

Journal Club

The Effect of Host Genetics on the Gut Microbiome

October 20th, 2016

Marc Sze

Introduction



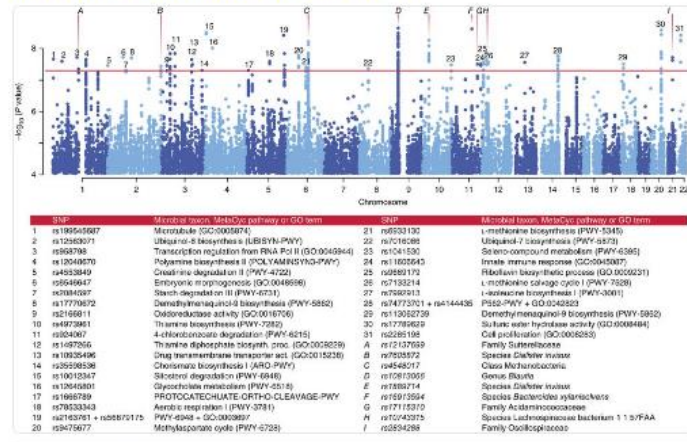
Ran Blekhman
@blekhman



Follow

The effect of host genetics on the gut microbiome

[nature.com/ng/journal/vao](https://www.nature.com/ng/journal/vao) ...



RETWEETS
26

LIKES
22

8:26 PM - 5 Oct 2016



26



22



Ran Blekhman @blekhman · Oct 6

@ewanbimey @WvSchaik @dgmcarthur for sure there are false positives, but I trust the LCT-bifido link. Validated in 3 independent papers



2

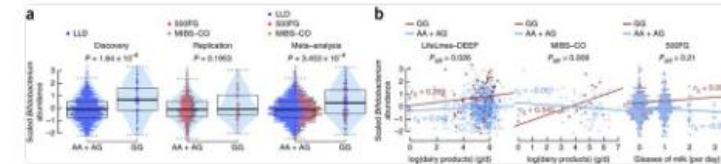


9



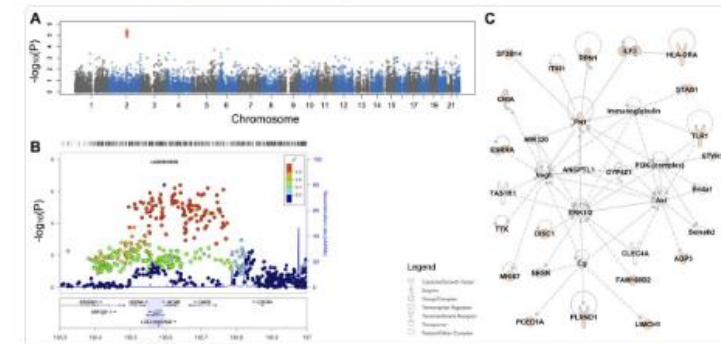
Ran Blekhman @blekhman · Oct 5

Great work looking into the interaction between diet, abundance of bifidobacterium, and genetic variation in the human LCT gene



Ran Blekhman @blekhman · Oct 5

We identified the LCT-bifido link here: genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0944-4, & was recently validated by Goodrich et al: [cell.com/cell-host-microbiome](https://www.cell.com/cell-host-microbiome)



Introduction

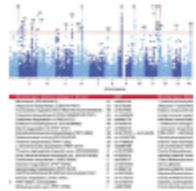


WvSchaik
@WvSchaik



Following

I would be really interested to hear what GWAS experts (@dgmacarthur?) make of this paper (which I like, but fear for spurious correlations)



Ran Blekhman @blekhman

The effect of host genetics on the gut microbiome
nature.com/ng/journal/vao...

RETWEETS

7

LIKES

6



4:24 PM - 6 Oct 2016



7



6



Daniel MacArthur @dgmacarthur · Oct 6

@blekhman @ewanbirney @WvSchaik I couldn't see any big red flags. Good people involved.



1



4



Ewan Birney @ewanbirney · Oct 6

@WvSchaik @dgmacarthur @blekhman however instinctively they need a higher multiple testing penalty than $5e-8$ - will dig in next week



1



3



David Eccles @gringene_bio · Oct 6

@WvSchaik @BioMickWatson @dgmacarthur I'd like to see a bootstrapped version of this. That plot looks a little bit population-ey.



1



Paul Gordon @YYCist · Oct 6

@ewanbirney @BioMickWatson @WvSchaik @dgmacarthur @blekhman based on 3 Dutch cohorts. Poss SNP/microbiome confounding by mother-child pairs



2



2



Paul Gordon @YYCist · Oct 6

@ewanbirney @BioMickWatson @WvSchaik @dgmacarthur @blekhman LifeLines-DEEP cohort explicitly multigeneration, methods used to ctl for this?



