**General information**

**GWAS summary statistics of PUD and its subtypes in BioBank Japan 1st Cohort (BBJ1-180K).**

|  |  |
| --- | --- |
| **Genotyping array**: | The Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips. |
| **Sample QC**: | We excluded samples with low genotyping call rates (call rate < 98%), samples of the estimated East Asian ancestry using PCA, and samples with high heterozygosity rate. |
| **Variant QC**: | We excluded the genotyped variants based on the following criteria for BBJ1-180K: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium P < 1.0 × 10-6, and (4) concordance rate < 99.5% or non-reference discordance rate ≥ 0.5% between array genotypes and whole-genome-sequence dataset using overlapping participants (N = 939). |
| **Phasing and imputation**: | Eagle2 and Minimac4 software |
| **Imputation reference**: | 1KGP p3v5 ALL |
| **Post-imputation QC**: | We here report results of imputed variants with Rsq > 0.3 and MAC>20. |
| **Association test**: | SAIGE software was used with age,　sex, and top 10 principal components as covariates. Leave-one-chromosome-out (LOCO) was applied. |

**GWAS summary statistics of PUD and its subtypes in BioBank Japan 1st Cohort (BBJ1-12K).**

|  |  |
| --- | --- |
| **Genotyping array**: | The Illumina Infinium Asian Screening Array BeadChips |
| **Sample QC**: | We excluded samples with low genotyping call rates (call rate < 98%), samples of the estimated East Asian ancestry using PCA, samples with extreme heterozygosity rate, and samples with amyotrophic lateral sclerosis to its comparatively high proportion. |
| **Variant QC**: | We excluded the genotyped variants based on the following criteria: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium P < 1.0 × 10-6 |
| **Phasing and imputation**: | Eagle2 and Minimac4 software |
| **Imputation reference**: | 1KGP p3v5 ALL |
| **Post-imputation QC**: | We here report results of imputed variants with Rsq > 0.3 and MAC>20. |
| **Association test**: | SAIGE software was used with age,　sex, and top 10 principal components as covariates. Leave-one-chromosome-out (LOCO) was applied. |

**GWAS summary statistics of PUD and its subtypes in BioBank Japan 2nd Cohort (BBJ2-42K).**

|  |  |
| --- | --- |
| **Genotyping array**: | The Illumina Infinium Asian Screening Array BeadChips |
| **Sample QC**: | We excluded samples with low genotyping call rates (call rate < 98%), samples of the estimated East Asian ancestry using PCA, and samples with extreme heterozygosity rate. |
| **Variant QC**: | We excluded the genotyped variants based on the following criteria: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium P < 1.0 × 10-6 |
| **Phasing and imputation**: | Eagle2 and Minimac4 software |
| **Imputation reference**: | 1KGP p3v5 ALL |
| **Post-imputation QC**: | We here report results of imputed variants with Rsq > 0.3 and MAC>20. |
| **Association test**: | SAIGE software was used with age,　sex, and top 10 principal components as covariates. Leave-one-chromosome-out (LOCO) was applied. |

**GWAS summary statistics of PUD and its subtypes in Biobank Japan and Tohoku Medical Megabank (BBJ1-180K, BBJ1-12K, BBJ2-42K and TMM-50K).**

|  |  |
| --- | --- |
| **Genotyping array**: | BBJ1-180K: The Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips.  BBJ1-12k and BBJ2-42K: The Illumina Infinium Asian Screening Array BeadChips  TMM-50K: Axiom Japonica Array JPAv2 (Thermo Fisher Scientific, MA, USA) |
| **Sample QC**: | BBJ1-180K: We excluded samples with low genotyping call rates (call rate < 98%), samples of the estimated East Asian ancestry using PCA, and samples with high heterozygosity rate.  BBJ1-12K: We excluded samples with low genotyping call rates (call rate < 98%), samples of the estimated East Asian ancestry using PCA, samples with extreme heterozygosity rate, and samples with amyotrophic lateral sclerosis to its comparatively high proportion.  BBJ2-42K: We excluded samples with low genotyping call rates (call rate < 98%), samples of estimated East Asian ancestry using PCA, and samples with extreme heterozygosity rate.  TMM-50K: we excluded samples with call rate < 95%, and samples of estimated East Asian ancestry using PCA. |
| **Variant QC**: | BBJ1-180K: We excluded the genotyped variants based on the following criteria for BBJ1-180K: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium P < 1.0 × 10-6, and (4) concordance rate < 99.5% or non-reference discordance rate ≥ 0.5% between array genotypes and whole-genome-sequence dataset using overlapping participants (N = 939).  BBJ1-12K, BBJ2-42K and TMM-50K: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium P < 1.0 × 10-6 |
| **Phasing and imputation**: | Eagle2 and Minimac4 software |
| **Imputation reference**: | 1KGP p3v5 ALL |
| **Post-imputation QC**: | We here report results of imputed variants with Rsq > 0.3 and MAC>20. |
| **Association test**: | SAIGE software was used with age,　sex, and top 10 principal components as covariates. Leave-one-chromosome-out (LOCO) was applied. |
| **Meta-analysis:** | METAL was used to perform fixed-effect inverse variance weighted (IVW) meta-analysis. |

