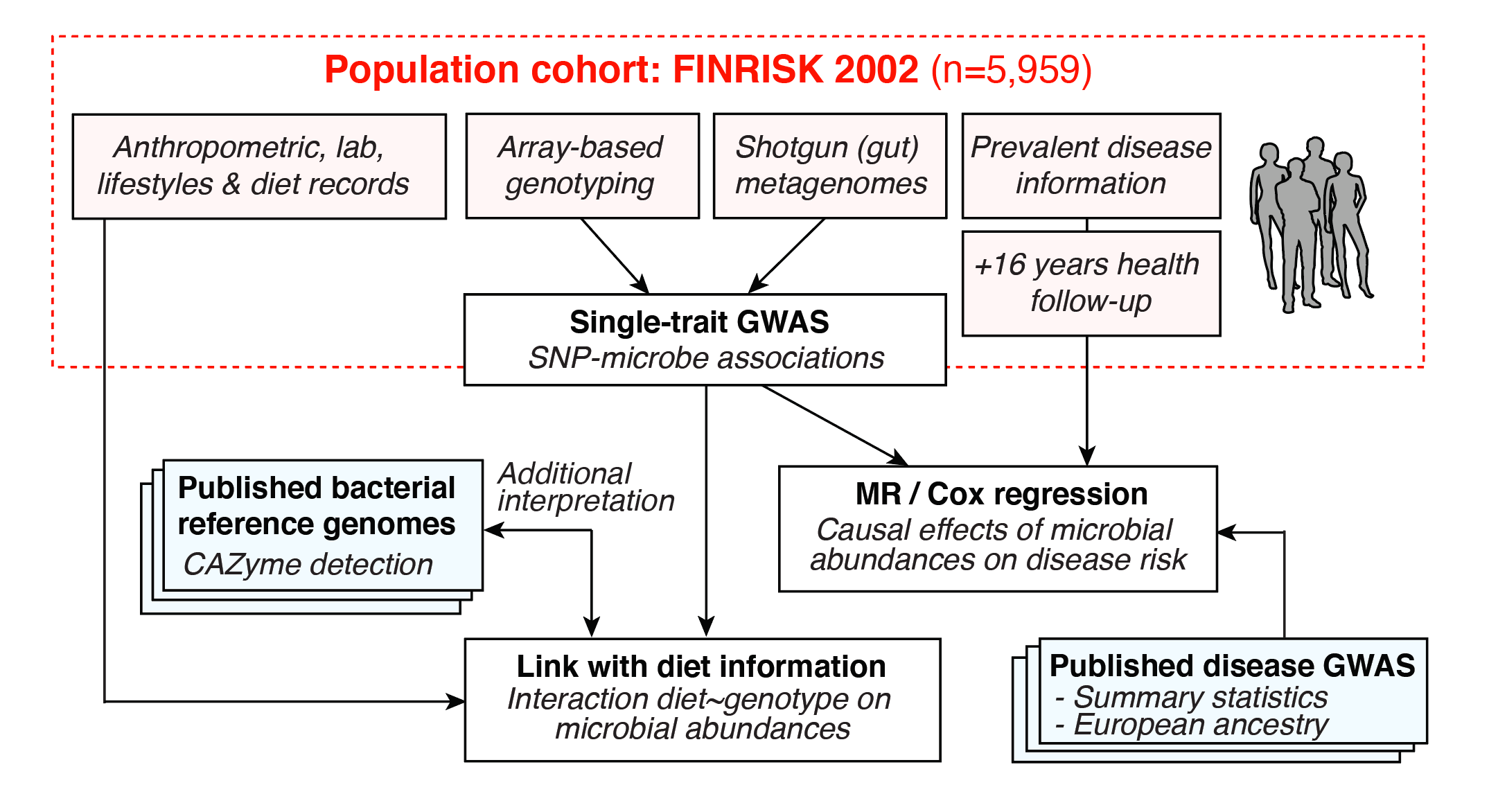
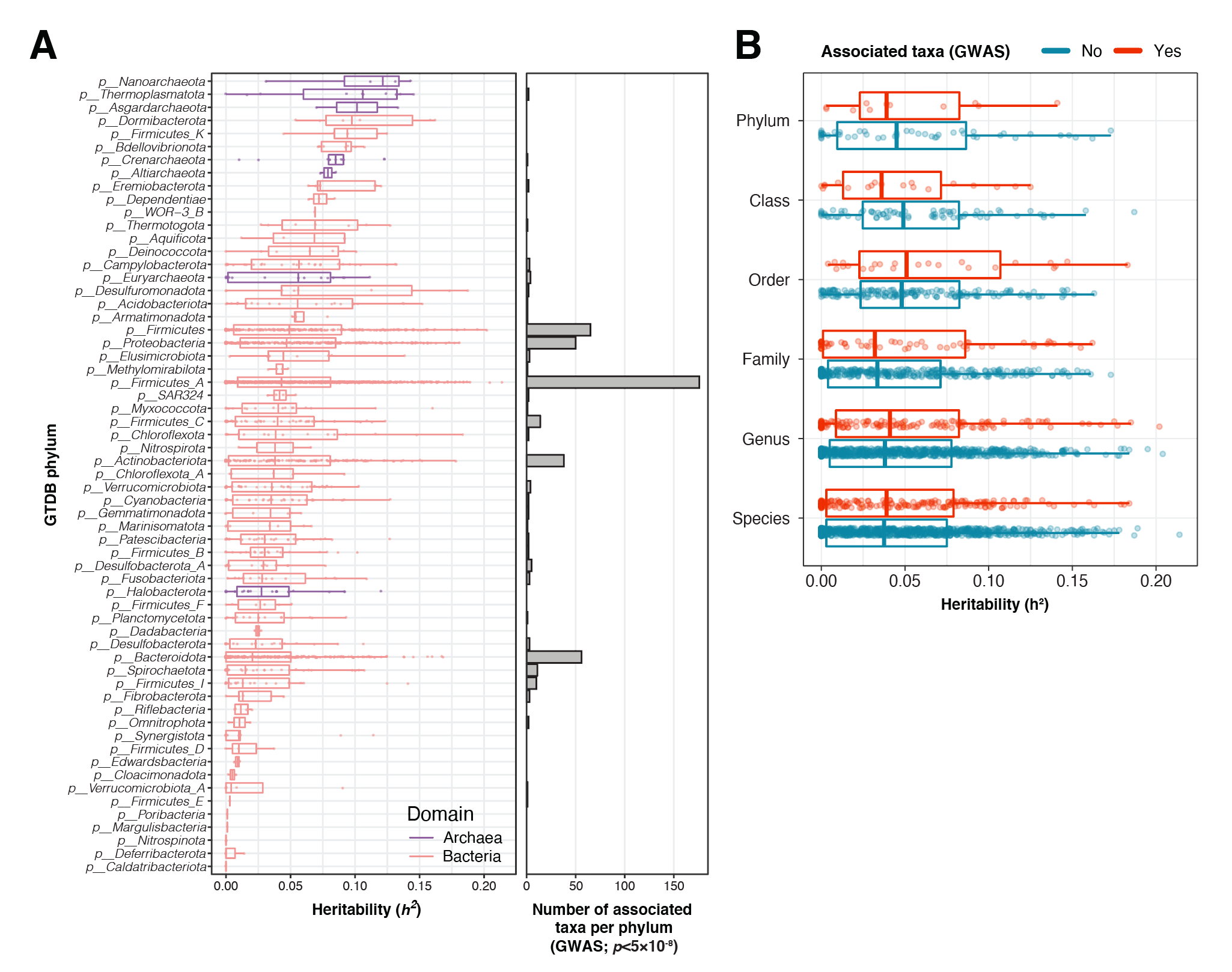
**Supplementary material to Qin et al. (2020)**