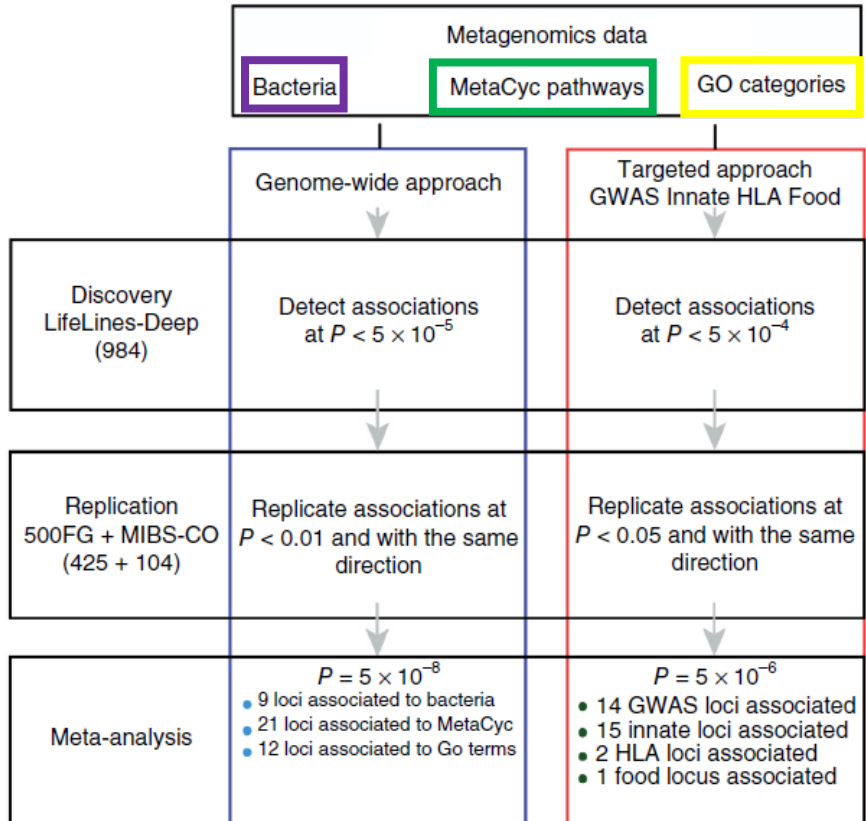
1



2

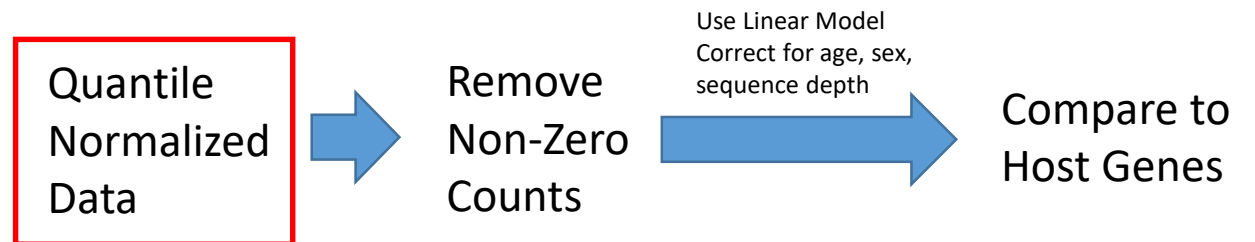


Overview



- MetaPhlan2.2 to generate abundance data
- HUMAN2 to generate 636 metabolic pathway counts
- HUMAN2 to generate 661 non-redundant GO terms for analysis

General Analysis Pipeline:



