

博士論文

East Asian-specific and cross-ancestry genome-wide meta-analyses provide mechanistic insights into peptic ulcer disease

(東アジアおよび祖先系が異なる集団の横断的なゲノムワイド関連メタ解析による消化性潰瘍の遺伝的メカニズム解明)

He Yunye

賀 云野

**East Asian-specific and cross-ancestry genome-wide meta-analyses
provide mechanistic insights into peptic ulcer disease**

(東アジアおよび祖先系が異なる集団の横断的なゲノムワイド関連メタ解析による消化性潰瘍の遺伝的メカニズム解明)

賀 云野 (He Yunye)

Laboratory of Complex Trait Genomics
Department of Computational Biology and Medical Sciences
Graduate School of Frontier Sciences
The University of Tokyo
2023.6

Table of contents

List of Figures	1
List of Tables	3
Abbreviations	5
Abstract	7
Chapter I - Introduction	8
1. 1 Peptic ulcer disease.....	8
1. 2 Risk factors of Peptic ulcer disease	8
1. 3 Previous GWASs of PUD	8
1. 4 Study Aim.....	8
Chapter II - Methodology.....	10
2. 1 Samples, genotyping, and imputation.....	10
2. 2 Genome-wide association analysis	12
2. 3 Post-GWAS analyses	13
2. 4 GWASLab: a Python package for processing and visualizing GWAS summary statistics	19
2. 5 Ethics approval	22
2. 6 Resources.....	22
2. 7 Code availability.....	22
Chapter III - Results	23
3. 1 The three-stage study design	23
3. 2 Association analyses of PUD and its subtypes in BBJ-180K	25
3. 3 Replication and East Asian-specific meta-analysis.....	29
3. 4 Cross-ancestry meta-analysis, effect size, and genetic correlation comparison.....	34
3. 5 Conditional and fine-mapping analysis	45
3. 6 Overlap of eQTL and pQTL	51
3. 7 Genetic correlation and pleiotropic effects	56
3. 8 H.pylori stratified analysis.....	63
3. 9 Genetic analyses revealed the heterogeneity of GU	65
3. 10 Pleiotropy of PUD risk variants on GC	73
3. 11 Gene-based and gene-set analysis.....	73
3. 12 Tissue- and cell-type specificity analysis.....	81
3. 13 Sensitivity analyses.....	88
Chapter IV - Discussion	95
4. 1 Genetic architecture of PUD across ancestries	95
4. 2 Roles of cell differentiation and hormone regulation in PUD etiology	95
4. 3 The similarities and differences between GU and DU.....	96
4. 4 The correlation between DU and GC	96
4. 5 Limitations of this study	97
4. 6 Summary and prospects.....	97
References	98
Acknowledgments	104

List of Figures

FIGURE 1. OVERVIEW OF GWASLAB PACKAGE DESIGN.	20
FIGURE 2. OVERVIEW OF THE STUDY DESIGN.	23
FIGURE 3. STUDY WORKFLOW.	24
FIGURE 4. VENN DIAGRAM OF CASES USED IN THE DISCOVERY-STAGE GWAS.	24
FIGURE 5. MANHATTAN PLOTS AND Q-Q PLOTS FOR PUD AND ITS SUBTYPES FROM THE DISCOVERY-STAGE GWAS IN BBJ1-180K.	25
FIGURE 6. MANHATTAN PLOTS AND Q-Q PLOTS FOR PUD AND ITS SUBTYPES FROM THE EAST ASIAN-SPECIFIC META-ANALYSIS.	29
FIGURE 7. MANHATTAN PLOT OF THE CROSS-ANCESTRY META-ANALYSIS OF PUD.	34
FIGURE 8. MANHATTAN PLOTS AND Q-Q PLOTS FOR PUD AND ITS SUBTYPES FROM THE CROSS-ANCESTRY META-ANALYSIS.	35
FIGURE 9. PHENOGRAM OF GENOME-WIDE SIGNIFICANT LOCI FOR PUD AND ITS SUBTYPES.	36
FIGURE 10. CROSS-ANCESTRY EFFECT SIZE COMPARISON AND GENETIC CORRELATION ANALYSIS.	37
FIGURE 11. EFFECT SIZE COMPARISON OF LEAD VARIANTS WITH AND WITHOUT THE WINNER’S CURSE CORRECTIONS.	38
FIGURE 12. CROSS-ANCESTRY EFFECT SIZE COMPARISON OF LEAD VARIANTS FOR PUD AND ITS SUBTYPES.	39
FIGURE 13. EAS-SPECIFIC SECONDARY SIGNALS AT <i>PDX1</i> AND <i>JUP</i>	45
FIGURE 14. OVERLAP BETWEEN PUD SIGNALS AND SIGNIFICANT CIS-eQTL VARIANTS OF THE GTEx DATABASE.	52
FIGURE 15. EFFECT SIZE COMPARISON OF DISTINCT VARIANTS AND GENETIC CORRELATIONS ACROSS PUD- RELATED TRAITS IN THE EAST ASIAN POPULATION.	56
FIGURE 16. GENETIC CORRELATION HEATMAP OF PUD AND DIETARY HABITS IN EAST ASIAN POPULATION.	57
FIGURE 17. GENETIC CORRELATION HEATMAP OF PUD AND QUANTITATIVE TRAITS IN EAST ASIAN POPULATION.	57
FIGURE 18. GENETIC CORRELATION HEATMAP OF PUD AND BINARY TRAITS IN EAST ASIAN POPULATION.	58
FIGURE 19. PHEWAS HEATMAP OF PUD RISK VARIANTS WITH ATC CODES AND QUANTITATIVE TRAITS.	58
FIGURE 20. PHEWAS HEATMAP OF PUD RISK VARIANTS WITH BINARY TRAITS (PART 1/2).	59
FIGURE 21. PHEWAS HEATMAP OF PUD RISK VARIANTS WITH BINARY TRAITS (PART 2/2).	60
FIGURE 22. SUMMARY OF SIGNIFICANT ASSOCIATIONS IDENTIFIED IN PHEWAS.	61
FIGURE 23. EFFECT SIZE COMPARISON OF DISTINCT SIGNALS FOR DU AND GU IN EAST ASIAN ANCESTRY INDIVIDUALS.	66
FIGURE 24. EFFECT SIZE COMPARISON OF DISTINCT SIGNALS FOR DUONLY AND GUONLY IN BBJ1-180K.	66
FIGURE 25. EFFECT SIZE COMPARISON OF DISTINCT SIGNALS FOR DU AND GU IN EUROPEAN ANCESTRY INDIVIDUALS.	67
FIGURE 26. CROSS-COHORT EFFECT SIZE COMPARISON OF DISTINCT SIGNALS FOR DU AND GU IN EAST ASIAN ANCESTRY INDIVIDUALS.	68
FIGURE 27. POLYGENICITY ESTIMATION IN EAST ASIANS USING SBAYES.	69
FIGURE 28. POSTERIOR DISTRIBUTION OF POLYGENICITY ESTIMATES FOR GU AND DU IN EAS USING SBAYES.	69
FIGURE 29. STATISTICAL POWER ESTIMATION FOR MENDELIAN RANDOMIZATION ANALYSIS.	73
FIGURE 30. EFFECT SIZE COMPARISON OF DISTINCT SIGNALS BETWEEN PUD AND GC IN EAST ASIAN ANCESTRY INDIVIDUALS.	74
FIGURE 31. TISSUE- AND CELL-TYPE SPECIFICITY ANALYSIS.	81

FIGURE 32. CELL-TYPE SPECIFICITY ANALYSIS IN EAST ASIAN ANCESTRY INDIVIDUALS USING LDSC.....	83
FIGURE 33. CELL-TYPE SPECIFICITY ANALYSIS IN EAST ASIAN ANCESTRY INDIVIDUALS USING MAGMA.....	83
FIGURE 34. CELL-TYPE SPECIFICITY ANALYSIS IN EUROPEAN ANCESTRY INDIVIDUALS USING LDSC.....	84
FIGURE 35. CELL-TYPE SPECIFICITY ANALYSIS IN EUROPEAN ANCESTRY INDIVIDUALS USING MAGMA.	84
FIGURE 36. CROSS-ANCESTRY META-ANALYSIS OF CELL-TYPE SPECIFICITY USING LDSC.....	85
FIGURE 37. CROSS-ANCESTRY META-ANALYSIS OF CELL-TYPE SPECIFICITY USING MAGMA.	85
FIGURE 38. VENN PLOT OF THE POTENTIAL SAMPLE OVERLAP WITHIN BIOBANK JAPAN COHORTS.....	88
FIGURE 39. MANHATTAN PLOTS AND Q-Q PLOTS FOR EAS-SPECIFIC META-ANALYSIS OF PUD USING DIFFERENT MODELS.....	89
FIGURE 40. MANHATTAN PLOTS AND Q-Q PLOTS FOR CROSS-ANCESTRY META-ANALYSES BASED ON DIFFERENT VERSIONS OF THE REFERENCE GENOME.....	90
FIGURE 41. COMPARISON OF $-\log_{10}(P)$ FOR SNPs IN GWASS BASED ON DIFFERENT VERSIONS OF THE REFERENCE GENOME.	90
FIGURE 42. POWER ANALYSIS FOR GWAS OF HP-NEGATIVE PUD IN TMM-50K.....	91
FIGURE 43. COMPARISON OF MAGMA RESULTS USING DIFFERENT WINDOW SIZES AROUND GENES FOR GENE- BASED ANALYSIS.	92
FIGURE 44. COMPARISON OF MAGMA RESULTS USING DIFFERENT WINDOW SIZES AROUND GENES FOR TISSUE- SPECIFICITY ANALYSIS.....	93
FIGURE 45. COMPARISON OF MAGMA RESULTS USING DIFFERENT WINDOW SIZES AROUND GENES FOR CELL- TYPE-SPECIFICITY ANALYSIS.....	94

