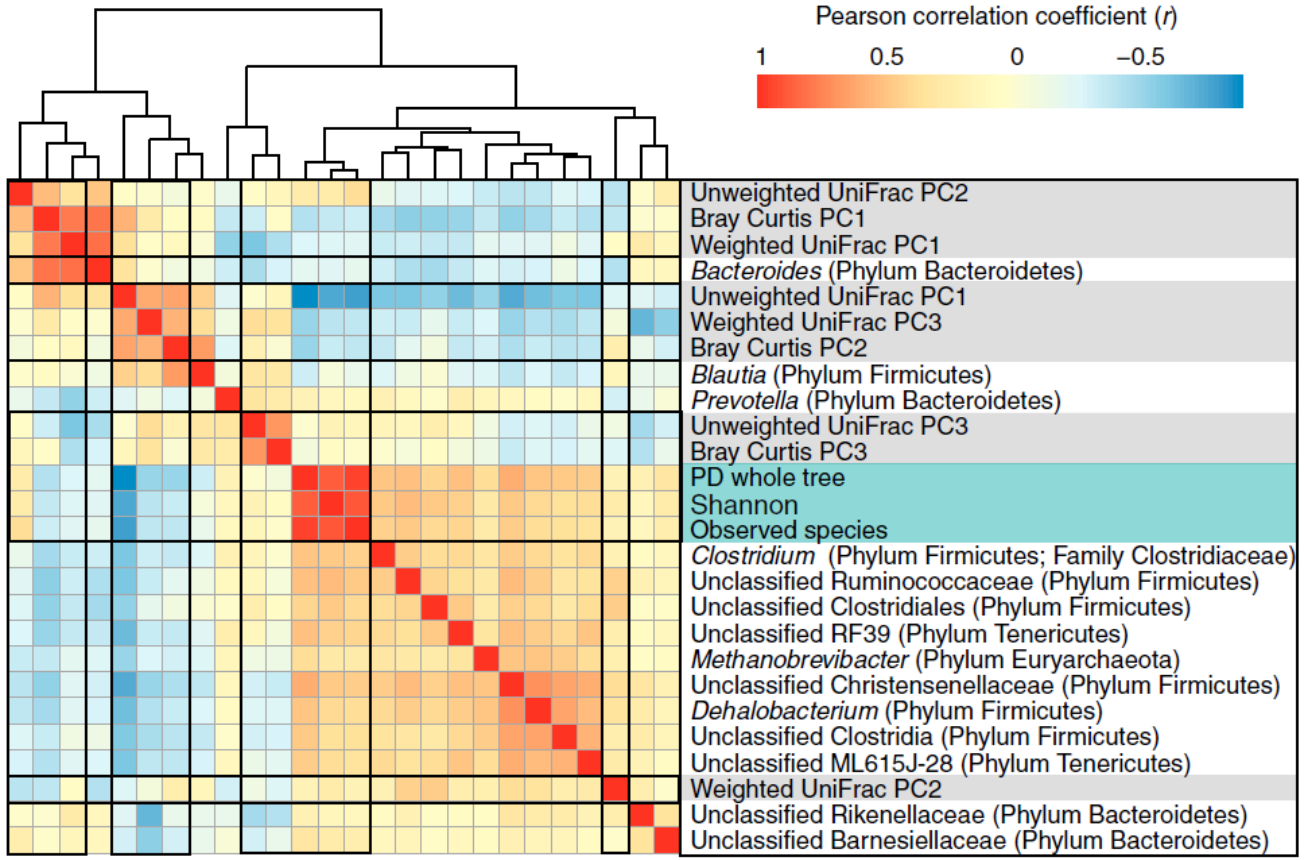
1st FDR with all 945 tests



2nd FDR with 163 tests

After removal of duplicates from 2nd test P-values < 0.05 went from 58 to 40

Correlation between Heritable bacteria and Alpha and Beta Diversity

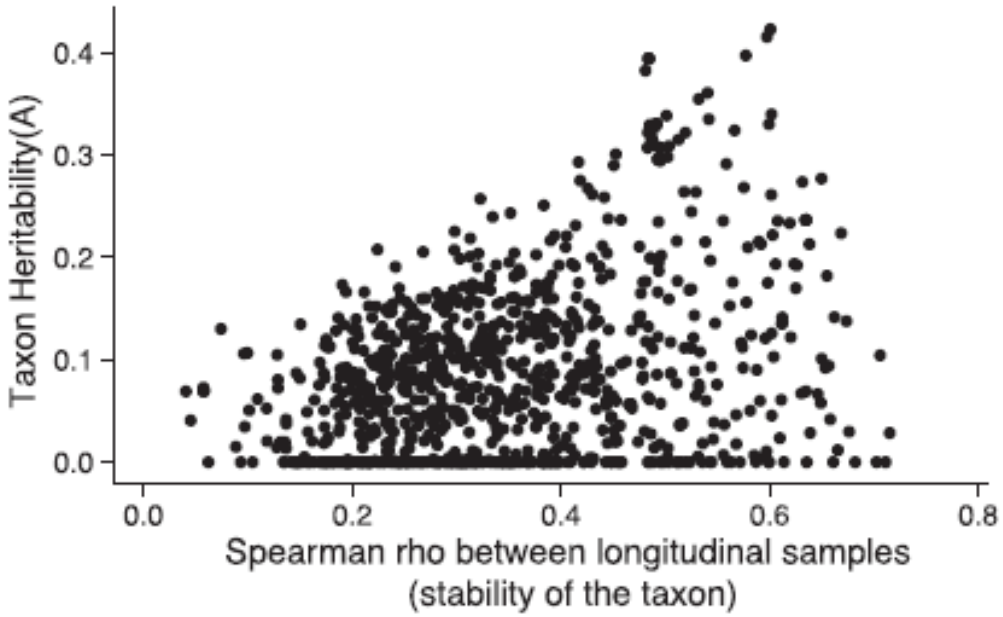


Heritability: Stability and Diversity

Table S2. Genetic and environmental contribution to Alpha- and Beta- diversity metrics. Related to Figure 2.

	Trait	A (95% CI)	*P*-value	C (95% CI)	E (95% CI)	ICC_{MZ}	ICC_{DZ}	Falconer h^2
Alpha Diversity	PD whole tree	0.37 (0.17-0.44)	0.0004	0.01 (0-0.19)	0.62 (0.56-0.68)	0.39	0.19	0.40
	Shannon	0.32 (0.26-0.38)	0.0005	0 (0-0.13)	0.68 (0.62-0.74)	0.33	0.14	0.38
	Observed species	0.3 (0.24-0.37)	0.0018	0 (0-0.14)	0.7 (0.63-0.76)	0.31	0.13	0.36
Beta Diversity	Unweighted UniFrac PC1	0.47 (0.29-0.58)	2.1×10^{-07}	0.06 (0-0.21)	0.47 (0.42-0.53)	0.53	0.29	0.48
	Unweighted UniFrac PC2	0.25 (0.09-0.31)	0.0053	0 (0-0.1)	0.75 (0.69-0.82)	0.25	0.10	0.32