Quantile Normalization

Gene	S1	S2	S3
A	5	4	3
B	2	1	4
C	3	4	6
D	4	2	8

2



Rearrange
values from
lowest to highest

Gene	S1	S2	S3
A	2	1	3
B	3	2	4
C	4	4	6
D	5	4	8

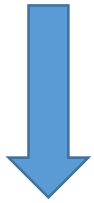
3



Find mean
of each
row

Rank	Mean
I	2
II	3
III	4.67
IV	5.67

1



Rank
lowest to
highest

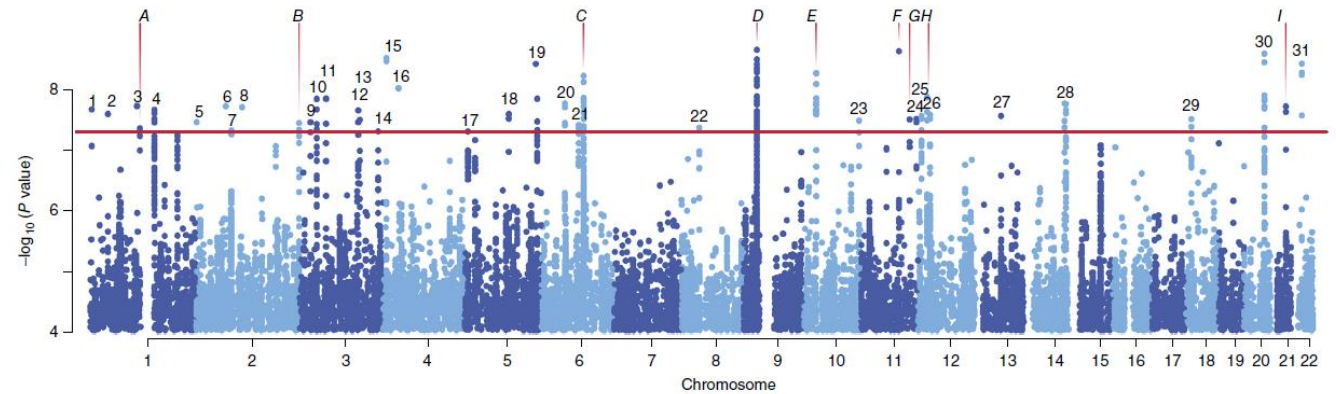
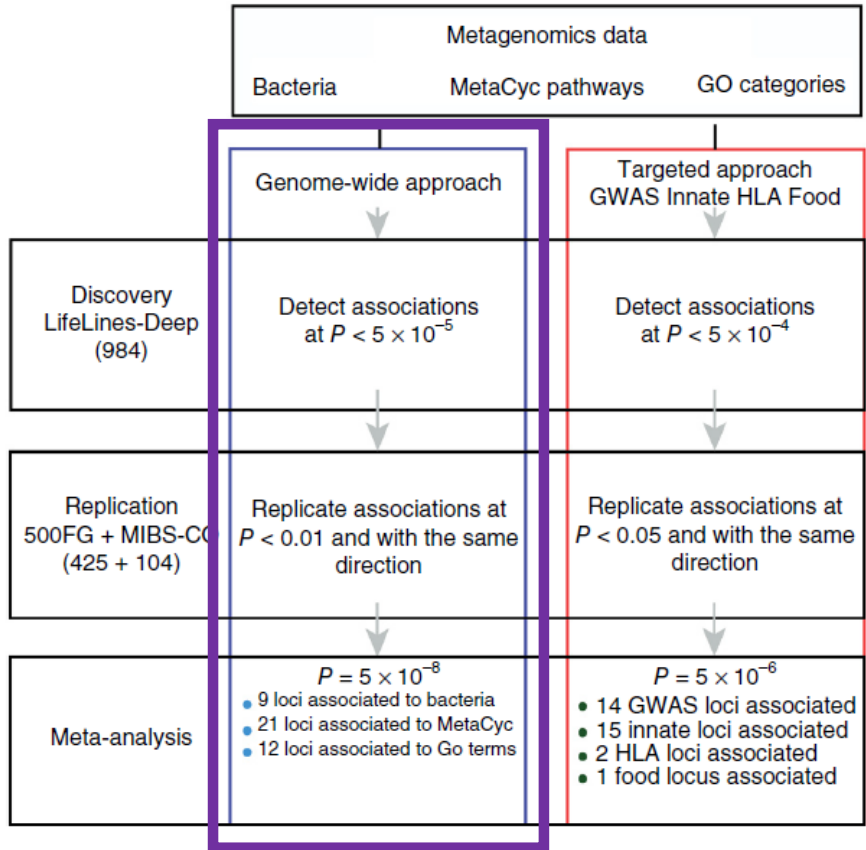
Gene	S1	S2	S3
A	IV	III	I
B	I	I	II
C	II	III	III
D	III	II	IV

4

Use rankings to assign
mean values to gene and sample.