List of Tables

TABLE 1. SUMMARY OF COHORTS AND PHENOTYPES ANALYZED IN THE GENOME-WIDE ASSOCIATION ANALYSIS IN THIS STUDY.	11
TABLE 2. VARIANTS WITH SIGNIFICANT NON-RANDOM MISSINGNESS IN BBJ COHORTS.....	11
TABLE 3. CHARACTERISTICS OF PARTICIPANTS IN BBJ1-180K, BBJ1-12K, BBJ2-42K, AND TMM-50K.....	26
TABLE 4. SIGNIFICANT LOCI ASSOCIATED WITH PUD OR PUD SUBTYPES IN DISCOVERY DISCOVER-STAGE GWAS DISCOVER DISCOVER-STAGE GWAS.....	27
TABLE 5. SIGNIFICANT LOCI ASSOCIATED WITH PUD OR PUD SUBTYPES FROM SEX-STRATIFIED AND CHR X ANALYSES IN BBJ1-180K.....	28
TABLE 6. REPLICATION OF NOVEL VARIANTS IDENTIFIED IN DISCOVERY STAGE GWAS IN BBJ1-12K, BBJ2-42K, AND TMM-50K.....	30
TABLE 7. ESTIMATIONS OF CONFOUNDING BIAS AND SNP-HERITABILITY BY LD SCORE REGRESSION.	31
TABLE 8. SIGNIFICANT LOCI ASSOCIATED WITH PUD OR PUD SUBTYPES IN EAS-SPECIFIC META-ANALYSIS.	32
TABLE 9. SIGNIFICANT LOCI ASSOCIATED WITH PUD OR PUD SUBTYPES FROM GENOME-WIDE META-ANALYSES.	33
TABLE 10. SUMMARY OF PUBLICLY AVAILABLE SUMMARY STATISTICS USED IN META-ANALYSES AND EFFECT SIZE COMPARISONS.	40
TABLE 11. SIGNIFICANT LOCI ASSOCIATED WITH PUD OR PUD SUBTYPES FROM INVERSE-VARIANCE-WEIGHTED CROSS-ANCESTRY META-ANALYSIS.	41
TABLE 12. SIGNIFICANT LOCI ASSOCIATED WITH PUD OR PUD SUBTYPES FROM CROSS-ANCESTRY META- REGRESSION USING MR-MEGA.....	42
TABLE 13. EFFECT COMPARISON OF PUD SIGNALS BETWEEN EAST ASIANS AND EUROPEANS.....	43
TABLE 14. CROSS-ANCESTRY GENETIC CORRELATION OF PUD AND PUD SUBTYPES.....	44
TABLE 15. INDEPENDENT VARIANTS FROM COJO ANALYSIS IN EAST ASIANS.....	47
TABLE 16. NONSYNONYMOUS VARIANTS IN CREDIBLE SETS FROM FINE-MAPPING ANALYSIS USING SUSIE IN EAST ASIANS.....	48
TABLE 17. LOGISTIC REGRESSION ANALYSIS FOR ASSOCIATION OF ABO BLOOD GROUP AND FUT2 SECRETOR STATUS WITH PEPTIC ULCERS.....	49
TABLE 18. LOGISTIC REGRESSION ANALYSIS FOR ABO-FUT2 INTERACTIONS.....	50
TABLE 19. OVERLAP OF LEAD VARIANTS ASSOCIATED WITH PUD AND PUD SUBTYPES IN EAS WITH PQTL SIGNALS.....	53
TABLE 20. SIGNIFICANT PROTEOMIC ASSOCIATIONS WITH ABO BLOOD AND FUT2 SECRETOR STATUS IN CONCORDANT DIRECTIONS.....	55
TABLE 21. GENETIC CORRELATION AMONG PUD, PUD-RELATED PHENOTYPES, AND ITS RISK FACTORS.....	62
TABLE 22. H.PYLORI INFECTION STATUS OF PARTICIPANTS IN TMM-50K.	63
TABLE 23. COLOCALIZATION ANALYSIS FOR HP-STRATIFIED ANALYSIS.	64
TABLE 24. EFFECT SIZE COMPARISON OF PUD SIGNALS BETWEEN DU AND GU IN EAST ASIANS.	70
TABLE 25. POLYGENICITY ESTIMATION USING SBAYESS IN EAST ASIANS.....	71
TABLE 26. ASSOCIATIONS OF POLYGENIC RISK SCORES WITH PUD AND PUD SUBTYPES IN BBJ1-180K.	72
TABLE 27. TWO-SAMPLE MENDELIAN RANDOMIZATION STUDY OF THE CAUSALITY OF PUD OR ITS SUBTYPES ON GC.....	75
TABLE 28. TESTS FOR DIRECTIONAL PLEIOTROPY AND HETEROGENEITY IN MR.	76

TABLE 29. OUTLIER-CORRECTED MR USING THE MR-PRESSO METHOD.	77
TABLE 30. EFFECT SIZE COMPARISON OF PUD SIGNALS BETWEEN DU AND GC IN EAST ASIANS.....	78
TABLE 31. GENE-BASED ANALYSIS FOR PUD AND PUD SUBTYPES IN EAST ASIAN INDIVIDUALS.	79
TABLE 32. GENE-SET ENRICHMENT ANALYSIS FOR PUD AND PUD SUBTYPES IN EAST ASIAN INDIVIDUALS.	80
TABLE 33. CELL TYPE SPECIFICITY ANALYSIS USING LDSC.....	86
TABLE 34. CELL TYPE SPECIFICITY ANALYSIS USING MAGMA.	87

Abbreviations

Abbreviation	Description
1KG	1000 Genome Project
BBJ	BioBank Japan
BU	Both duodenal and gastric Ulcers
CADD	Combined Annotation Dependent Depletion
COJO	COnditional & JOint association analysis
CS	Credible Sets
D cell	(somatostatin-producing) Delta cells
DU	Duodenal ulcers
EAS	East Asians
EC cell	Enterochromaffin cell
eQTL	expression QTL
EUR	Europeans
FUMA	Functional Mapping and Annotation of Genome-Wide Association Studies
GC	Gastric Cancers
GTE _x	Genotype-Tissue Expression project
GU	Gastric Ulcers
GWAMA	Genome-Wide Association Meta-Analysis
GWAS	Genome-Wide Association Studies
HLA	Human Leukocyte Antigen
HP	Helicobacter Pylori
IVW	Inverse Variance Weighted
JPT	Japanese in Tokyo
LD	Linkage Disequilibrium
LIA	Latex Agglutination Immunoassay
LOCO	Leave One Chromosome Out
MAC	Minor allele count
MAF	Minor allele frequency
MAGMA	Multi-marker Analysis of GenoMic Annotation
MCMC	Markov chain Monte Carlo
MD	Major depression
MHC	Major Histocompatibility Complex
MR	Mendelian randomization
MR-PRESSO	Mendelian Randomization Pleiotropy RESidual Sum and Outlier
NSAID	Non-steroidal anti-inflammatory drugs
OR	Odds ratio
pQTL	protein QTL
PRS	Polygenic Risk Scores
PUD	Peptic ulcer disease
QC	Quality control

Abbreviation	Description
QTL	Quantitative trait locus
SAIGE	Scalable and Accurate Implementation of GEneralized mixed model
scRNA-seq	Single-cell RNA sequencing
SNP	Single-nucleotide polymorphism
SuSiE	Sum of Single Effects
TMM	Tohoku Medical Megabank
UKB	UK Biobank
WC	Winner's curse

Abstract

Peptic ulcer disease (PUD) refers to the acid-induced injury of the digestive tract, mainly occurring in the stomach (Gastric ulcer; GU) or duodenum (Duodenal ulcer; DU). Despite several risk loci for PUD identified in previous genome-wide association studies (GWAS), the underlying genetics largely remain unclear. A large-scale cross-ancestry meta-analysis was conducted for PUD GWAS with four Japanese and two European studies (52,032 cases and 905,344 controls), discovering 25 novel, independent loci that were highly concordant across ancestries. GU shared the same risk loci with DU but showed smaller genetic effect sizes than DU, indicating higher heterogeneity of GU. Downstream analyses highlighted plausible biologies, such as blood coagulation by integrating expression and protein quantitative trait locus analyses. Moreover, the gene-level analysis showed that genetic factors are enriched in highly expressed genes in stomach tissue, especially in somatostatin-producing D cells and EC cells from stomach and duodenum single-cell RNA sequencing datasets. Variants in *EFNA1*, *PTGER4*, and *PSCA* showed strong but opposite effects on DU and gastric cancers (GC), suggesting the potential mechanisms contributing to the decreased risk of GC in DU patients. *H.pylori* (HP)-stratified analysis found HP-positive-specific host genetic locus at CCKBR, marking its role in HP-mediated PUD etiology. In summary, this study revealed potential genes and provided evidence that cell differentiation and hormone regulation are involved in genetic susceptibility to PUD and its subtypes.