	Unweighted UniFrac PC3	0.13 (0-0.32)	0.2559	0.13 (0-0.28)	0.74 (0.67-0.81)	0.26	0.19	0.15
	Weighted UniFrac PC1	0 (0-0.16)	1.0000	0.2 (0.06-0.26)	0.8 (0.74-0.85)	0.20	0.21	-0.03
	Weighted UniFrac PC2	0.15 (0-0.22)	0.1094	0 (0-0.15)	0.85 (0.78-0.92)	0.17	0.05	0.23
	Weighted UniFrac PC3	0.11 (0-0.29)	0.3388	0.12 (0-0.25)	0.77 (0.7-0.84)	0.23	0.17	0.12
	Bray Curtis PC1	0.15 (0-0.36)	0.1597	0.17 (0-0.33)	0.68 (0.61-0.75)	0.32	0.25	0.15
	Bray Curtis PC2	0.08 (0-0.3)	0.4830	0.17 (0-0.28)	0.75 (0.69-0.82)	0.26	0.20	0.12
	Bray Curtis PC3	0.17 (0-0.24)	0.0708	0 (0-0.16)	0.83 (0.76-0.9)	0.18	0.07	0.24

For each Metric, we report the A, C and E estimates from the ACE model (A = additive genetic influence; C = shared environmental influence; E = nonshared environmental influence) and *P*-value for A (likelihood ratio test comparing the ACE model to the CE model). Intraclass coefficient of correlation for MZ (ICC_{MZ}) and DZ (ICC_{DZ}), Falconer ($h^2 = 2*(ICC_{MZ}-ICC_{DZ})$) heritability.



