**General information**

**GWAS summary statistics of gut microbial traits**

|  |  |
| --- | --- |
| **Genotyping array**: | We performed SNP array-based genotyping using Infinium Asian Screening Array (Illumina, San Diego, CA, USA). This genotyping array was built using an EAS reference panel including whole genome sequences, which enabled effective genotyping in EAS populations. |
| **Sample QC**: | We excluded individuals with a genotyping call rate of < 0.98. All the individuals were estimated to be of EAS ancestry, based on the PCA with the samples of the 1KG dataset using EIGENSTRAT (version 6.1.4). For pairs of closely related individuals (PI\_HAT calculated by PLINK > 0.185), we removed either of the related individuals. |
| **Variant QC**: | We excluded SNPs with (i) call rate < 0.99, (ii) minor allele count < 5, and (iii) P-values for Hardy-Weinberg equilibrium < 1.0 × 10-5. |
| **Phasing and imputation**: | SHAPEIT4 and Minimac4 software |
| **Imputation reference**: | A combined panel of WGS data from the BioBank Japan project (N=1,037) and 1KGP p3v5 ALL (N=2,504). |
| **Post-imputation QC**: | We report results of whole-genome imputed variants with Rsq > 0.7 and MAF > 0.01. |
| **Association test**: | Plink2 was used. Age, age2, sex, sequencing batches, facility, diseases, three genotype-based principal components, and unobserved confounders were regressed out with PEER before the association analysis. |
| **Meta-analysis:** | METASOFT software was used for the meta-analysis. |

**GWAS summary statistics of plasma metabolite traits**

|  |  |
| --- | --- |
| **Whole genome sequencing**: | All libraries were constructed using the TruSeq DNA PCR-Free Library Preparation Kit according to the manufacturer’s protocols. Libraries were sequenced on HiSeqX (Illumina, San Diego, CA, USA) with a mean coverage of 16.4×. Briefly, sequenced reads were aligned against the reference human genome with the decoy sequence (GRCh37, human\_g1k\_v37\_decoy) using BWA-MEM (version 0.7.13). Duplicated reads were removed using Picard MarkDuplicates (version 2.10.10). After Base-quality score recalibration implemented in GATK (versions 3.8-0), we generated individual variant call results using HaplotypeCaller and performed multi-sample joint-calling of the variants via GenotypeGVCFs. |
| **Sample QC**: | For pairs of closely related individuals (PI\_HAT calculated by PLINK > 0.185), we removed either of the related individuals. |
| **Variant QC**: | We set genotypes satisfying any of the following criteria as missing: (i) DP < 5, (ii) GQ < 20, or (iii) DP > 60 and GQ < 95, then removed variants with low genotyping call rates (< 0.90). We performed Variant Quality Score Recalibration for SNVs and short indels according to the GATK Best Practice recommendations and adopted the variants, which passed the QC criteria. We further removed the variants (i) located in the low complexity regions, (ii) with ExcessHet >60, or (iii) with Hardy–Weinberg P-value < 1.0 × 10-10. We kept only those presenting a non-significant difference in allele frequency (*P* > 1.0 × 10-10 provided by chi-square test) in the following representative reference datasets of Japanese ancestry: the combined reference panel of 1KG Phase 3 version 5 genotype (*N*Japanese = 104) and Japanese WGS data (*N* = 1037) used for the aforementioned genotype imputation, and the allele frequency panel of Tohoku Medical Megabank Project (ToMMo 8.3KJPN Allele Frequency Panel, *N* = 8,380). |
| **Genotype refinement:** | Beagle5 software |
| **Post-imputation QC**: | We report results of whole-genome variants with MAF > 0.01. |
| **Association test**: | Plink2 was used. Age, age2, sex, facility, diseases, three genotype-based principal components, and unobserved confounders were regressed out with PEER before the association analysis. |
| **Meta-analysis:** | METASOFT software was used for the meta-analysis. |

**Uploaded files**

**GWAS summary statistics of gut microbial taxa**

|  |  |
| --- | --- |
| File name | Descriptions |
| microbiome\_{TRAITNAME}\_QCed\_sumstats.tsv.gz | Summary results for autosomal variants |

**GWAS summary statistics of gut microbial gene orthologues**

|  |  |
| --- | --- |
| File name | Descriptions |
| KO\_1e\_4\_GWAS.tsv.gz | Summary results for the association between 4,644 gene orthologues and autosomal variants  Associations satisfying *P* < 1×10-4 are reported |

**GWAS summary statistics of gut microbial pathways**

|  |  |
| --- | --- |
| File name | Descriptions |
| KO\_1e\_4\_PATH.tsv.gz | Summary results for the association between 146 gene orthologues and autosomal variants  Associations satisfying *P* < 1×10-4 are reported |

**GWAS summary statistics of plasma metabolites**

|  |  |
| --- | --- |
| File name | Descriptions |
| metabo\_{TRAITNAME}\_QCed\_sumstats.tsv.gz | Summary results for autosomal variants |

**Columns**

|  |  |  |
| --- | --- | --- |
| # | Column name | Description |
| 1 | ID | SNP ID |
| 2 | CHR | chromosome |
| 3 | POS | position (hg19) |
| 4 | OBJ | Phenotype |
| 5 | PVALUE\_FE | P-value in the fixed-effect analysis |
| 6 | BETA\_FE | Effect size of the alternative allele in the fixed-effect analysis |
| 7 | STD\_FE | Standard error of BETA\_FE |
| 8 | PVALUE\_RE | P-value in the random-effect analysis |
| 9 | BETA\_RE | Effect size of the alternative allele in the random-effect analysis |
| 10 | STD\_RE | Standard error of BETA\_RE |
| 11 | PVALUE\_RE2 | P-value in the Han and Eskin’s Random Effects analysis |
| 12 | BETA\_RE2 | Effect size of the alternative allele in the Han and Eskin’s Random Effects analysis |
| 13 | STD\_RE2 | Standard error of BETA\_RE2 |
| 14 | STAT1\_RE2 | Statistic 1 in Han and Eskin’s Random Effects analysis |
| 15 | STAT2\_RE2 | Statistic 2 in Han and Eskin’s Random Effects analysis |
| 16 | I\_SQUARE | heterogeneity index I2 |
| 17 | Q | Cochran's Q statistic |
| 18 | PVALUE\_Q | P value for Q |
| 19 | TAU\_SQUARE | Tau-square heterogeneity estimator |
| 20 | REF | Reference allele |
| 21 | ALT | Alternative allele |
| 22 | ALT\_FREQS | Frequency of the alternative allele |
| 23 | BETA\_SET1 | Effect size of the alternative allele in the dataset 1 |
| 24 | SE\_SET1 | Standard error of the effect size in the dataset 1 |
| 25 | P\_SET1 | P-value in the dataset 1 |
| 26 | ALT\_FREQS\_SET1 | Frequency of the alternative allele in the dataset 1 |
| 27 | BETA\_SET2 | Effect size of the alternative allele in the dataset 2 |
| 28 | SE\_SET2 | Standard error of the effect size in the dataset 2 |
| 29 | P\_SET2 | P-value in the dataset 2 |
| 30 | ALT\_FREQS\_SET2 | Frequency of the alternative allele in dataset 2 |
| 31 | Annotation (only for KEGG gene orthologues and pathways) | Description of the phenotype |