Chapter I - Introduction

1. 1 Peptic ulcer disease

Peptic ulcer disease (PUD), referring to the acid-induced injury of digestive tract, is one of the most common gastrointestinal disorders, with bleeding, perforation, or gastric outlet obstruction as the major complication. The major symptom of PUD is upper abdominal pain while chronic peptic ulcers can be asymptomatic. PUD can be diagnosed with upper gastrointestinal endoscopy or barium x-ray. Conventionally, based on the location where the ulcer occurred, PUD can be categorized into two major subtypes: gastric ulcers (GU; stomach) and duodenal ulcers (DU; duodenum).

The lifetime prevalence rate of PUD has been estimated to be 5-10% in the general population¹. Epidemiological studies showed substantial geographical variations in terms of the prevalence of PUD and *H. pylori* (HP) infection^{1,2}. It has been reported that the prevalence of PUD is substantially higher in East Asians (EAS) than in Europeans (EUR), and GU is more common in the Japanese population whereas DU is more common in Europeans³. In terms of sex difference in PUD risk, a previous epidemiological study has also shown that both the number of prevalent cases and age-standardized prevalence rate (ASR) were higher in males than in females¹.

1. 2 Risk factors of Peptic ulcer disease

The most well-known causes of PUD include HP infection and the use of non-steroidal anti-inflammatory drugs (NSAIDs)⁴. HP infection is observed in up to 90% of individuals with duodenal ulcers and in around 70% of patients with gastric ulcers⁵. Previous studies have shown that genetic factors also play an important role in PUD etiology. A study of Finnish twins reported that around 40% of the liability to peptic ulcer disease was explained by genetic factors⁶. In addition, smoking⁷, alcohol drinking⁸, and depression⁹ have been reported to be associated with an increased risk of peptic ulcer. Epidemiological studies have also suggested that DU is a protective factor against gastric cancers (GC)¹⁰. However, whether genetic factors for PUD and GC are concordant and can explain the epidemiological findings is still unclear.

1. 3 Previous GWASs of PUD

Genome-wide association study (GWAS) has been shown to be a powerful approach to identifying risk loci associated with complex traits¹¹. Previous GWASs have identified several risk loci associated with PUD or its subtypes, mostly in Europeans. A previous study in a Japanese population identified 2 risk loci for duodenal ulcers³; a study in UK Biobank (UKB) identified 6 novel loci for PUD¹² and a recent study in a Japanese population identified an additional locus for gastric ulcers¹³.

1. 4 Study Aim

Given the relatively high prevalence of PUD and HP infection in EAS and the limited number of risk loci identified in EAS, genome-wide association studies (GWAS) with a larger sample size would be necessary to improve our understanding of the genetic etiology of PUD. Despite the progress in revealing the genetic architecture for peptic ulcer disease, there are still remaining issues to be addressed. First, the host genetic factors that interact with HP are still largely unknown; second, differences across PUD subtypes should be investigated given that GU and DU differ in various aspects; third, the key cell types

contributing to PUD etiology should be investigated; and finally, the difference in genetic architecture across ancestries has not been systematically investigated.

This study aims to address the above-mentioned issues by conducting the largest-to-date EAS-specific and cross-ancestry analysis of PUD and its subtypes, along with more than 52,022 PUD cases and 905,344 controls from four Japanese studies and two European cohorts.

Chapter II - Methodology

2. 1 Samples, genotyping, and imputation

Cohorts and samples

The datasets used in this study were obtained from two of the largest biobanks in Japan, BioBank Japan Project¹⁴ and Tohoku Medical Megabank Project¹⁵. BioBank Japan (BBJ, <https://biobankjp.org/en/>) Project was one of the largest hospital-based biobanks around the world, which was founded in 2003 and consisted of two cohorts. BioBank Japan first cohort (denoted as BBJ1) enrolled more than 199,998 patients of mainly Japanese ancestry with at least one of 47 target complex diseases from 2003 to 2008, with collaborations from 66 participating hospitals (12 medical institutions) across Japan. BioBank Japan second cohort (denoted as BBJ2) additionally recruited more than 67,334 new individuals with at least one of 38 target complex diseases from 2013 to 2017, with support from 52 participating hospitals (12 institutions) across Japan. Tohoku Medical Megabank Project was established in 2011 and involves two main cohorts: a family-based cohort and a population-based cohort. The population-based cohort (the Tohoku Medical Megabank Project Community-Based Cohort Study, The TMM CommCohort Study) recruited more than 84,000 participants from 2013 to 2017.

Genotype data of the samples employed in the discovery-stage GWAS was obtained from the primary dataset of BBJ1, which includes 181,927 patients (denoted as BBJ1-180K). Clinical information including age and sex for BBJ1-180K was obtained from the participating hospitals. Replication was performed in three independent Japanese cohorts with available genotype data and clinical information: an additional and independent cohort of 11,715 patients from BBJ1 (denoted as BBJ1-12K), an independent cohort from BBJ2 of 42,689 individuals (denoted as BBJ2-42K) and a population-based cohort of 49,621 samples from the Tohoku Medical Megabank Project (denoted as TMM-50K).

In this study, samples of age ≥ 18 were included. Principal component analysis (PCA) using samples from the 1000 Genomes Project¹⁶ was conducted and then all samples were projected onto the same space. Outliers from the East Asian cluster were excluded.

Phenotype definition

In this study, we assessed peptic ulcer disease and its subtypes including gastric ulcers and duodenal ulcers. Clinical information for GU and DU cases was retrieved from interviews and reviews of medical records using a standardized questionnaire in the three BBJ cohorts and by self-administered questionnaires in TMM-50K¹⁵. PUD cases were obtained from the combination of DU and GU. Patients with both duodenal and gastric ulcers were additionally categorized into a group BU (short for both ulcers). Individuals without a given diagnosis of peptic ulcers or any HP-related diseases (Gastric cancers) were used as control samples (Table 3). The cohorts and phenotypes that were first reported in this study were summarized in Table 1.

Table 1. Summary of cohorts and phenotypes analyzed in the genome-wide association analysis in this study.

Cohort	Target trait	First reported in this study	Datasets used in this study	Reference	Major differences with reference
BBJ1-180K	PUD	TRUE	Analyzed independently	/	/
BBJ1-180K	DU	FALSE	Re-analyzed independently	Tanikawa et al. PMID 22387998	a. ~127,000 additional controls b. 132 additional cases c. Imputation panel
BBJ1-180K	GU	FALSE	Re-analyzed independently	Sakaue et al. PMID 34594039	a. Imputation panel
BBJ1-180K	BU	TRUE	Analyzed independently	/	/
BBJ1-12K	PUD	TRUE	Analyzed independently	/	/
BBJ1-12K	DU	TRUE	Analyzed independently	/	/
BBJ1-12K	GU	TRUE	Analyzed independently	/	/
BBJ1-12K	BU	TRUE	Analyzed independently	/	/
BBJ1-42K	PUD	TRUE	Analyzed independently	/	/
BBJ1-42K	DU	TRUE	Analyzed independently	/	/
BBJ1-42K	GU	TRUE	Analyzed independently	/	/
BBJ1-42K	BU	TRUE	Analyzed independently	/	/
TMM-50K	PUD	TRUE	Analyzed independently	/	/
TMM-50K	DU	TRUE	Analyzed independently	/	/
TMM-50K	GU	TRUE	Analyzed independently	/	/
TMM-50K	BU	TRUE	Analyzed independently	/	/
TMM-50K	HP-stratified	TRUE	Analyzed independently	/	/
UKB	PUD	FALSE	Publicly available summary statistics	Wu et al. PMID 33608531	/
UKB	GU	FALSE	Publicly available summary statistics	Zhou et al. PMID 30104761	/
UKB	DU	FALSE	Publicly available summary statistics	Zhou et al. PMID 30104761	/
FinnGen	PUD	FALSE	Publicly available summary statistics	Kurki et al. PMID 36653562	/
FinnGen	GU	FALSE	Publicly available summary statistics	Kurki et al. PMID 36653562	/
FinnGen	DU	FALSE	Publicly available summary statistics	Kurki et al. PMID 36653562	/

Genotyping and quality control

Samples from BBJ1-180K were genotyped with either Illumina HumanOmniExpressExome BeadChips or a combination of Illumina HumanOmniExpress and HumanExome BeadChips; samples from the BBJ1-12K and BBJ2-42K cohorts were genotyped with Infinium Asian Screening Array BeadChips. Samples from the TMM-50K cohort were genotyped with Axiom Japonica Array JPAv2 (Thermo Fisher Scientific, MA, USA).