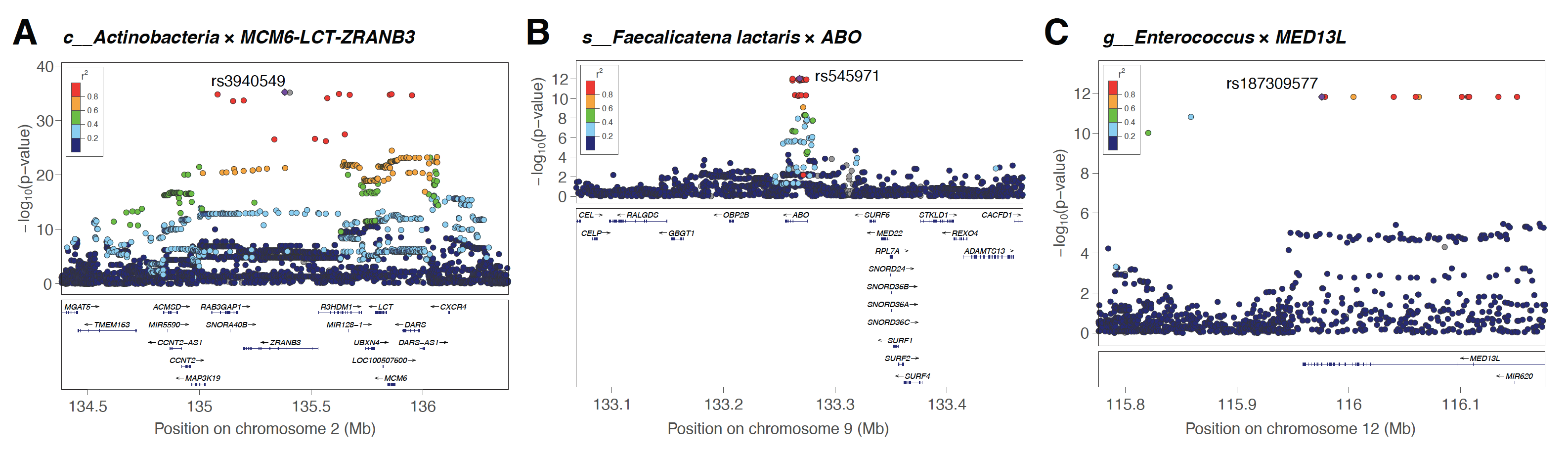
**Figure S1. Study flowchart.**

****

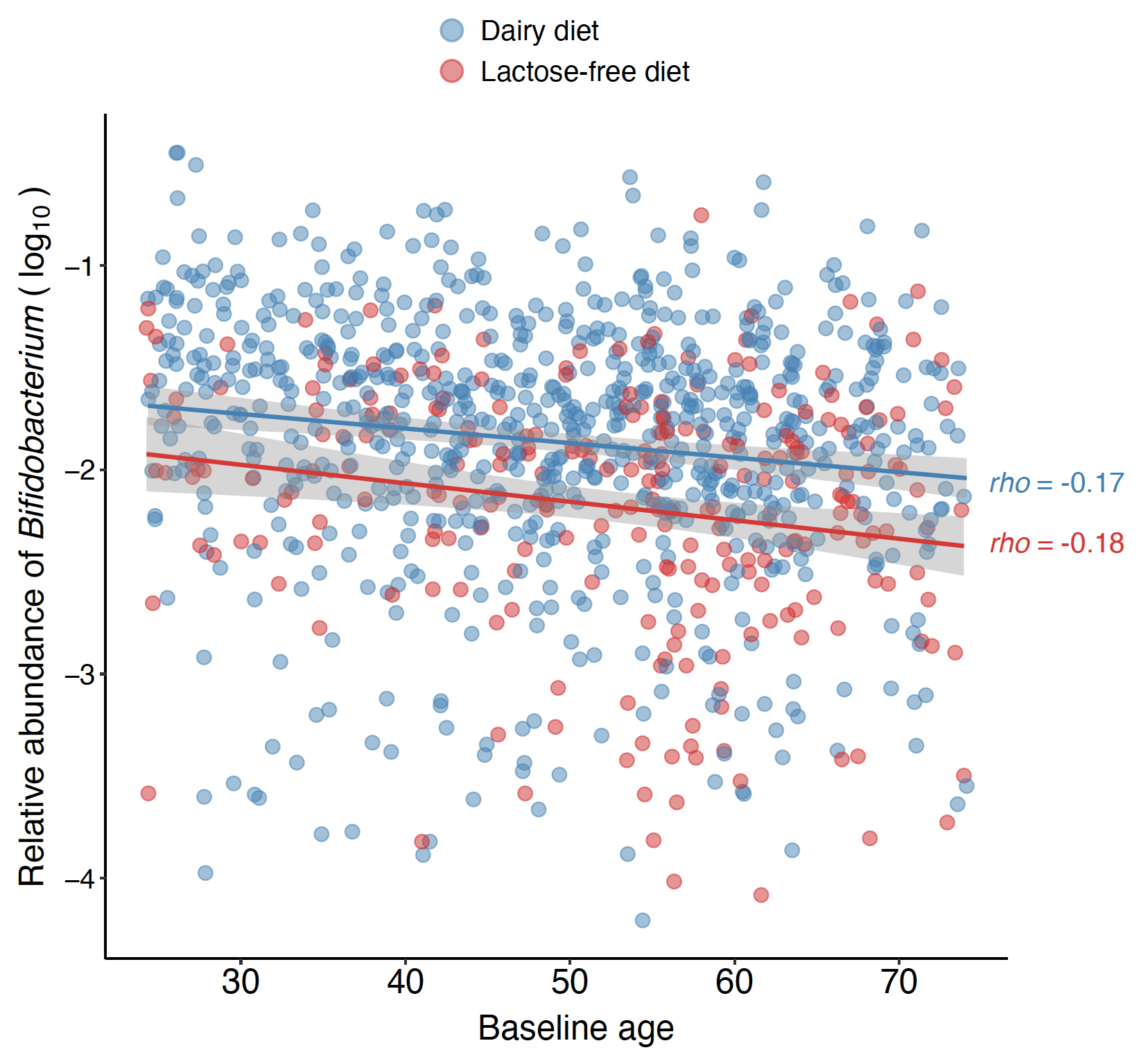
**Figure S2. SNP-microbial taxon heritability.** (A) SNP-taxon heritability (*h2*) for all 2,801 taxa included in the genome-wide association analysis, grouped into their 61 corresponding GTDB phyla, and ordered by median heritability per phylum. Red denotes bacterial phyla, and purple denotes archaeal phyla. The right panel indicates the number of genome-wide significant associated taxa for each phylum (*p*<5×10-8). (B) SNP-taxon heritability is shown for each associated taxon (*p*<5×10-8), grouped by its taxonomic rank.