Falconer's formula to calculate Heritability

$H2 = A = 2 * (r_{mz} - r_{dz})$

Gene, SNPs, and Heritable Taxa

Table 1. SNP and Gene Candidate Sets with Suspected Roles in Shaping the Gut Microbial Community

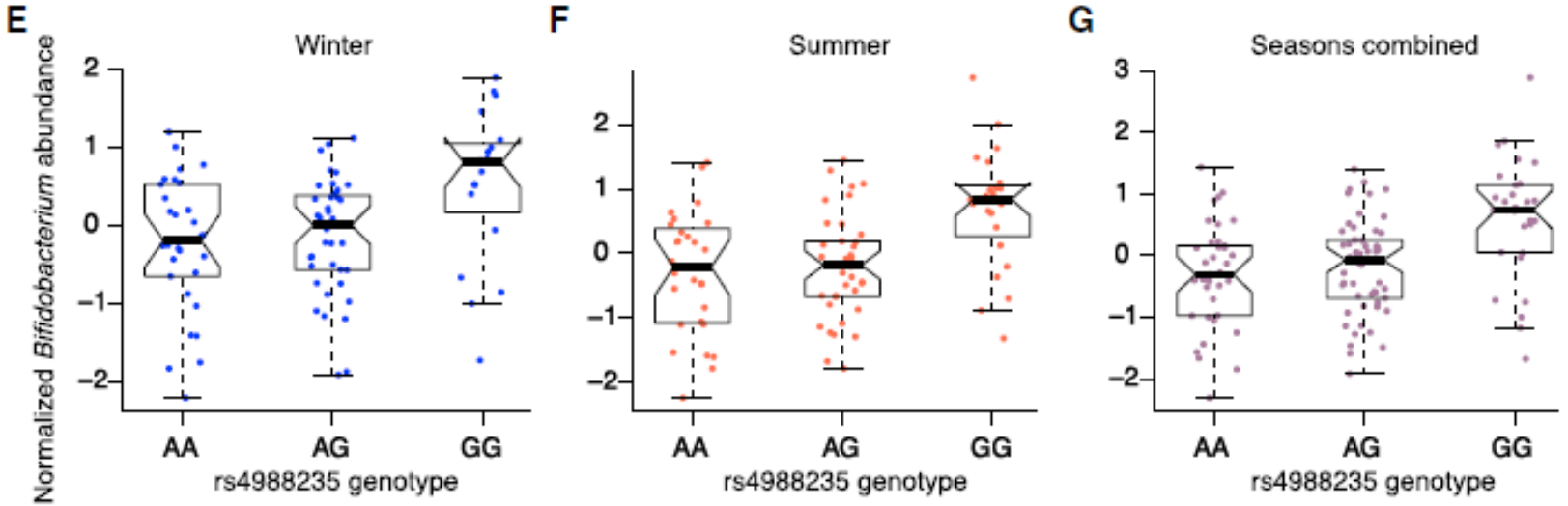
Candidate Set Name	Gene or SNP[a]	Number of Genes[b]	Number of SNPs[c]	Min p Value of Tested SNPs[d]	Min FDR Adjusted p Value of Tested SNPs[e]	Candidate Set Permutation p Value[f]	References
Gastric Cancer Risk	SNP		25	8.9×10^{-4}	0.446	0.286	Mocellin et al., 2015
Gastric Cancer Risk	Gene	29	611	8.0×10^{-5}	0.555	0.307	
Inflammatory Bowel Disease	SNP		140	2.4×10^{-5}	0.045	0.056	Jostins et al., 2012
Inflammatory Bowel Disease	Gene	109	4,337	6.8×10^{-6}	0.126	0.203	
Rheumatoid Arthritis	SNP		57	6.9×10^{-4}	0.443	0.511	Okada et al., 2014
Rheumatoid Arthritis	Gene	129	5,555	2.4×10^{-5}	0.599	0.585	
Type 2 Diabetes	SNP		65	0.001	0.527	0.703	Mahajan et al., 2014
Type 2 Diabetes	Gene	78	7,731	4.0×10^{-6}	0.602	0.214	
Non-HLA Associations with Cholestatic Disorders	Gene	87	5,350	2.5×10^{-5}	0.795	0.537	Hirschfield et al., 2013
Cilantro Soapy Taste	SNP		1	0.084	0.446	0.667	Eriksson et al., 2012
Cilantro Soapy Taste	Gene	1	13	9.1×10^{-5}	0.024	0.011	
Bitter Taste	Gene	21	136	0.001	0.717	0.367	Bachmanov and Beauchamp, 2007
Lipid Taste	Gene	1	51	6.1×10^{-5}	0.022	0.009	
Salty Taste	Gene	5	45	0.01	0.965	0.988	
Sour Taste	Gene	7	1,048	1.6×10^{-4}	0.492	0.659	
Sweet Taste	Gene	2	15	0.014	0.508	0.816	
Signal of Selection in Ancient Eurasians	SNP		9	0.003	0.603	0.37	Mathieson et al., 2015
Signal of Selection in Ancient Eurasians	Gene	43	1,122	3.5×10^{-6}	0.063	0.022	
Innate Immunity Genes [illegible] Selection	Gene	42	2,217	5.1×10^{-5}	0.425	0.565	Deschamps et al., 2016
LCT	Gene	1	21	3.5×10^{-6}	0.001	<0.001	
Davenport SNPs	SNP		187	1.7×10^{-4}	0.369	0.311	Davenport et al., 2015
NOD2	Gene	1	17	0.007	0.609	0.493	
Org Mouse eQTL	Gene	12	728	0.001	0.987	0.949	Org et al., 2015
Spor et al. (2011) Previous Microbiome Associations	Gene	27	461	6.2×10^{-4}	0.859	0.827	Spor et al., 2011