Gene	S1	S2	S3
A	5.67	4.67	2
B	2	2	3
C	3	4.67	4.67
D	4.67	3	5.67

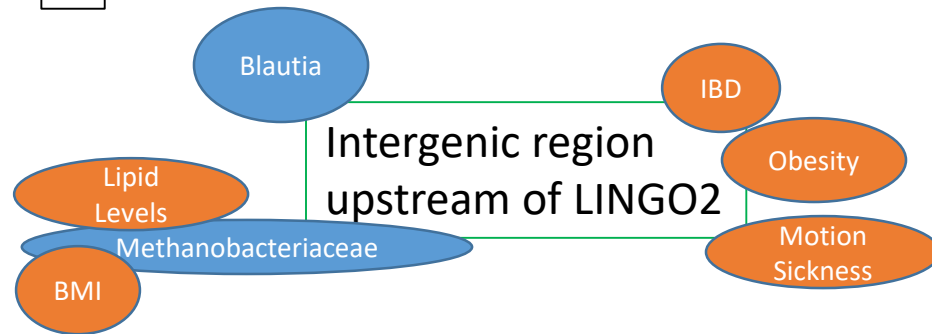
Results: GWAS and Microbiome



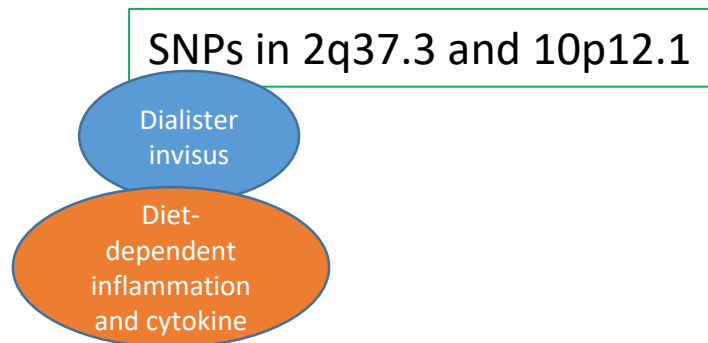
	SNP	Microbial taxon, MetaCyc pathway or GO term		SNP	Microbial taxon, MetaCyc pathway or GO term
1	rs199545687	Microtubule (GO:0005784)	21	rs6933130	L-methionine biosynthesis (PWY-5345)
2	rs12563071	Ubiquinol-8 biosynthesis (UBISYN-PWY)	22	rs7016086	Ubiquinol-7 biosynthesis (PWY-5873)
3	rs958798	Transcription regulation from RNA Pol II (GO:0045944)	23	rs1041530	Seleno-compound metabolism (PWY-6395)
4	rs12048670	Polyamine biosynthesis II (POLYAMINSYN3-PWY)	24	rs11606643	Innate immune response (GO:0045087)
5	rs4553849	Creatinine degradation II (PWY-4722)	25	rs9669179	Riboflavin biosynthetic process (GO:0009231)
6	rs6546647	Embryonic morphogenesis (GO:0048598)	26	rs7133214	L-methionine salvage cycle I (PWY-7528)
7	rs2084597	Starch degradation III (PWY-6731)	27	rs7992913	L-isoleucine biosynthesis I (PWY-3001)
8	rs17770672	Demethylmenaquinol-9 biosynthesis (PWY-5862)	28	rs74773701 + rs4144435	P562-PWY + GO:0042823
9	rs2166811	Oxidoreductase activity (GO:0016706)	29	rs113062739	Demethylmenaquinol-9 biosynthesis (PWY-5862)
10	rs4973961	Thiamine biosynthesis (PWY-7282)	30	rs17789629	Sulfur ester hydrolase activity (GO:0008484)
11	rs924067	4-chlorobenzoate degradation (PWY-6215)	31	rs2285198	Cell proliferation (GO:0008283)
12	rs1497266	Thiamine diphosphate biosynth. proc. (GO:0009229)	A	rs12137699	Family Sutterellaceae
13	rs10935496	Drug transmembrane transporter act. (GO:0015238)	B	rs7605872	Species <i>Dialister invisus</i>
14	rs35598536	Chorismate biosynthesis I (ARO-PWY)	C	rs4548017	Class Methanobacteria
15	rs10012347	Sitosterol degradation (PWY-6948)	D	rs10813066	Genus <i>Blautia</i>
16	rs12645801	Glycocholate metabolism (PWY-6518)	E	rs1889714	Species <i>Dialister invisus</i>
17	rs1666789	PROTOCATECHUATE-ORTHO-CLEAVAGE-PWY	F	rs16913594	Species <i>Bacteroides xylanisolvens</i>
18	rs78533343	Aerobic respiration I (PWY-3781)	G	rs17115310	Family Acidaminococcaceae
19	rs2163761 + rs56879175	PWY-6948 + GO:0003697	H	rs10743315	Species <i>Lachnospiraceae bacterium 1157FAA</i>
20	rs9475677	Methylaspartate cycle (PWY-6728)	I	rs2834288	Family Oscillospiraceae