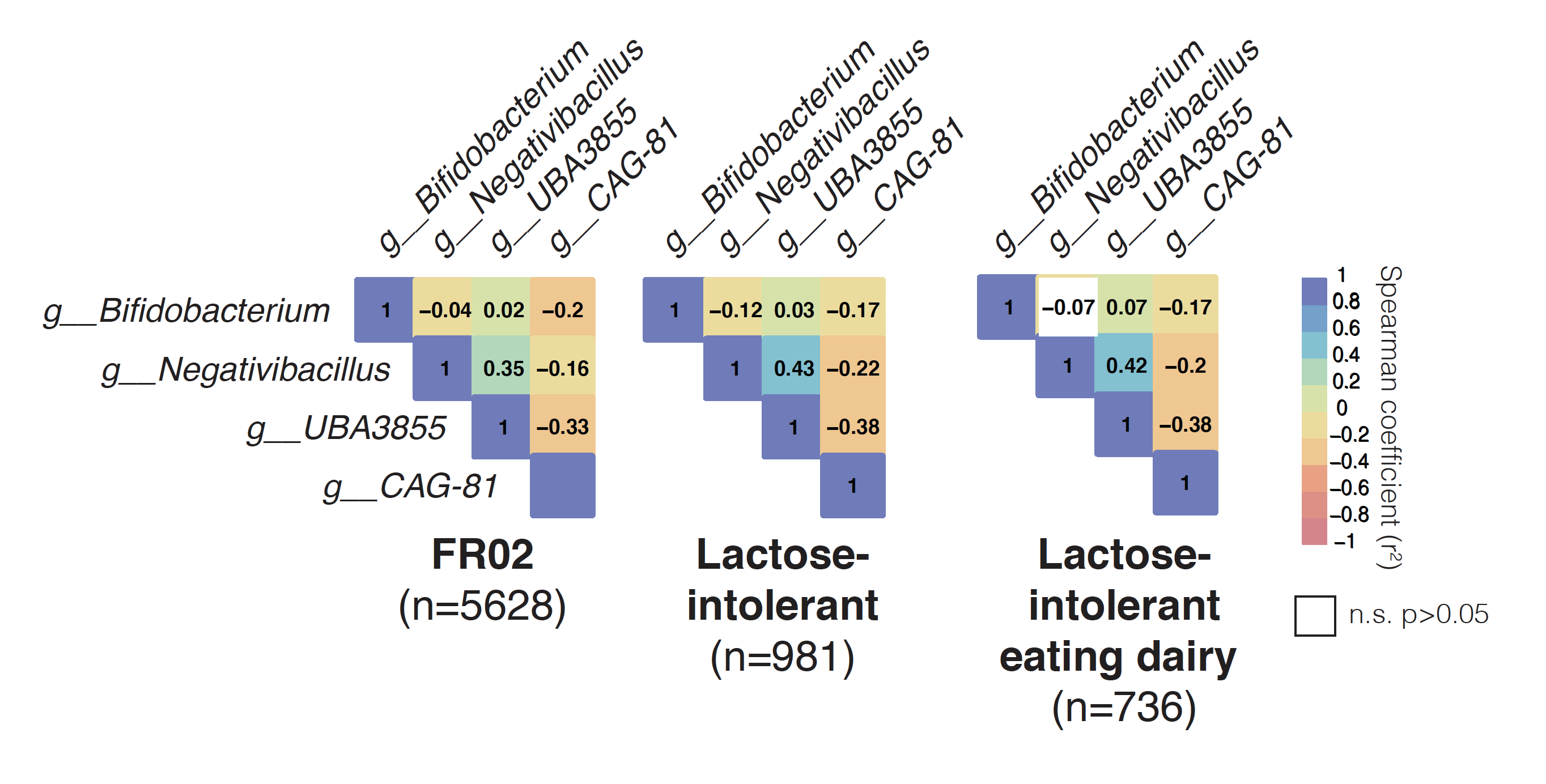
**Figure S3. LocusZoom plots for three loci with study-wide significant associations (*p*<3.8×10-11).** Associations with top taxa are shown. Top SNPs are indicated in purple diamond. Other SNPs are coloured by their linkage disequilibrium (LD) values with the top SNPs. Genes covered by the region are indicated in the bottom. (A) Associated SNPs at the *LCT* locus spans over a 2 Mbp genomic region, while they are grouped on a 400 kbp region for both (B) *ABO* and (C) the *MED13L* loci.