Candidate Set Name	Gene or SNP[a]	Number of Genes[b]	Number of SNPs[c]	Min p Value of Tested SNPs[d]	Min FDR Adjusted p Value of Tested SNPs[e]	Candidate Set Permutation p Value[f]	References
Blood Lipid Traits	SNP		86	2.1×10^{-7}	3.7×10^{-04}	<0.001	Teslovich et al., 2010
Sugar Transporters	Gene	2	66	0.006	0.789	0.68	
Tight Junctions	Gene	2	207	3.7×10^{-4}	0.262	0.32	
TLRs NLRPs ILs	Gene	15	430	6.0×10^{-4}	0.845	0.773	
SCFA receptors	Gene	2	6	0.005	0.409	0.378	
Sphingolipids	Gene	3	284	2.7×10^{-4}	0.445	0.335	

- Choose 20 taxa with heritability > 0.20 (all from figure 1)
- Significance for trait determined by taking lowest p-value and comparing it to 1000 iterations of p-values of the taxa residuals
  - If there were 3 tests done for a group (p-values = 0.05, 0.000001, and 0.4) the lowest of the three would then be compared to the distribution of the 1000 p-values of the taxa residuals. This would determine if the trait was significant or not.

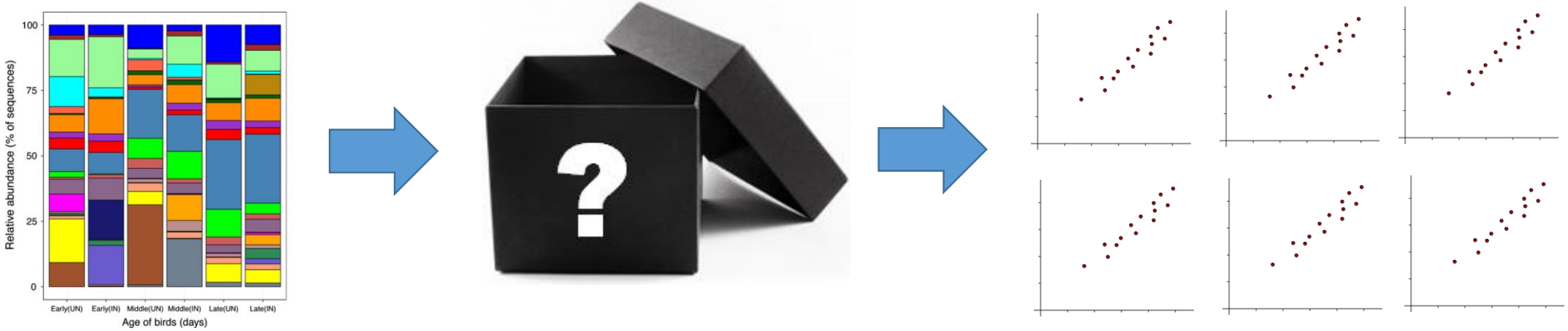
GWAS and Microbiome Part 1

- Tested all SNPs (1,300,091) against the 20 heritable taxa identified
  - Found no associations that achieved study wide significance (FDR < 0.2)
  - Potential correlation between Bifidobacterium and SNPs in genes involved with lactase persistence
    - Used Hutterite data set to try and validate this finding (Davenport, 2015)
      - Used **custom algorithm** to impute genotypes 98 individuals to 1,317 (>99% accurate in test cases)