Results: GWAS and Microbiome - Taxonomy

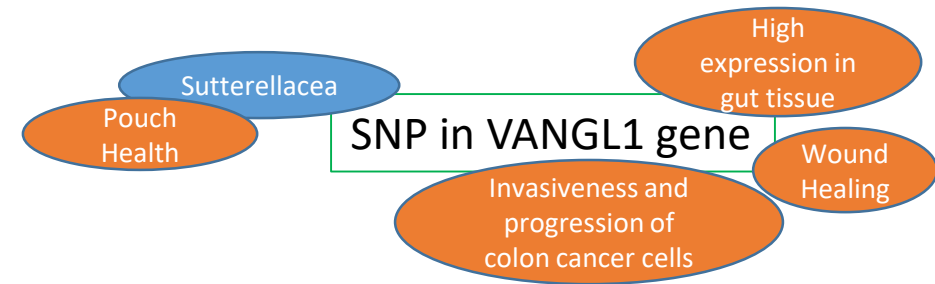
1



2

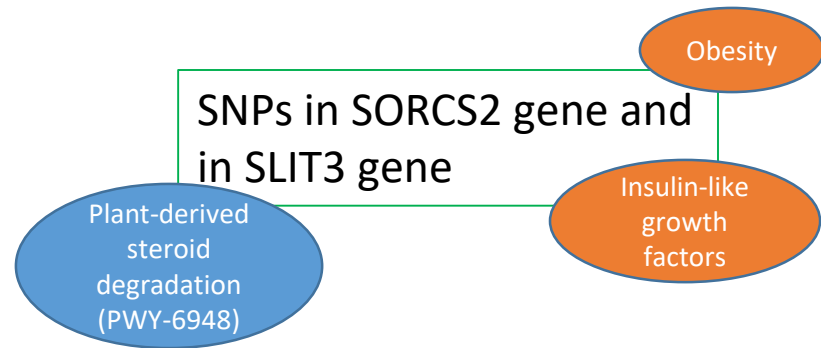


3

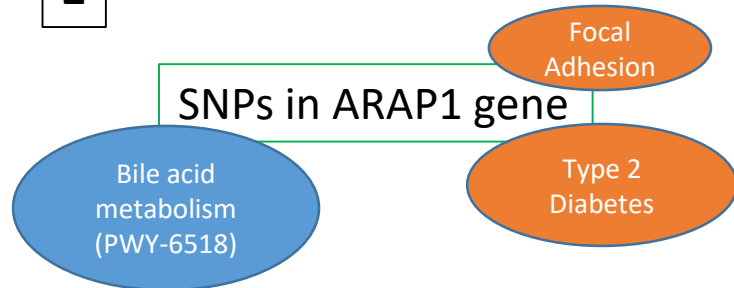


Results: GWAS and Microbiome - Pathways

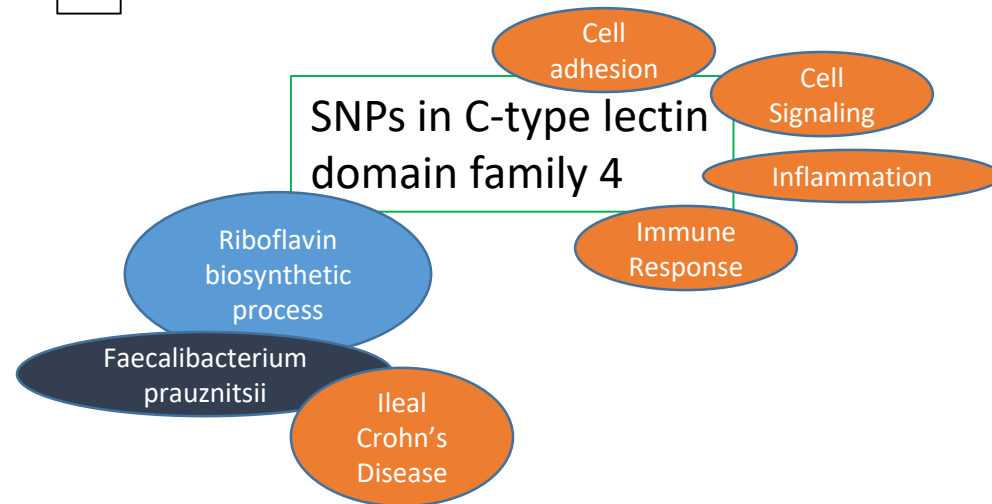
1



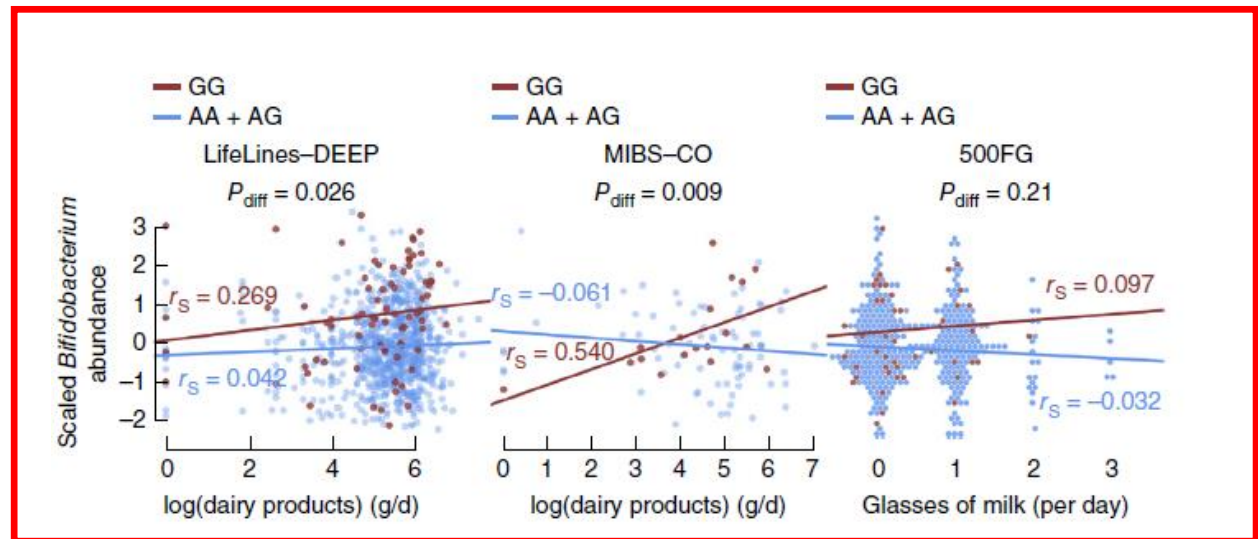
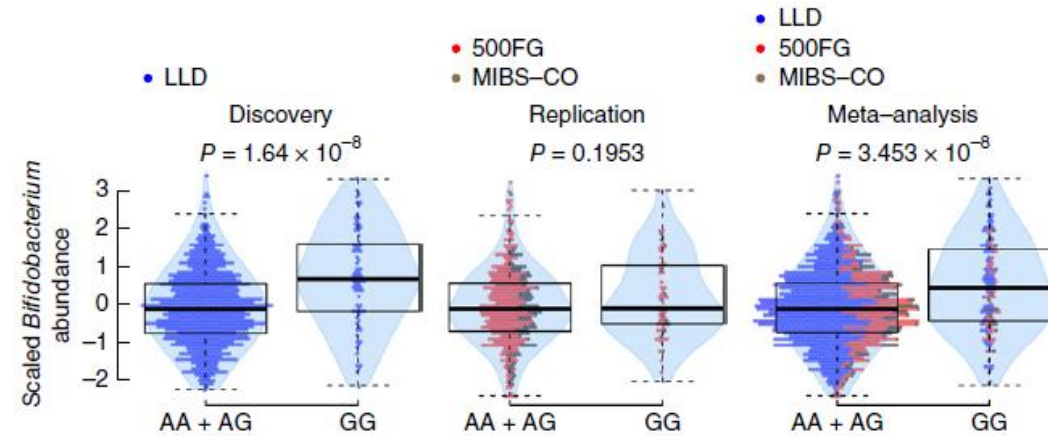
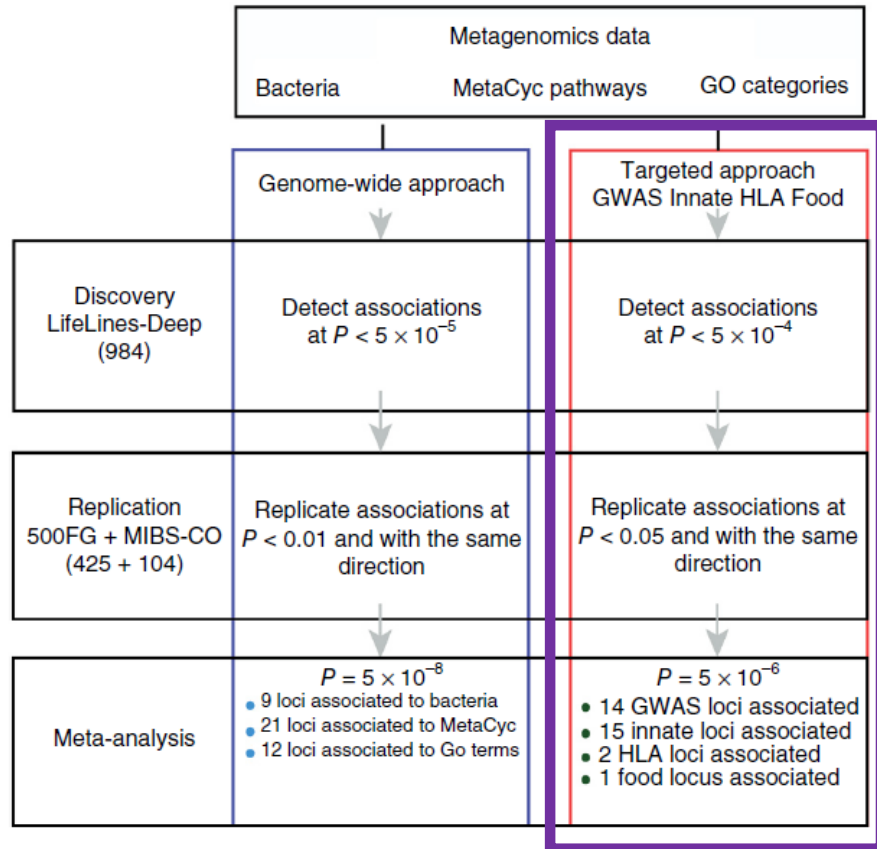
2