****

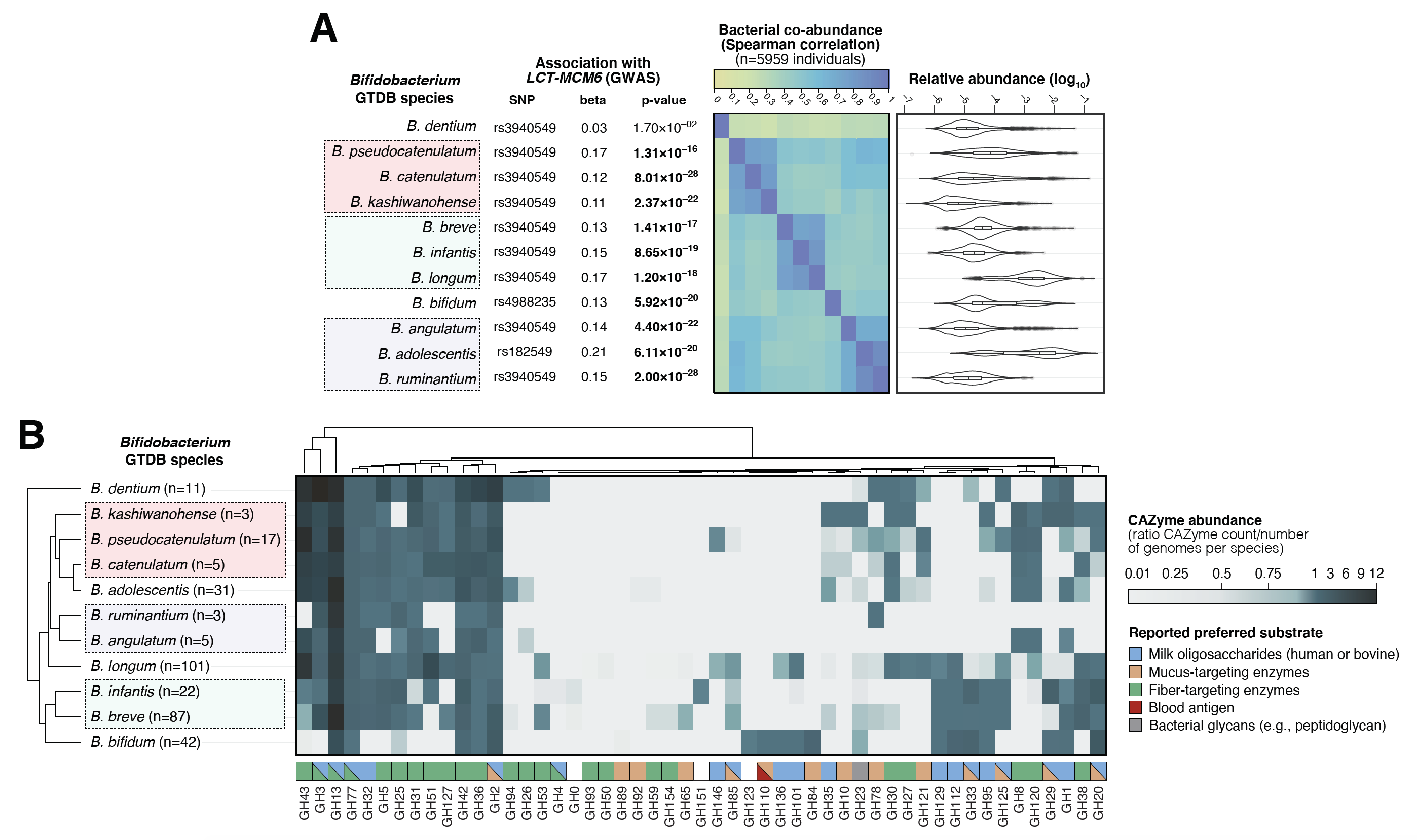
**Figure S4**. **Correlation between individual baseline age and the relative abundance of bacteria from the *Bifidobacterium* genus in lactose intolerant individuals.** Only genetically lactose intolerance individuals (rs4988235:CC) are shown, and coloured by dietary dairy habits (blue: self-reported regular consumption of dairy, n=763; and red: self-reported irregular dairy diet or lactose-free diet, n=253). Best fitted lines and 95% confidence intervals are indicated. Spearman correlation coefficients are given.