GWAS and Microbiome Part 2

- Used program called microbiomeGWAS (Hua et al., 2015)
  - Two problems with initial manuscript (still in biorxiv)
    - No benchmarking against traditional methods
    - Correction removed the signal
  - Trying to correct the large skewness and positive kurtosis that arises from these types of comparisons (i.e. make normally distributed)
    - Typically use permutations or bootstrap



GWAS and Microbiome Part 2

- SNP (rs563779) within UHRF2 was associated with weighted UniFrac
  - P-value = 9.77×10^{-9}
- Two SNPs associated with Bray Curtis dissimilarity

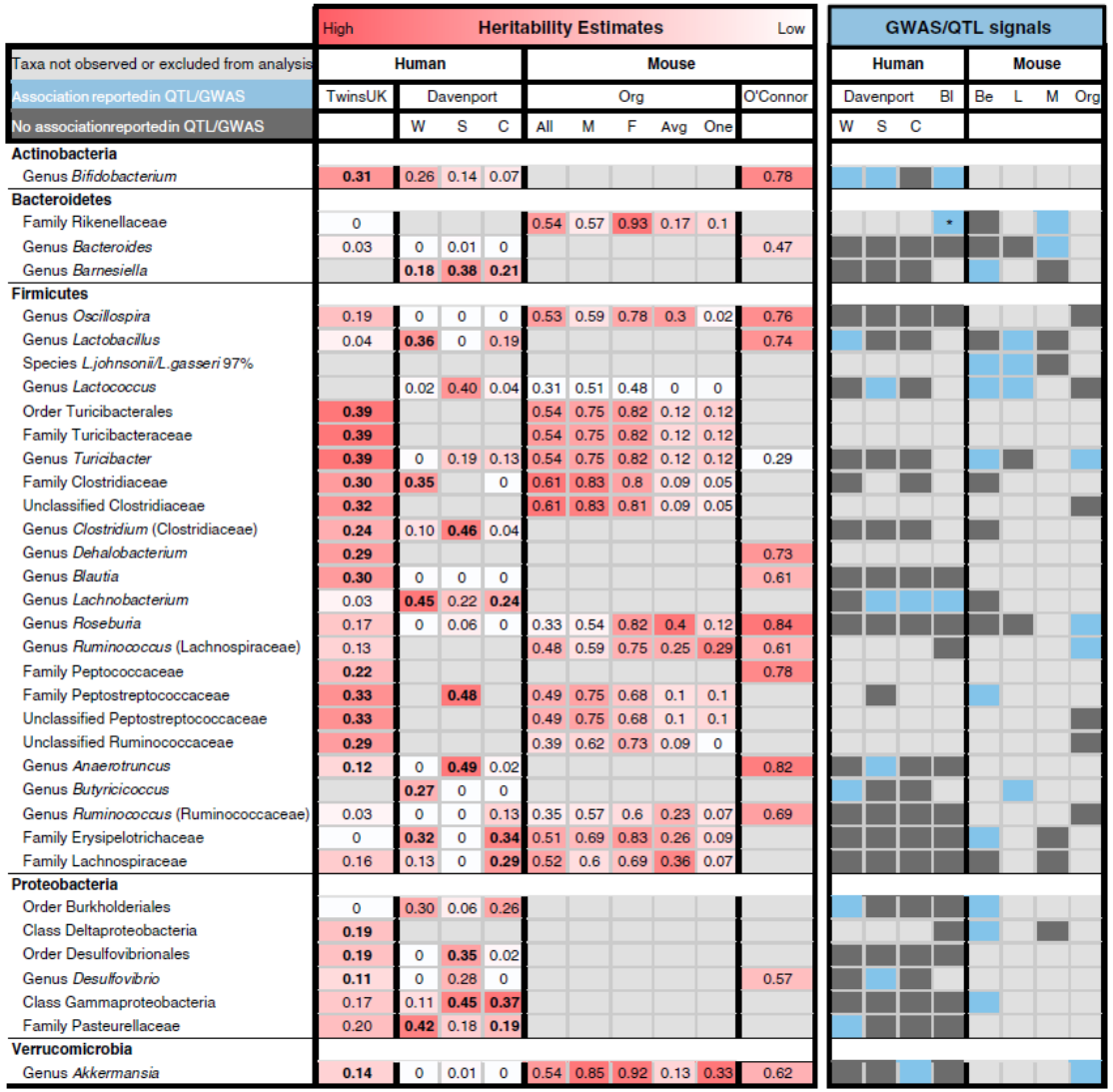
Imputed Gene Expression and Microbiome

- Used PrediXcan
  - Imputes expression in 40 different tissues based on SNPs and Genotype
- Compared full 945 taxa set
- Had two separate significance cutoffs
  - Study-wide
    - $0.05/(40x945x338372) = 3.91x10^{-12}$
  - Tissue-wide
    - $0.05/(945 \times \#\ \text{genes imputed in that tissue})$ or $5x10^{-8}$

Imputed Gene Expression and Microbiome

- Transverse colon SIGLEC15 expression associated with Akkermansia ($p = 6.21 \times 10^{-9}$)
- Number of other gene-microbiome under cut-off of $5x10^{-8}$
  - ZDHHC11B in Brain (cerebellum and Hypothalamus) – *ML615J.28*
  - RAB4B in Artery (Tibial) – *Lachnospiraceae (4331360)*
  - RPS27L Colon (Transverse) – *Ruminococcaceae (352347)*
  - HSF2 Heart (Left-Ventricle) – *Bacteroides*
  - RPS-468k18.5 Nerve (Tibial) – *Ruminococcaceae (44151)*
  - INSL3 Skin (Sun exposed lower leg) – *Lachnospiraceae (179384)*
  - JPH4 Spleen – *Clostridiales (193075)*
  - PPP1R3E Cells (Transformed Fibroblasts) – *Clostridiales (193075)*