**GWAS summary statistics of PUD and its subtypes in Biobank Japan, Tohoku Medical Megabank, FinnGen and UKB (BBJ1-180K, BBJ1-12K, BBJ2-42K, TMM-50K, FinnGen and UKB).**

Here, we provide summary statistics of the cross-population and multi-trait meta-analysis result of BioBank Japan and UK Biobank.

|  |  |
| --- | --- |
| **Genotyping array**: | BBJ1-180K: The Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips.  BBJ1-12k and BBJ2-42K: The Illumina Infinium Asian Screening Array BeadChips  TMM-50K: Axiom Japonica Array JPAv2 (Thermo Fisher Scientific, MA, USA)  For datasets from UKB and FinnGen, the genotyping arrays were described in PMID: 33608531, PMID: 30104761 and PMID: 36653562. |
| **Sample QC**: | BBJ1-180K: We excluded samples with low genotyping call rates (call rate < 98%), samples of the estimated East Asian ancestry using PCA, and samples with high heterozygosity rate.  BBJ1-12K: We excluded samples with low genotyping call rates (call rate < 98%), samples of the estimated East Asian ancestry using PCA, samples with extreme heterozygosity rate, and samples with amyotrophic lateral sclerosis to its comparatively high proportion.  BBJ2-42K: We excluded samples with low genotyping call rates (call rate < 98%), samples of estimated East Asian ancestry using PCA, and samples with extreme heterozygosity rate.  TMM-50K: we excluded samples with call rate < 95%, and samples of estimated East Asian ancestry using PCA.  For datasets from UKB and FinnGen, the sample QC methods were described in PMID: 33608531, PMID: 30104761, and PMID: 36653562. |
| **Variant QC**: | BBJ1-180K: We excluded the genotyped variants based on the following criteria for BBJ1-180K: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium P < 1.0 × 10-6, and (4) concordance rate < 99.5% or non-reference discordance rate ≥ 0.5% between array genotypes and whole-genome-sequence dataset using overlapping participants (N = 939).  BBJ1-12K, BBJ2-42K and TMM-50K: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium P < 1.0 × 10-6  For datasets from UKB and FinnGen, the variant QC methods were described in PMID: 33608531, PMID: 30104761, and PMID: 36653562. |
| **Phasing and imputation**: | BBJ1-180K , BBJ1-12K, BBJ2-42K and TMM-50K: Eagle2 and Minimac4 software.  For datasets from UKB and FinnGen, the methods were described in PMID: 33608531, PMID: 30104761, and PMID: 36653562. |
| **Post-imputation QC**: | Only autosomal variants in the 1KGp3v5 dataset were included in the meta-analysis; all variants were normalized, duplicate and multiallelic variants were removed for each dataset, and variants with imputation quality scores less than 0.3 were removed. For summary statistics from UKB-SAIGE (imputation quality scores not available), variants included in the previously published GWAS of PUD in UKB were kept, missing information of chromosome and base pair positions were assigned according to rsID, variants with extreme effect size values (|log(OR)| > 10) were removed, variants with minor allele count (MAC) < 5 were removed, and the strand of palindromic variants with MAF < 0.40 was further inferred using the allele frequencies obtained from each population in 1KGp3v5 dataset. Finally, we compared the effect allele frequencies in summary statistics and the population-specific alternative allele frequencies in 1KGp3v5. Variants with deviation in allele frequencies > 0.16 were excluded. |
| **Association test**: | BBJ1-180K, BBJ1-12K, BBJ2-42K, and TMM-50K: SAIGE software was used with age,　sex, and top 10 principal components as covariates. Leave-one-chromosome-out (LOCO) was applied.  For datasets from UKB and FinnGen, the methods were described in PMID: 33608531, PMID: 30104761, and PMID: 36653562. |
| **Meta-analysis:** | METAL was used to perform fixed-effect inverse variance weighted (IVW) meta-analysis. |