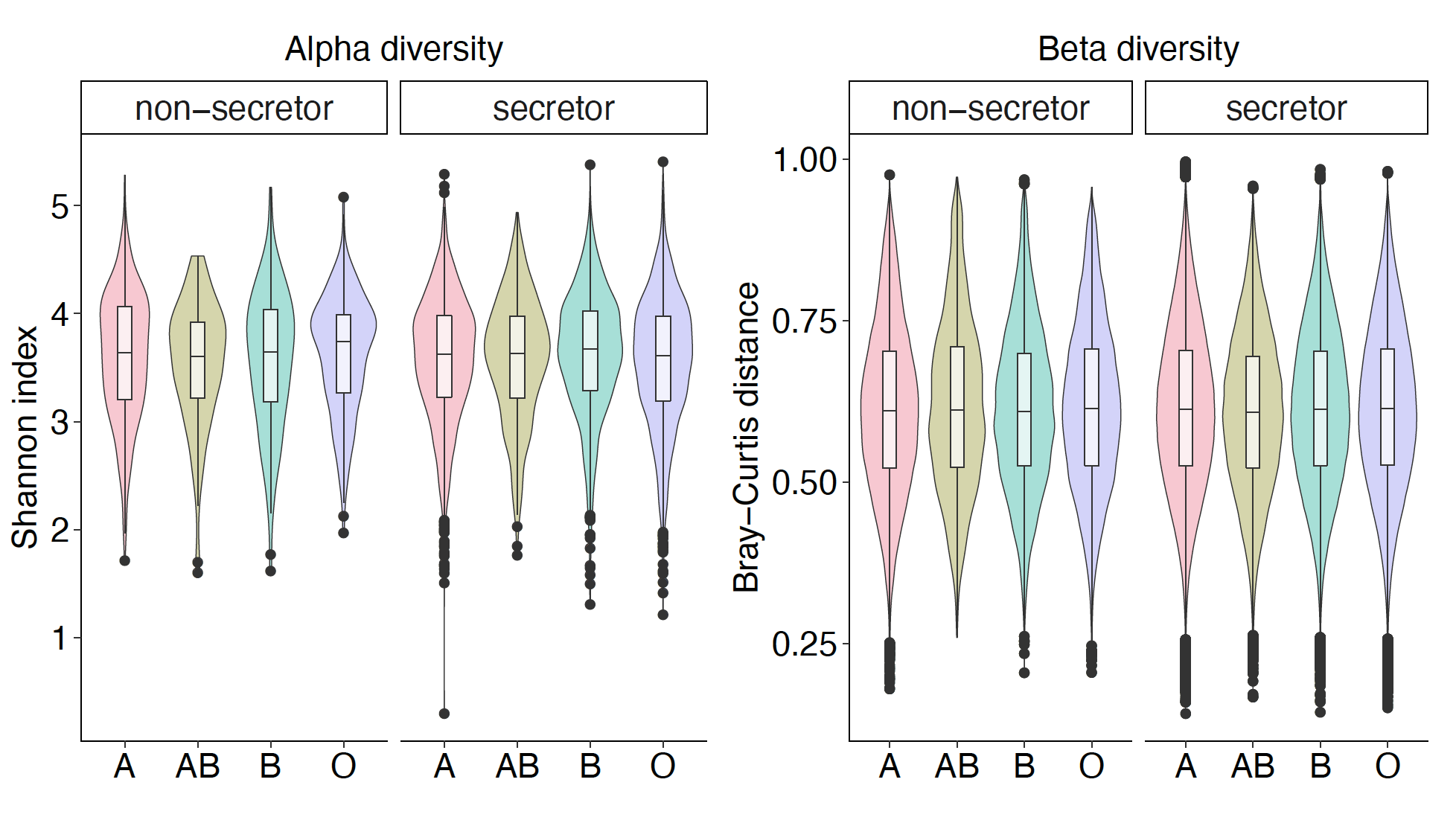
**Figure S5. Spearman correlation of relative abundances in 4 taxa associated with the *LCT* locus.** Abundances of the *Bifidobacterium, Negativibacillus, UBA3855* and *CAG-81* genera are compared. Abundances in the entire FR02 cohort is compared to those in a subset of genetically lactose-intolerant individuals, and to a subset of genetically lactose-intolerant individuals who reported a regular dairy diet. Coloured boxes denote the strength of correlation, while a white square denotes a non-significant p-value for the correlation (*p*>0.05).