3



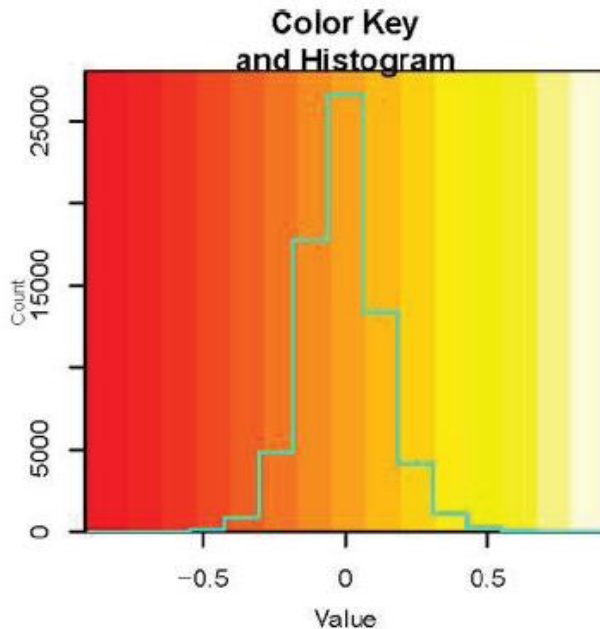
Results: Targeted Approach and Microbiome



Some Important Caveats

Ewan Birney @ewanbirney · Oct 6
@WvSchaik @dgmacarthur @blekman however instinctively they need a higher multiple testing penalty than $5e-8$ - will dig in next week

David Eccles @gringene_bio · Oct 6
@WvSchaik @BioMickWatson @dgmacarthur I'd like to see a bootstrapped version of this. That plot looks a little bit population-ey.



- Some “Tricksy” Math

- Based on authors chosen P-value the expectation is for 50,000 False Positives in first stage
 - No report of total comparisons under P-value to judge how many in first stage could be expected to be not false positives (e.g. 50,004 -50,000)
- Judge their False Discovery Rate for second stage (validation) based on the 50,000 from the first stage.
 - $FDR = 50,000 \times 0.01 = 500$
 - My thoughts are validation FDR should be independent not dependent on first set for true validation.
- Have a higher trust in combined data
 - Cutoff p-value = 5×10^{-8}
 - FDR ranges from 1.5 – 5 (so 2%-12%)
 - Accepted loose FDR now accepted as 5%

Some Important Caveats



Paul Gordon @YYCist · Oct 6

@ewanbirney @BioMickWatson @WvSchaik @dgmacarthur @blekhman based on 3 Dutch cohorts. Poss SNP/microbiome confounding by mother-child pairs