Heritability across Studies

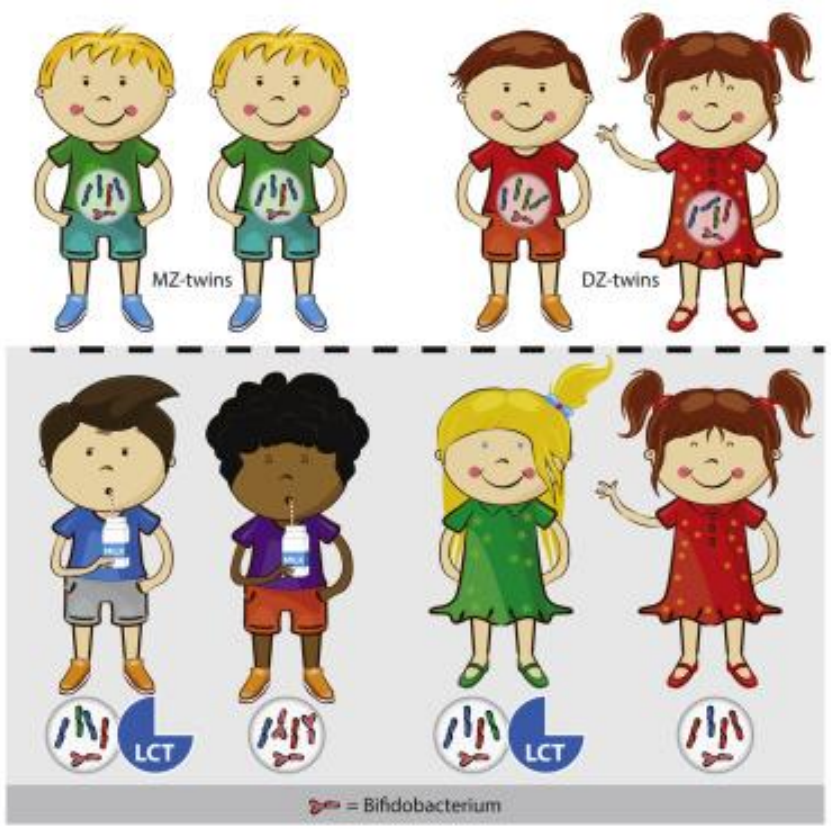


	High — Heritability Estimates — Low										GWAS/QTL signals						
Taxa not observed or excluded from analysis	Human				Mouse						Human			Mouse			
Association reported in QTL/GWAS	TwinsUK	Davenport			Org					O'Connor	Davenport		Bl	Be	L	M	Org
No association reported in QTL/GWAS		W	S	C	All	M	F	Avg	One		W	S	C				
Actinobacteria																	
Genus *Bifidobacterium*	**0.31**	0.26	0.14	0.07						0.78							
Bacteroidetes																	
Family Rikenellaceae	0				0.54	0.57	0.93	0.17	0.1					*			
Genus *Bacteroides*	0.03	0	0.01	0						0.47							
Genus *Barnesiella*		**0.18**	**0.38**	**0.21**													
Firmicutes																	
Genus *Oscillospira*	0.19	0	0	0	0.53	0.59	0.78	0.3	0.02	0.76							
Genus *Lactobacillus*	0.04	**0.36**	0	0.19						0.74							
Species *L.johnsonii/L.gasseri* 97%																	
Genus *Lactococcus*		0.02	0.40	0.04	0.31	0.51	0.48	0	0								
Order Turicibacterales	**0.39**				0.54	0.75	0.82	0.12	0.12								
Family Turicibacteraceae	**0.39**				0.54	0.75	0.82	0.12	0.12								
Genus *Turicibacter*	**0.39**	0	0.19	0.13	0.54	0.75	0.82	0.12	0.12	0.29							
Family Clostridiaceae	**0.30**	**0.35**		0	0.61	0.83	0.8	0.09	0.05								
Unclassified Clostridiaceae	**0.32**				0.61	0.83	0.81	0.09	0.05								
Genus *Clostridium* (Clostridiaceae)	**0.24**	0.10	**0.46**	0.04													
Genus *Dehalobacterium*	**0.29**									0.73							
Genus *Blautia*	**0.30**	0	0	0						0.61							
Genus *Lachnobacterium*	0.03	**0.45**	0.22	**0.24**													
Genus *Roseburia*	0.17	0	0.06	0	0.33	0.54	0.82	0.4	0.12	0.84							