****

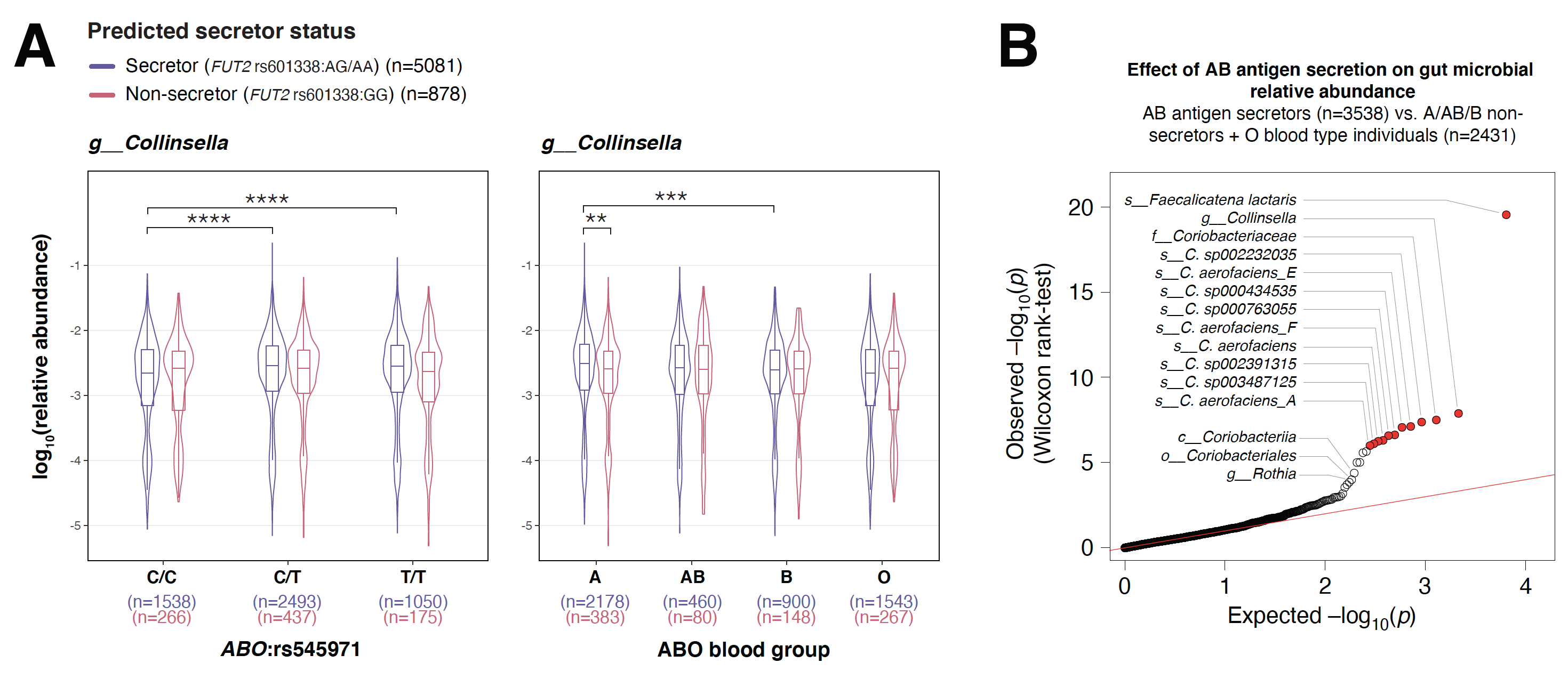
**Figure S6. Co-abundance and carbohydrate-active enzymes (CAZyme) distribution patterns in 11 *Bifidobacterium* species harboured by >25% of individuals in the FR02 cohort.** (A) Associations between the *LCT-MCM6* locus and 11 *Bifidobacterium* species; (left) top association results between variation of 11 *Bifidobacterium* species and the *LCT* locus, with study-wide significant associations (*p*<3.8×10-11) highlighted in bold; (middle) Spearman coefficients calculated on CLR-transformed abundances; (right) relative abundances across the FR02 cohort. (B) Carbohydrate-active enzymes (CAZyme) distribution patterns in 327 previously published reference genomes from 11 *Bifidobacterium* GTDB species which were included in the GTDB release 89 index used to classify metagenomes in this study. The heatmap indicates abundance of corresponding CAZyme families in species, corresponding to the total count of detected families for each species divided by the number of reference genomes examined for the same species. Values <1 indicate that less than one copy per genome of the corresponding CAZyme family was detected for each species, values >1 indicate that more than one copy per genome was detected. Preferred substrate groups are based on literature search and descriptions on CAZypedia.org.