**Uploaded files**

**GWAS summary statistics of PUD and its subtypes in BioBank Japan 1st Cohort (BBJ1-180K).**

|  |  |
| --- | --- |
| File name | Descriptions |
| BBJ1\_180K\_{trait}\_1KG.txt.gz | Summary results for autosomal variants |

**GWAS summary statistics of PUD and its subtypes in BioBank Japan 1st Cohort (BBJ1-12K).**

|  |  |
| --- | --- |
| File name | Descriptions |
| BBJ1\_12K\_{trait}\_1KG.txt.gz | Summary results for autosomal variants |

**GWAS summary statistics of PUD and its subtypes in BioBank Japan 2nd Cohort (BBJ2-42K).**

|  |  |
| --- | --- |
| File name | Descriptions |
| BBJ2\_42K\_{trait}\_1KG.txt.gz | Summary results for autosomal variants |

**GWAS summary statistics of PUD and its subtypes in Biobank Japan and Tohoku Medical Megabank (BBJ1-180K, BBJ1-12K, BBJ2-42K, and TMM-50K).**

|  |  |
| --- | --- |
| File name | Descriptions |
| {trait}\_EAS.tsv.gz | Summary results for autosomal variants |

**GWAS summary statistics of PUD and its subtypes in Biobank Japan, Tohoku Medical Megabank, FinnGen and UKB (BBJ1-180K, BBJ1-12K, BBJ2-42K, TMM-50K, FinnGen, and UKB).**

|  |  |
| --- | --- |
| File name | Descriptions |
| {trait}\_EAS\_EUR.tsv.gz | Summary results for autosomal variants |

PUD: peptic ulcer disease (DUonly + GUonly + BU)

DU: duodenal ulcers (DUonly + BU)

DUonly: duodenal ulcers only

GU: gastric ulcers (GUonly + BU)

GUonly: gastric ulcers only

BU: both duodenal and gastric ulcers

**Columns**

Individual trait

BBJ1\_12K\_{trait}\_1KG.txt.gz

BBJ2\_42K\_{trait}\_1KG.txt.gz

BBJ1\_180K\_{trait}\_1KG.txt.gz

| **#** | **column name** | **Descriptions** |
| --- | --- | --- |
| 1 | SNPID | Variant ID  (CHR:POS:reference\_allele: alternative\_allele) |
| 2 | CHR | chromosome |
| 3 | POS | position (hg19) |
| 4 | EA | effect allele |
| 5 | NEA | Non-effect allele |
| 6 | EAF | effect allele frequency |
| 7 | BETA | effect size |
| 8 | SE | standard error |
| 9 | P | P value |
| 10 | INFO | Imputation Rsq |
| 11 | N | Sample size |

EAS-specific Meta-analysis1

| **#** | **column name** | **Descriptions** |
| --- | --- | --- |
| 1 | SNPID | Variant ID  (CHR:POS:reference\_allele: alternative\_allele) |
| 2 | CHR | chromosome |
| 3 | POS | position (hg19) |
| 4 | EA | effect allele |
| 5 | NEA | Non-effect allele |
| 6 | EAF | effect allele frequency |
| 7 | BETA | effect size |
| 8 | SE | standard error |
| 9 | P | P value |
| 10 | Direction | Effect size directions for ach dataset |
| 11 | HetIsq | Heterogeneity I2 |
| 12 | HetPval | Heterogeneity test P value |
| 13 | N | Sample size |

{trait}\_EAS.tsv.gz

Cross-ancestry Meta-analysis

| **#** | **column name** | **Descriptions** |
| --- | --- | --- |
| 1 | rsID | dbSNP rsID |
| 2 | CHR | chromosome |
| 3 | POS | position (hg19) |
| 4 | EA | effect allele |
| 5 | NEA | Non-effect allele |
| 6 | BETA | effect size |
| 7 | SE | standard error |
| 8 | P | P value |
| 9 | N | Sample size |
| 10 | Direction | Effect size directions for each dataset |
| 11 | HetIsq | Heterogeneity I2 |
| 12 | HetPval | Heterogeneity test P value |

{trait}\_EAS\_EUR.tsv.gz