In the BBJ1-180K dataset, samples with call rates <98% and excess heterozygosity (4 standard deviations (SD) from the mean) were excluded. QC of autosomal genotypes was performed as described¹⁷. In brief, genotyped variants were excluded based on the following criteria: (1) call rate < 99%, (2) heterozygote count < 5, (3) Hardy–Weinberg-equilibrium $P < 1.0 \times 10^{-6}$, and (4) concordance rate < 99.5% or non-reference discordance rate $\geq 0.5\%$ between array genotypes and whole-genome-sequence dataset using overlapping participants ($n = 939$), as described previously¹⁷. The same sample and variant QC criteria (except (4) for autosomal genotype QC) as in the discovery stage to the replication sets of BBJ1-12K and BBJ2-42K. Additionally, samples with extreme heterozygosity rate (± 4 SD from the mean) in BBJ1-12K and BBJ2-42K were removed. Samples with amyotrophic lateral sclerosis in BBJ1-12K were excluded due to the comparatively high proportion. Non-random missingness in cases and controls was tested in BBJ1-180K, BBJ1-12K, and BBJ2-42K (Table 2). In the replication using TMM-50K, samples with call rate < 95% were excluded, and variants with (1) call rate < 99%, (2) heterozygote count < 5, or (3) Hardy–Weinberg-equilibrium $P < 1.0 \times 10^{-6}$ were excluded. For QC of variants on chromosome X, genotyped variants with (1) call rate < 99% in males, females, or both, and (2) Hardy–Weinberg-equilibrium $P < 1.0 \times 10^{-6}$ in females were excluded.

Table 2. Variants with significant non-random missingness in BBJ cohorts.

Cohort	CHR	rsID	Missing rate in cases	Missing rate in controls	Fisher's exact test P
BBJ1-180K	16	rs4258597	0.00137	0.000222	1.04E-10

Phasing and imputation

Pre-phasing was performed using Eagle2 (v2.4.1)¹⁸ and imputation was performed with Minimac4 (v1.0.2) using the 1000 Genomes Project Phase 3¹¹ version 5 (1KGp3v5) ALL panel. Additionally, BBJ1-12K and BBJ2-42K were imputed with the 1KG high-coverage reference panel¹⁹ (GRCh38). For chromosome X, haplotypes were pre-phased for males and females, and variants were imputed separately in males and females using the same software described above. Imputed variants with $R_{sq} > 0.3$ were included. Around 13 million variants were included in the discovery-stage association analysis.

2. 2 Genome-wide association analysis

Discovery-stage genome-wide association analysis

Single variant association test was performed using SAIGE(v0.44)²⁰. SAIGE is a statistical method implementing a generalized mixed model with SPA correction controlling for case-control imbalance and cryptic relatedness. Age, sex, and top 10 PCs were used as covariates in the regression. For step 1 of SAIGE, LD-pruned genotyped variants (PLINK²¹ --indep-pairwise 50 5 0.2) with $MAF > 1\%$ were used to estimate the null models with leave one chromosome out (LOCO). Variants with $MAC < 20$ were excluded from the association tests for step 2. For sex-stratified analysis, association analyses were performed with SAIGE adjusting for the same set of covariates other than sex in males and females.

Variants of the X chromosome were tested separately in males and females using the corresponding null models estimated by autosomes in each sex. Haploid-based dosages of the non-pseudo autosomal (non-PAR) region of males were multiplied by 2. The statistics for each sex were then meta-analyzed using inverse-variance weighted methods implemented in METAL²².

In this study, genome-wide significant loci were determined by iteratively extending 500-kb flanking regions around the most significant variant until no genome-wide significant variant ($P < 5.0 \times 10^{-8}$) was detected within the extended regions. The lead variant was defined as the most significant variant in each locus. Loci for different traits with lead variants within 500 kb of each other were considered the same, denoted by the most significant lead variant from the locus. Significant variants in the major histocompatibility complex (MHC) region (chromosome 6: 25–34 Mb) were counted as one locus due to complexity of the region.

Replication and Meta-analysis in East Asians

The directions and effect sizes were compared with the replication GWAS sets for the lead variants of significant loci identified in the discovery-stage GWAS. The statistics of GWAS at the discovery and replication stages were combined using the fixed-effect inverse-variance method implemented in METAL. Heterogeneity was estimated by Cochran's Q test. We considered the lead variants identified in discovery-stage GWAS as replicated if the variants reached the nominal significance threshold ($P_{rep} < 0.05$) in the same direction in at least two of the replication GWAS.

Cross-ancestry meta-analysis

Summary statistics of PUD and PUD subtypes for individuals of European ancestry were obtained from a published study of PUD in UKB¹² (<https://cnsgenomics.com/content/data>) and publicly available

databases of FinnGen²³ (release 6 for PUD, DU, and GU; https://www.finnngen.fi/en/access_results) and PheWeb UKB-SAIGE²⁰ (DU and GU; <https://pheweb.org/UKB-SAIGE/>).

For summary statistics that were not based on GRCh37, the genome coordinates of summary statistics were converted to GRCh37 (hg19) using the UCSC LiftOver tool²⁴. Additional quality control and harmonization were performed for all summary statistics prior to meta-analyses. The following criteria were applied: (1) only autosomal variants in the 1KGp3v5 dataset were included in the meta-analysis; (2) all variants were normalized²⁵, (3) duplicate and multiallelic variants were removed for each dataset, and (4) variants with imputation quality scores less than 0.3 were removed.

For summary statistics from UKB-SAIGE (imputation quality scores not available), (1) variants included in the previously published GWAS of PUD in UKB (which was imputed using the Haplotype Reference Consortium (HRC) and UK10K as reference) were kept, (2) missing information of chromosome and base pair positions were assigned according to rsID, (3) variants with extreme effect size values ($|\log(\text{OR})| > 10$) were removed, (4) variants with minor allele count (MAC) < 5 were removed, (5) and the strand of palindromic variants with $\text{MAF} < 0.40$ was further inferred using the allele frequencies obtained from each population in 1KGp3v5 dataset. Lastly, the effect allele frequencies in summary statistics and the population-specific alternative allele frequencies in 1KGp3v5 were compared, and variants with deviation in allele frequencies > 0.16 were excluded. In total, more than 19 million variants were included in the meta-analysis.

The cross-ancestry meta-analyses integrating GWAS results in EAS and EUR populations were conducted using the fixed-effect inverse-variance method implemented in METAL²². EUR-specific summary statistics were generated by fixed-effect meta-analyses combining results from UKB and FinnGen. The effect sizes of lead variants identified in cross-ancestry meta-analyses were compared between EAS and EUR populations using the population-specific meta-analyses.

Moreover, MR-MEGA²⁶ (v0.2) was used to perform cross-ancestry meta-regression with four axes of genetic variation derived via multi-dimensional scaling. P values were recalculated using chi-squared statistic due to the lack of support in MR-MEGA for $P < 1.0 \times 10^{-14}$.

2. 3 Post-GWAS analyses

LD score regression

LD score regression (v1.0.0)²⁷ was performed to estimate the bias caused by confounding factors including population stratification or cryptic relatedness. The LD scores provided by authors for the East Asian population were utilized, which were estimated from 1000 Genomes Project EAS individuals. For the conversion of observed-scale heritability to liability-scale heritability, the population lifetime prevalence rates in East Asian populations were set to 6.2%, 6.9%, 10.8%, and 1.8% for DU, GU, PUD, and BU, respectively. The prevalence rates were estimated from the population-based TMM-50K and were similar to those in previous epidemiological studies.

Genetic correlation estimation

Cross-trait LDSC (v1.0.0)²⁸ with LD scores estimated from 1KG EAS individuals were used to evaluate the genetic correlation between PUD and common binary and quantitative traits in East Asian populations.

East Asian summary statistics were obtained from previous GWAS conducted in BBJ. The MHC region was excluded from this analysis.

Popcorn (v1.0)²⁹ was utilized with pre-computed cross-ancestry LD scores estimated from 1KG EUR and EAS populations to estimate the cross-ancestry correlations of genetic effect for PUD and subtypes between EAS and EUR. For these analyses, meta-analyzed summary statistics of PUD in EAS and EUR for HapMap 3 SNPs (without the MHC region) were used.

Conditional analysis and fine-mapping

Conditional analysis was performed using GCTA-COJO (v1.93.2)³⁰ in each significant locus identified in EAS-specific meta-analysis to test for secondary signals. An LD reference panel was constructed using the best-guess imputed genotype of 20,000 randomly selected and unrelated individuals of East Asian ancestry from BBJ1-180K. Stepwise model selection was conducted first to select independent association signals ($P < 5.0 \times 10^{-8}$), and a joint analysis of these selected signals was performed next. Variants with MAF > 0.01 were included in the analysis.