Genus *Ruminococcus* (Lachnospiraceae)	0.13				0.48	0.59	0.75	0.25	0.29	0.61							
Family Peptococcaceae	**0.22**									0.78							
Family Peptostreptococcaceae	**0.33**		**0.48**		0.49	0.75	0.68	0.1	0.1								
Unclassified Peptostreptococcaceae	**0.33**				0.49	0.75	0.68	0.1	0.1								
Unclassified Ruminococcaceae	**0.29**				0.39	0.62	0.73	0.09	0								
Genus *Anaerotruncus*	**0.12**	0	**0.49**	0.02						0.82							
Genus *Butyricicoccus*		**0.27**	0	0													
Genus *Ruminococcus* (Ruminococcaceae)	0.03	0	0	0.13	0.35	0.57	0.6	0.23	0.07	0.69							
Family Erysipelotrichaceae	0	**0.32**	0	**0.34**	0.51	0.69	0.83	0.26	0.09								
Family Lachnospiraceae	0.16	0.13	0	**0.29**	0.52	0.6	0.69	0.36	0.07								
Proteobacteria																	
Order Burkholderiales	0	0.30	0.06	0.26													
Class Deltaproteobacteria	**0.19**																
Order Desulfovibrionales	**0.19**	0	**0.35**	0.02													
Genus *Desulfovibrio*	**0.11**	0	0.28	0						0.57							
Class Gammaproteobacteria	0.17	0.11	**0.45**	**0.37**													
Family Pasteurellaceae	0.20	**0.42**	0.18	**0.19**													
Verrucomicrobia																	
Genus *Akkermansia*	**0.14**	0	0.01	0	0.54	0.85	0.92	0.13	0.33	0.62							

**Alistipes* was in high correlation with Rikenellaceae so only the genus was tested

- Want to show that there is consistency between studies in both mouse and humans with respect to heritable taxa within the microbiome.

Summary



- The cartoon is supposed to emphasize that Bifdobacterium is a heritable taxa and that it is linked to polymorphisms in the LCT gene. So those with a specific SNP in the LCT gene will have more Bifdobacterium then those that do not.

- What do you think?