****

**Figure S7. ABO blood type and secretor status are not associated with gut microbial alpha- and beta-diversity measures.** (left)Alpha diversity represented by Shannon indices; (right) beta diversity, represented by Bray-Curtis distances. Alpha and beta diversity were calculated from individual taxonomic profiles at the genus level. Individuals were segregated according to their predicted blood type and secretor status, both predicted from genotype data.

****

**Figure S8. Combined effects of *ABO*:rs545971 genotype and ABO blood group with secretor status on gut levels of *Collinsella* associated with *ABO* variation across participants of the FR02 cohort.** (A)Abundances are compared across stratified groups of individuals from the FR02 cohort according to (left panel): *ABO*:rs545971 genotype and predicted secretor status (blue: secretor status conferred by *FUT2* rs601338:AG/AA genotype; red: non-secretor status conferred by *FUT2* rs601338:GG genotype) and (right panel) according to predicted A, AB, B and O blood types, and predicted secretor status. All statistical comparisons denote the p-values of Wilcoxon rank test on the distributions. *P*-values thresholds are abbreviated as follow: \*:*p*≤0.05; \*\*:*p*≤0.01; \*\*\*:*p*≤0.001; \*\*\*\*:*p*≤0.0001. (B) Effect of AB antigen secretion on gut microbial relative abundance, using the 2,801 taxa considered for GWAS in our study. Taxa with FDR adjusted *p* value <0.05 are highlighted in red. Red line indicates the expected distribution of *p* values under the null hypothesis. P values were calculated using Wilcoxon rank test.

****

**Table S1. Details of 583 genome-wide significant associations between bacterial taxa abundances and genetic variants**. These were independent associations between 464 unique SNPs and 478 microbial taxa. Study-wide significant associations are highlighted in blue. Associated bacterial taxa are following the following nomenclature (d\_\_ for domain, p\_\_ for phylum, c\_\_ for class, o\_\_ for order, f\_\_ for family, g\_\_ for genus and s\_\_ for species). The list of human gut microbiome keystone taxa was taken from Banerjee et al. (2018) *Nature Reviews Microbiology* 16(9):567-576.

**Table S2. Effect of before and after the adjustment of *Bifidobacterium* abundance on GWAS results.** Associations between the *LCT* locus and three *Firmicutes\_A* taxa were compared before and after the adjustment of *Bifidobacterium* abundance. Spearman's rank correlation coefficient (ρ) was calculated using CLR-transformed bacterial abundances.

**Table S3: Details of disease GWAS data used for two-sample Mendelian randomization.** In total, 46 diseases are included. Diseases with significant causal relationships with gut microbes are highlighted in bold.

**Table S4. Predicted causal effects of gut microbial abundances on disease risks using five Mendelian Randomization methods.** These causal effects were supported by at least three out of five tested methods. Red color denotes significant test of causal effect at *p*<0.05 threshold. Horizontal pleiotropy, occurring when a variant would have an effect on disease outside of its effect on the exposure in Mendelian randomization (MR) was tested with MR-PRESSO and is reported for each test.

**Table S5. Predicted causal effects of disease risks on gut microbial abundances using five Mendelian Randomization methods.** These causal effects were supported by at least three out of five tested methods. Red color denotes significant test of causal effect at *p*<0.05 threshold. Horizontal pleiotropy, occurring when a variant would have an effect on disease outside of its effect on the exposure in Mendelian randomization (MR) was tested with MR-PRESSO and is reported for each test.

**Table S6. Summary statistics of genetic instruments used in the Mendelian Randomization.** Only genetic instruments for the inference of the significant causal effects are shown.