Fine-mapping was conducted using SuSiE (v0.11.92)³¹ with default configurations while allowing ten putative causal variants within each locus. Unrelated individuals (KING kingship coefficient value³² < 0.0884 , $N = 171,085$) from BBJ1-180K were used as LD reference, computed by LDstore (v2.0)³³ based on the imputed dosages. The regions were defined based on the 3 Mb window centered at the lead variants and merged if the window overlapped. Only variants with $Rsq \geq 0.5$ were included in fine-mapping. The credible sets (CS) were reported, which have a 95% probability of harboring one causal variant.

Variant annotation

Variants identified in this study were annotated using ANNOVAR³⁴ (v2020-06-07; -protocol refGene,avsnp150,clinvar_20200316). Chromatin states (core 15-state model) for stomach mucosa and duodenum mucosa were obtained from the Roadmap Epigenomics Project³⁵. Allele frequencies for variants not available in the population-specific meta-analysis were obtained from the Genome Aggregation Database (gnomAD)³⁶ or the 1000 Genomes Project if not available in gnomAD.

Blood group and secretor status interaction analysis

In this study, the interaction between blood group and secretor status was investigated using samples in BBJ1-180K. ABO blood groups were inferred based on genotyped variants and the secretor status was inferred based on imputed variants. Specifically, ABO blood groups for 164,613 unrelated individuals used in discovery-stage GWAS in BBJ1-180K (KING kingship coefficient < 0.0884) were inferred from two genotyped variants (rs505922 and rs8176746) described previously³. Secretor status was inferred using the best-guess genotype of imputed variants rs1047781 (p.Ile140Phe), where AA or AT genotypes are secretors, and TT are non-secretors³⁷.

Logistic regression analyses adjusting for age, sex, and top 10 PCs were performed to examine the association of blood group or secretor status with PUD or subtypes. Blood group-specific effect sizes were estimated using the target blood group as exposure and the combination of the other three blood groups as non-exposure (for example, individuals with A vs. B, AB, and O). Secretor status effect sizes were estimated considering the non-secretor status as exposure and the secretor status as non-exposure.

To investigate the interaction of blood group O with non-secretor status, logistic regression analyses were further performed for blood group O–secretor interaction, adjusting for O blood group, FUT2 secretor status, age, sex, and top 10 PCs. Similarly, the blood group O–non-secretor status interaction was tested using imputed dosages of the variants determining O antigen and secretor status. All of the above-mentioned logistic regressions were conducted in R v4.1.0. Additionally, a previous pQTL study was explored, which had investigated proteomic associations with ABO blood groups and FUT2 secretor status for proteins associated with secretor status and A, B, and AB blood groups.

H.pylori stratified analysis in TMM-50K

To investigate the interaction of *H. pylori* (HP) with host genetic factors on the risk for PUD, we performed HP-stratified analyses in HP-positive and HP-negative individuals in TMM-50K due to the availability of HP infection information in TMM (Clinical information for HP infection was not available in BBJ for most participants).

HP infection status was determined by anti-HP serum IgG antibody titer, which was measured by the latex agglutination immunoassay (LIA). Individuals with anti-HP serum IgG antibody titer ≥ 10 U/mL were categorized as HP-positive and Individuals with anti-HP serum IgG antibody titer < 10 U/mL were categorized as HP-negative.

Genome-wide association tests were performed with the same settings as in the discovery-stage GWAS. Cochran's Q and I^2 statistics (calculated by R package metafor³⁸ v3.4, <https://www.metafor-project.org/doku.php>) were used to test the effect size heterogeneity between HP-positive and HP-negative GWAS for each subtype.

To investigate if the signals for HP-positive PUD and HP-negative PUD colocalize, colocalization analysis was conducted using coloc package³⁹ (v5.1.0; <https://chr1swallace.github.io/coloc/>) for each significant locus identified in EAS meta-analysis under a single causal variant assumption. For loci with multiple independent signals identified in the conditional analysis, coloc was applied to the signals identified by SuSiE⁴⁰.

Two-sample Mendelian randomization

To evaluate the causality of PUD or its subtypes on GC, two-sample Mendelian randomization (MR) analysis was performed using TwoSampleMR R package⁴¹ in the study. An additional meta-analysis combining GWASs for PUD and its subtypes in BBJ1-12K, BBJ2-42K, and TMM-50K was performed to avoid sample overlap between the exposure (PUD or its subtypes) and outcome (GC). A total of 23 available independent variants identified in the EAS-specific meta-analysis were used as instrumental variables for MR. To avoid bias caused by weak instruments, per-variant F statistics were estimated for each exposure, and variants with $F < 10$ were removed from the instrumental variables for the exposure. To correct for the horizontal pleiotropic variants, MR-PRESSO⁴² (Mendelian Randomization Pleiotropy RESidual Sum and Outlier) was employed which is a method that detects horizontal pleiotropic outliers in multi-instrument summary-level MR by analyzing the observed and expected distributions of the variables. The summary statistics for GC were obtained from the previous study conducted in BBJ1-180K⁴³. Statistical power for MR was approximately estimated using methods described previously⁴⁴. Compared with MR using all available variants, the pleiotropic outlier correction may result in

insufficient power (Figure 29).

Polygenic risk estimation

Polygenic risk score (PRS) models for PUD and its subtypes were constructed in East Asians using the summary statistics derived from replication GWAS and HP-stratified analysis in the population-based Japanese cohort TMM-50K. PRS-CS⁴⁵ (v2021-Jun-4, <https://github.com/getian107/PRSCs>) were employed to compute PRS models using HapMap3 SNPs with an EAS-specific LD reference panel from the 1KG EAS individuals. Global shrinkage parameters were obtained from the data by PRS-CS using a fully Bayesian approach (PRS-CS-auto). The models were applied to individuals from BBJ1-180K and the associations of PRS with PUD or its subtypes were tested using logistic regression adjusted for age, sex, and top five PCs. The predictive ability of each PRS model was evaluated by its improvement of AUC and R2 on the liability scale⁴⁶ over a base model that includes age, sex, and top five PCs.

Phenome-wide association analysis

To investigate whether variants associated with PUD or its subtypes were also associated with other binary and quantitative traits in East Asians, summary statistics of the non-overlapping lead variants and secondary signals for 215 case-control and quantitative traits were obtained from BioBank Japan PheWeb (<https://pheweb.jp/>). LD proxy ($r^2 > 0.6$) with the highest r^2 estimated from 1KG EAS was used if a variant was unavailable in the datasets. The significance threshold was set to $P < 8.6 \times 10^{-6}$ after the Bonferroni correction.

Effect size comparison

For the pair-wise effect size (logarithm of odds ratios) comparison among PUD, PUD subtypes, and gastric cancers in the EAS population, the non-overlapping (interval between adjacent variants > 500 kb) lead variants identified in EAS-specific meta-analysis and the independent signals identified in COJO analysis were selected. For loci associated with two or more phenotypes, the most significant associations (lead variants with the lowest P-value) for comparison. For PUD and its subtypes, Effect sizes in EAS-specific meta-analysis were used; for GC, summary statistics were obtained from a previous study in BBJ1-180K. For cross-ancestry comparison of variant effect sizes, all non-overlapping lead variants associated with PUD or any subtypes in population-specific meta-analysis or cross-ancestry meta-analysis were included. Associations, with the lowest P values, of loci associated with more than one phenotype, were selected for comparison, and effect sizes in the population-specific meta-analysis were used. Cochran's Q test was used to test heterogeneity across the effect sizes. Additionally, we compared the winner's curse (WC)-corrected effect sizes of the lead variants identified in the EAS meta-analysis in the current study and the variants reported in UKB¹² using the methods as described previously^{47,48}.

Polygenicity estimation

To estimate the polygenicity (defined as the proportion of SNPs with nonzero effects), and the strength of negative selection (defined as the relationship between MAF and effect sizes, and denoted by S) for PUD, SBayesS⁴⁹ from GCTB software (v2.0) was used. Briefly, SBayesS employs a Bayesian mixed linear model and reports the posterior means of SNP-based heritability, polygenicity estimates, and a metric that indicates negative selection.

An LD reference panel for EAS was constructed using the approach described previously. Shortly, a full

LD matrix on HapMap3 SNPs was computed using 50,000 randomly selected and unrelated East Asian individuals from BBJ1-180K. The off-diagonal entries of full LD matrix were shrunk with the interpolated genetic map for the 1000 Genomes Project JPT population (<https://github.com/joepickrell/1000-genomes-genetic-maps>). The effective population size and genetic map sample size were set to 11,600 and 100, respectively, according to 1000 Genomes Project phased OMNI data (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_rates/). The sparse shrunk LD matrix used for SBayesS was created by setting elements of the shrunk matrix to zero if their chi-squared statistic under the sampling distribution of the correlation coefficient did not exceed 10.

For a cross-trait comparison of polygenicity estimates in EAS, 42 registered binary traits of BBJ1 were used in the analysis. The MHC region was excluded from the analysis. Four parallel MCMC chains were run with a length of 50,000 and a burn-in size of 20,000 for each trait. To evaluate the convergence in MCMC, potential scale reduction statistics for each parameter were computed and traits with potential scale reduction statistic < 1.2 for all three parameters, including SNP-based heritability, polygenicity estimates, and S, were considered to have good convergence as recommended by the author, and were, therefore, used in the analysis.