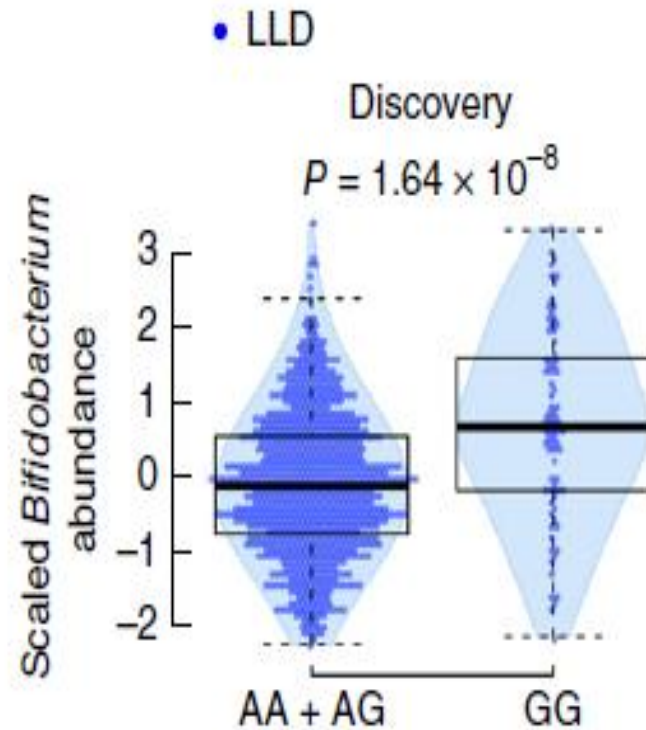
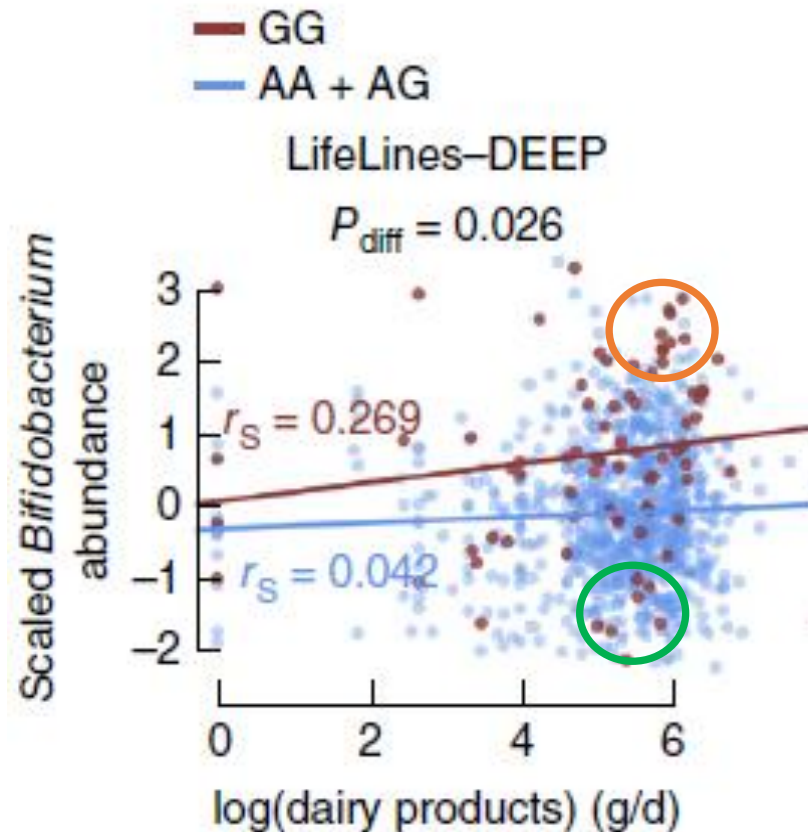
← ↻ 2 ❤️ 2 ⋮



Paul Gordon @YYCist · Oct 6

@ewanbirney @BioMickWatson @WvSchaik @dgmacarthur @blekhman LifeLines-DEEP cohort explicitly multigeneration, methods used to ctrl for this?