Gene-based and gene-set-based analysis

Gene-based and gene-set-based analyses were performed using MAGMA⁵⁰ (v1.08; <https://ctg.cncr.nl/software/magma>) implemented in FUMA⁵¹ (v1.3.8; <https://fuma.ctglab.nl/>). An LD reference panel constructed from 1000 Genomes EAS population was used. For the gene-based analysis, a total of 19,033 protein-coding genes (ENSEMBL v92⁵²) were tested for association. Statistics of the gene-based analysis were then utilized for gene-set enrichment analysis with a total of 15,485 curated gene sets and GO terms from MsigDB v7.0.

Tissue and cell-type specificity analysis

MAGMA gene-property analysis implemented in the SNP2GENE method of FUMA was employed for tissue-type specificity analysis with gene expression profile from GTEx v8 dataset. A total of 54 non-diseased tissue types and 30 general tissue types were tested. Tissue types with a false discovery rate (FDR) $< 5\%$ were considered significant.

To identify the cell types associated with PUD in the stomach and duodenum, processed single-cell RNA sequencing datasets of the human stomach and duodenum were obtained from a previous study⁵³ (Table S6. Human gastric cell types and Table S11. Human duodenal cell types), which filtered for cells with more than 1,500 transcripts per cell and genes expressed by at least three transcripts in at least one cell. A total of 13,980 genes for 19 cell types in the stomach and 17 cell types in the duodenum were included in the study. The top 10% most specifically expressed genes based on fold-change (defined as the average transcript counts of all cell types except the target cell type divided by the average transcript counts of the target cell type) were extracted for each cell type. SNPs in cell-type-specific genes were used to compute partitioned LD scores in 1KG phase 3 EAS or EUR population. The gene coordinates were extended by a window size of 100 kb to capture the effects of regulatory elements. Stratified LD score regressions were performed using the partitioned LD scores of cell-type-specific genes, partitioned LD

scores of all available genes in the dataset, and the baseline model of 53 annotations for each ancestry on HapMap3 SNPs excluding the MHC region (downloaded from <https://alkesgroup.broadinstitute.org/LDSCORE/>).

Gene-set enrichment analysis using MAGMA was performed with the cell-type-specific gene sets described above. The 1000 Genomes Project phase 3 EAS and EUR population datasets were used as reference panels. Variants with MAF < 0.01 or in the MHC region were excluded from the analysis. The gene coordinates were extended by window sizes of 35 kb upstream and 10 kb downstream. IVW meta-analysis was performed using statistics of both ancestries for each method to increase statistical power. P values were calculated using the one-tailed test. Cell types with FDR < 5% within each expression dataset were considered significant.

eQTL and pQTL analysis

To characterize the effect of variants on gene expression level, LD proxies (with $r^2 > 0.6$ in EAS or EUR 1KG Phase 3) with the lead variants and secondary signals in EAS were extracted. Only significant SNP-gene pairs with FDR < 5% (pre-computed by the authors) from the GTEx version were extracted. The overlap between the lead variants and secondary signals (including proxies) and cis-eQTL variants in GTEx v8 were examined. The most significant cis-eQTL association for each gene in each tissue was selected for interpretation.

To characterize the effect of variants on protein level, LD proxies (with $r^2 > 0.6$ in EAS or EUR 1KG Phase 3) with the lead variants and secondary signals in EAS were extracted. Genome-wide significant SNP-protein associations from five published large-scale pQTL studies in recent years were extracted⁵⁴⁻⁵⁸. The five studies were conducted on individuals of mainly European ancestry. The overlap between the lead variants (including LD proxies) with cis- and trans-pQTL were then checked. The most significant association for each protein was selected for interpretation.

2. 4 GWASLab: a Python package for processing and visualizing GWAS summary statistics

Based on the scripts used for the visualization and harmonization of summary statistics in this study, a Python (GWASLab) was developed for processing and visualizing GWAS summary statistics.

Introduction

GWAS summary statistics (SumStats) are accumulating at a rapid speed. As of April 2023, SumStats for more than 6,000 publications are publicly available on GWAS Catalog ⁵⁹. The availability of SumStats potentiates a wide range of post-GWAS analyses, such as LD Score regression ²⁷, genome-wide meta-analysis, Mendelian randomization, and polygenic risk scores. SumStats sharing greatly enhanced research in genetics, and it was estimated that SumStats sharing led to around 75% more citations ⁶⁰.

Despite the efforts to develop a standard SumStats format ⁶¹, the large number of existing unprocessed SumStats remain challenges for data sharing and efficient reuse. Additionally, extensive discrepancies exist between the formats GWAS software generates and the required formats for each post-GWAS software. The missingness of certain information (such as rsID), the inconsistent representation of statistics, and the ambiguity of column headers (such as the headers for effect allele and non-effect allele) often hinder the direct reuse of SumStats, especially for beginners, and were error-prone during data and format conversions without careful reading the manual. Furthermore, unexpected failure in processing SumStats can lead to adverse impacts on downstream analyses, which could be avoided by checking detailed log messages.

Existing tools for handling or visualizing GWAS SumStats, mostly implemented in R ⁶²⁻⁶⁴, focused on specific functionalities such as data munging or plotting. With the rapid increase in publicly available SumStats and post-GWAS tools, there is also an increasing need for a comprehensive and customizable tool that integrates functions for handling and visualizing GWAS summary statistics, which can serve as a bridge linking unprocessed SumStats to post-GWAS analysis tools seamlessly. GWASLab, a user-friendly Python package for the manipulation and visualization of GWAS summary statistics, was developed. This package provides functions including quality control (QC) of statistics, standardization of chromosome and allele notation, variant normalization, harmonization for meta-analysis, and data visualization. A summary statistics format conversion library along with standalone utilities was also developed, which ensure seamless compatibility with a wide range of post-GWAS tools. Moreover, the implementation of a logging system provided detailed reports on each process applied, which increased the interpretability and the reusability of SumStats.

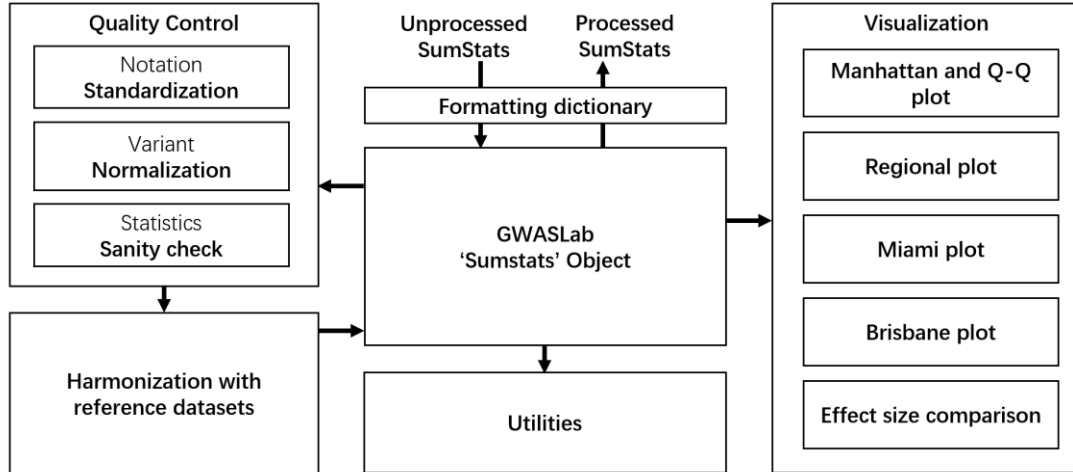


Figure 1. Overview of GWASLab package design.

Q-Q plot, quantile-quantile plot. SumStats, Summary statistics. Miami plot, mirrored Manhattan plots for two traits. Brisbane plot, Manhattan-like plot showing the genomic density of independent genetic associations.

Implementation

GWASLab was designed based on the following principles: (1) to remove ambiguity and minimize uncertainty in SumStats by specifying column definition, standardizing notations, and aligning with reference data; (2) to provide flexible, customizable, and interpretable functions for each manipulation so that users can customize their pipelines when the pre-defined pipeline does not suit for their SumStats; (3) to support detailed logging for every step of manipulation so that the pipeline is traceable and replicable; and (4) to provide compatibility and extensibility with existing tools and future ones.

GWASLab has been developed as an open-source Python package. Sumstats will be formatted and stored in the 'GWASLab.Sumstats' object. Functions to manipulate and visualize the summary statistics were implemented as methods of the 'GWASLab.Sumstats' object, enabling the construction of interpretable, flexible, and customizable SumStats processing pipelines.

GWASLab is composed of five main function groups (Fig.1): (1) SumStats formatting; (2) standardization and quality control without references, (3) harmonization with references, (4) visualization, and (5) standalone utilities. This package takes advantage of commonly used Python packages and can be readily integrated into any Python-based downstream analysis tools or pipelines. Additionally, a logging system was implemented along with an original status code system to log the manipulations applied to the SumStats and trace the status of each variant.