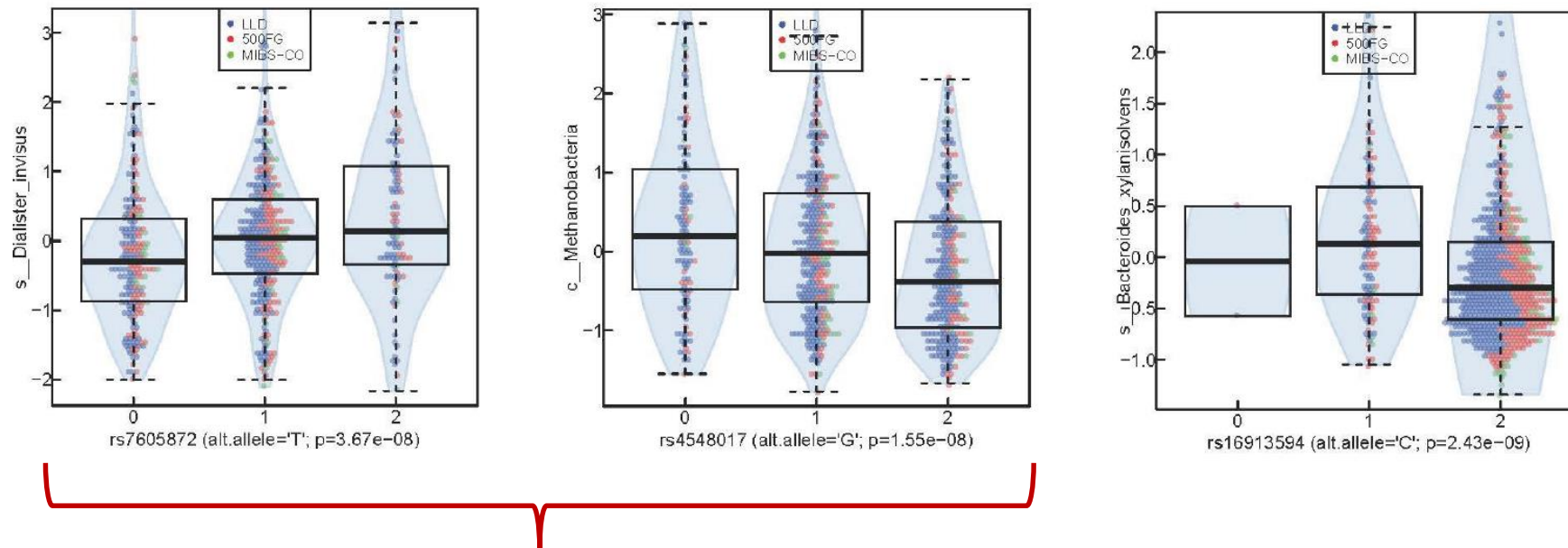
← ↻ 1 ❤️ 2 ⋮



Could be comparing
Extended families with
shared environments.
Confounds the results.

Some Important Caveats

- Three types of patterns emerge from significant data
 - Two out of three are easily biologically interpretable



These are easily interpretable based on biology

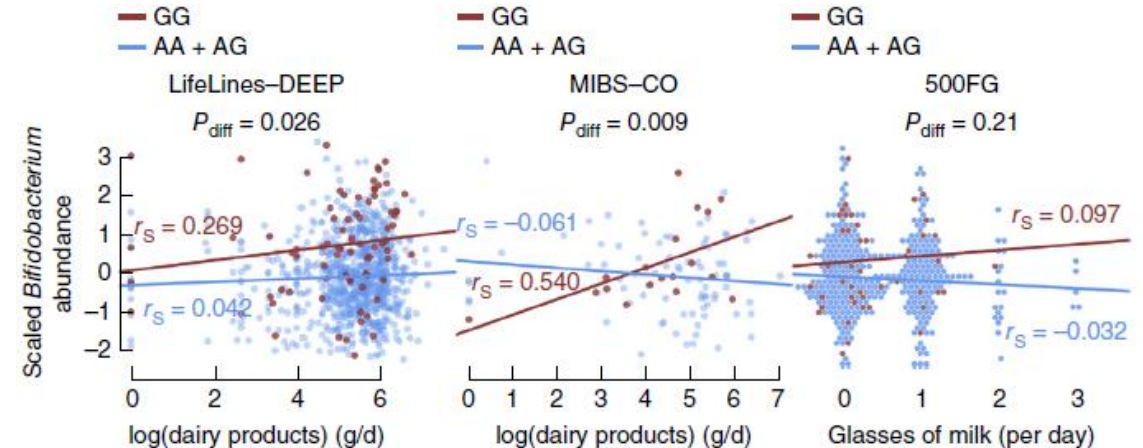
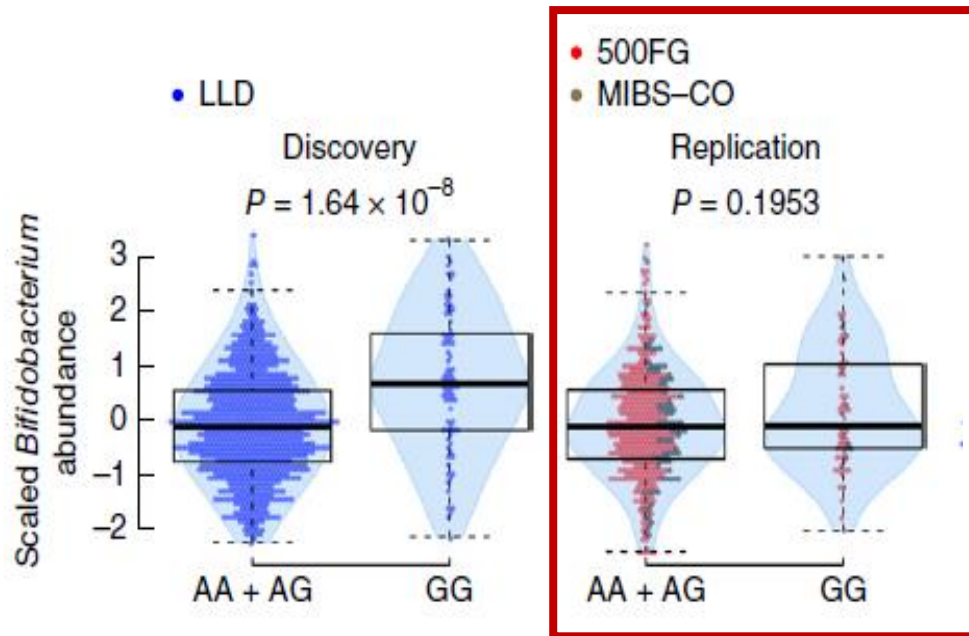
And Now the LCT Story



Ran Blekhman @blekhman · Oct 6

@ewanbirney @WvSchaik @dgmacarthur for sure there are false positives, but I trust the LCT-bifido link. Validated in 3 independent papers

👍 2 ❤️ 9 ...



- 1) Ideally qPCR for Bifidobacterium would have been good to complement the inferred abundances
- 2) Would categorize the evidence as still up for debate.

Thank you for your time and attention!!