Columns of SumStats data in GWASLab consist of two parts, core columns, and additional columns. Core columns comprise key information for variants and statistics generated from GWAS, including rsID, SNPID (preferably in the format of chromosome: position: reference allele: alternative allele), CHR (chromosome number), POS (base-pair position), EA (effect allele), NEA (non-effect allele), EAF (effect allele frequency), BETA (effect size), SE (standard error of effect size), N (sample size), OR (odds ratio), OR_95L (lower bound of 95% confidence interval), OR_95U (upper bound of 95% confidence interval),

P (p-value), MLOG10P (-log10(P-value)), Z (z score), CHISQ (chi-square statistic), DIRECTIONS (effect size directions). Additional columns consist of optional information for the variants such as annotation, providing extensibility for formatting or customized filtering.

Main Usage

1. Seamlessly importing and formatting files

Using a curated and expandable format conversion library ('formatbook'), GWASLab can read SumStats generated by widely used GWAS software such as PLINK ²¹, SAIGE ²⁰ and REGENIE ⁶⁵ and format SumStats to software-specific or widely accepted formats in line with sharing standards such as GWAS VCF ^{66,67} and GWAS-SSF ⁶¹. Users can also load customized formats by explicitly specifying the columns.

2. Standardization, normalization, quality control, and harmonization

GWASLab provides functions to standardize the notations and data types of variant ID, chromosomes, base pair positions, and alleles. After standardization, variants will be checked for normalization to ensure they are left aligned and parsimonious ²⁵. For statistics, GWASLab can filter extreme values, remove missing or duplicated records, and perform sanity checks as implemented in existing tools ^{62,68}. GWASLab provides functions for SumStats harmonization with references, which include allele alignment with a reference genome, converting genome coordinates (liftover) using appropriate chain file, strand checking of palindromic SNPs, and annotating rsID using reference files downloaded from commonly used sources such as 1000 Genomes Project ¹⁶ and dbSNP ⁶⁹, or self-prepared files for quick annotation.

3. Highly customized visualization and other utilities

Visualization functions were integrated into the GWASLab framework. GWASLab can create Q-Q plots, Manhattan plots, Miami plots ⁷⁰, Regional plots, and Brisbane plots ⁷¹. GWASLab also supports effect size comparison, data conversion, lead variant extraction, and position-based novel loci determination ⁷². For example, users can specify an EFO ID ⁷³, and GWASLab will extract known associations for the trait through GWAS Catalog API ⁵⁹, and compare the base pair distances between known loci and lead variants in user-provided SumStats, allowing users to check whether the known loci are potentially novel based on distances.

2. 5 Ethics approval

The research project was approved by the ethics committees at the Institute of Medical Science, the University of Tokyo (application number 29-74-A0215), and Iwate Tohoku Medical Megabank Organization, Iwate Medical University (application number HG H25-2).

2. 6 Resources

- UKB-SAIGE: <https://pheweb.org/UKB-SAIGE/>
- PUD GWAS in UKB: <https://cnsgenomics.com/content/data>
- FinnGen: https://www.finnngen.fi/en/access_results
- PheWeb.jp: <https://pheweb.jp/>
- JENGER: <http://jenger.riken.jp/result>
- The 1000 Genomes Project phased OMNI data: http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_rates/
- The interpolated genetic map for the 1000 Genomes Project: <https://github.com/joepickrell/1000-genomes-genetic-maps>

2. 7 Code availability

- PLINK1.9: <https://www.cog-genomics.org/plink/>
- PLINK2: <https://www.cog-genomics.org/plink/2.0/>
- MAGMA: <https://ctg.cncr.nl/software/magma>
- FUMA: <https://fuma.ctglab.nl>
- ANNOVAR: <https://annovar.openbioinformatics.org/en/latest/>
- LDSC: <https://github.com/bulik/ldsc>
- LiftOver: <https://genome.ucsc.edu/cgi-bin/hgLiftOver>
- SAIGE: <https://github.com/weizhouUMICH/SAIGE>
- METAL: <http://csg.sph.umich.edu/abecasis/Metal/index.html>
- GWAMA: <https://genomics.ut.ee/en/tools>
- MR-MEGA: <https://genomics.ut.ee/en/tools>
- GCTA-COJO: <https://yanglab.westlake.edu.cn/software/gcta/#COJO>
- GCTB: <https://cnsgenomics.com/software/gctb/#Overview>
- Coloc: <https://chr1swallace.github.io/coloc>
- SuSiE: <https://github.com/stephenslab/susieR/>
- GTEx: <https://www.gtexportal.org/home/datasets>
- LDstore: <http://www.christianbenner.com/>
- Popcorn: <https://github.com/brielin/Popcorn>
- TwosampleMR: <https://mrcieu.github.io/TwoSampleMR/>
- MR-PRESSO: <https://github.com/rondolab/MR-PRESSO>
- PRS-CS: <https://github.com/getian107/PRSes>
- GWASLab: <https://github.com/Cloufield/gwaslab>
- formatbook: <https://github.com/Cloufield/formatbook>

Chapter III - Results

3.1 The three-stage study design

In this study, a three-stage genome-wide analysis of PUD and its subtypes was conducted including (1) discovery stage GWASs, (2) East Asian-specific meta-analyses, and (3) cross-ancestry meta-analyses. An overview of the workflow was provided in Figure 2 and Figure 3. PUD cases in the East Asian populations were obtained by combining individuals with any of the PUD subtypes (namely GU and DU). Individuals with comorbidities of GU and DU were classified as BU (both GU and DU) cases (Figure 4). Detailed information on the number and characteristics of the cases and controls was summarized in Table 3.

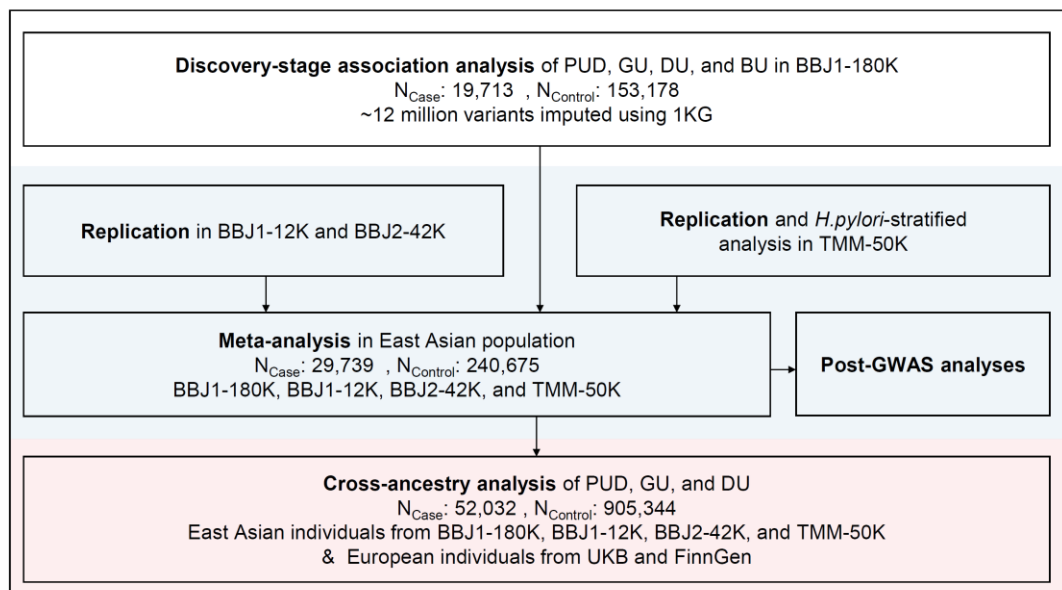


Figure 2. Overview of the study design.

1KG, 1000 Genomes Project panel; PUD, peptic ulcer diseases; GU, gastric ulcers; DU, duodenal ulcers; BU, comorbidities of DU and GU; BBJ1-180K, approximately 180K individuals from Biobank Japan 1st cohort; BBJ1-12K, approximately 12K individuals from BioBank Japan 1st cohort; BBJ2-42K, approximately 42K individuals from Biobank Japan 2nd cohort; TMM-50K, approximately 50K individuals from Tohoku Medical Megabank; UKB, UK Biobank.

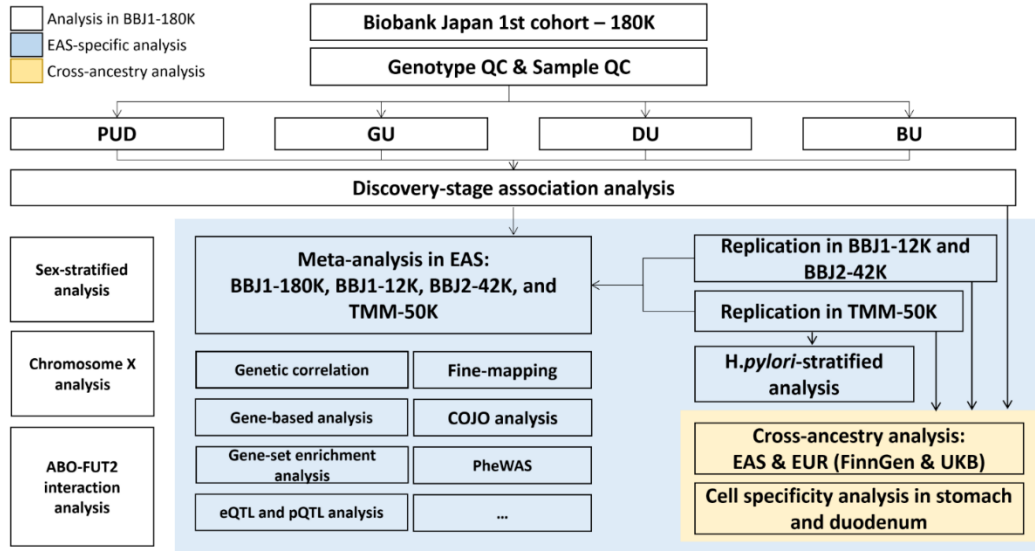


Figure 3. Study workflow.

The three-stage design of this study was shown. We first performed discovery-stage GWAS for PUD and its subtypes in BBJ1-180K and then conducted replication in three independent studies. Next, East Asian-specific meta-analyses were conducted for PUD and its subtypes combining the four studies, and post-GWAS analyses were performed mainly using East Asian-specific summary statistics. Finally, we carried out cross-ancestry meta-analyses for PUD, DU, and GU, combining the four East Asian studies and GWASs in FinnGen and UKB.

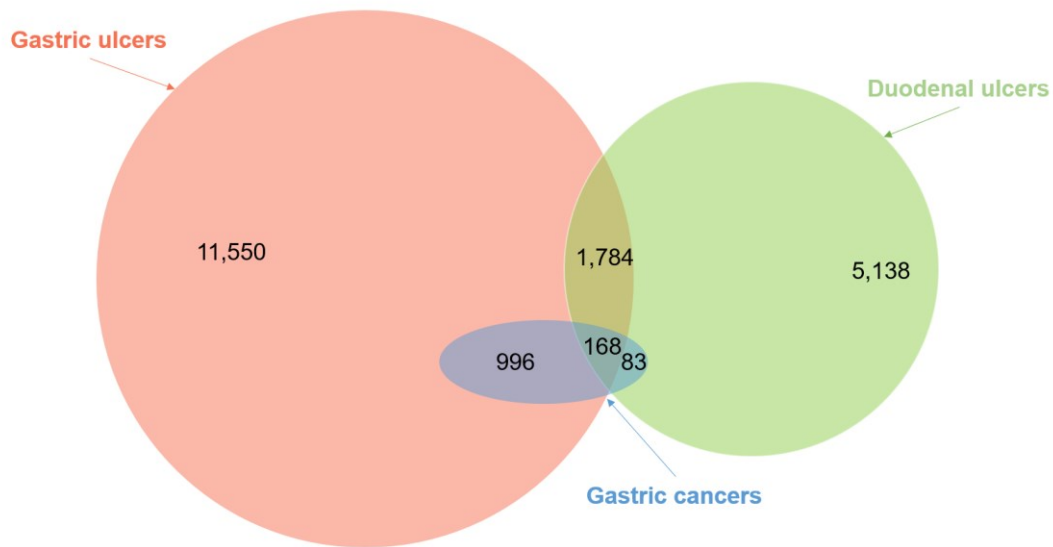


Figure 4. Venn diagram of cases used in the discovery-stage GWAS.

Phenotype overlap among individuals with gastric ulcers, duodenal ulcers, and gastric cancers was shown for PUD cases in BBJ1-180K.

3. 2 Association analyses of PUD and its subtypes in BBJ-180K

The discovery-stage GWAS of PUD and subtypes (GU, DU, and BU) was performed on the Biobank Japan first cohort¹⁴ (BBJ1)-180K dataset, which included 19,713 PUD cases and 153,178 controls (Table 3) and was imputed using 1000 Genomes Project¹⁶ phase 3 (1KG phase 3) reference panel. Around 12 million variants (minor allele count (MAC) >20 and imputation accuracy Rsq > 0.3) were tested for association with SAIGE²⁰, which is a generalized mixed model that controls for the case-control imbalance. In this BBJ-180K dataset, the case-to-control ratios ranged from 1:7.7 to 1:82 (Table 3).

In the discovery-stage GWAS, ten genome-wide significant risk loci ($P < 5.0 \times 10^{-8}$) were identified for PUD, five of which had not been reported as genome-wide significant loci in previous GWAS of PUD or any subtype. For DU, 14 risk loci reached the significance threshold, including seven novel loci (three of which overlapped with the novel PUD loci). For GU and BU, one previously reported locus at *PSCA*³ was identified. In the discovery-stage association analysis for PUD and its subtypes, a total of 15 non-overlapping risk loci reached the genome-wide significance threshold, of which nine were novel (Table 4; Figure 5). Analysis of the X chromosome identified one reported locus at *GUCY2F*¹³ for PUD, GU, and DU (Table 5). For the sex-stratified analysis of PUD and its subtypes, 13 non-overlapping significant loci were identified (13 for males and one for females; Table 5).

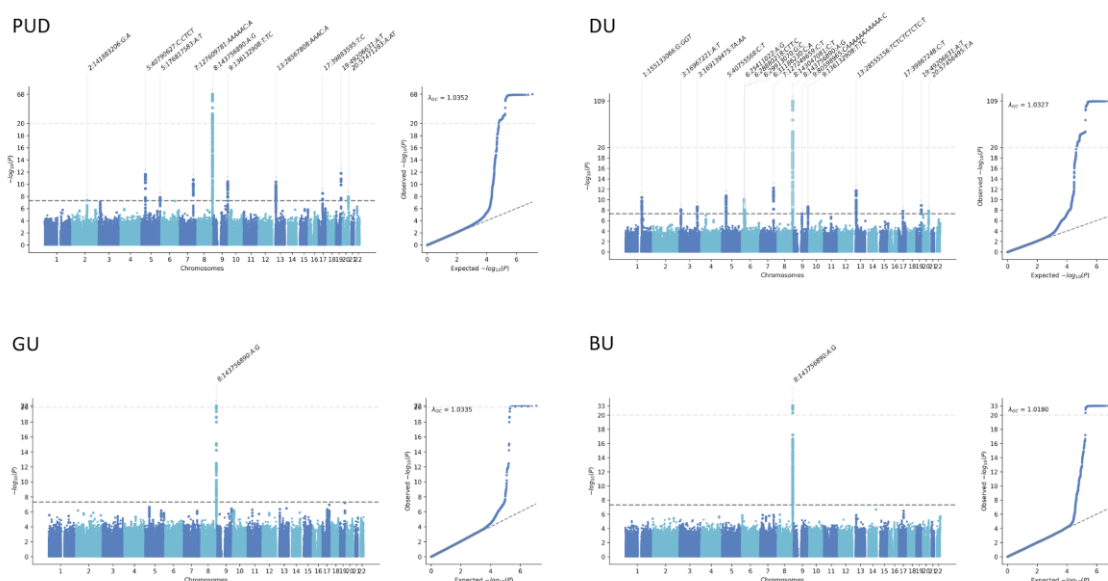


Figure 5. Manhattan plots and Q-Q plots for PUD and its subtypes from the discovery-stage GWAS in BBJ1-180K.

P values were derived from the discovery-stage GWAS in BBJ1-180K. Variants are plotted against GRCh37. For variants above the top light grey dashed line ($-\log_{10}(P) > 20$), P values are rescaled. Significant loci are annotated with the lead variant. The bottom dark grey dashed line indicates the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$).

Table 3. Characteristics of participants in BBJ1-180K, BBJ1-12K, BBJ2-42K, and TMM-50K.

Stage	Study	Sex	Control		PUD		DU		GU		BU		Prevalence rate				Case-control ratio			
			Age (mean±SD)	N	Age (mean±SD)	N	Age (mean±SD)	N	Age (mean±SD)	N	Age (mean±SD)	N	PUD	DU	GU	BU	PUD	DU	GU	BU
Discovery	BBJ1-180K	Male	62.66 ± 13.83	78211	65.48 ± 11.07	13383	63.58 ± 11.26	5218	66.28 ± 10.88	9528	63.78 ± 11.18	1363	14.6%	6.3%	10.9%	1.7%				
		Female	62.17 ± 15.18	74967	65.0 ± 13.09	6330	61.69 ± 12.99	1949	65.96 ± 12.97	4885	61.6 ± 13.02	504	7.8%	2.5%	6.1%	0.7%				
		Both	62.42 ± 14.51	153178	65.33 ± 11.76	19713	63.07 ± 11.78	7167	66.17 ± 11.63	14413	63.19 ± 11.74	1867	11.4%	4.5%	8.6%	1.2%	7.77	21.37	10.63	82
Replication	BBJ1-12K	Male	61.98 ± 14.22	3634	65.18 ± 11.0	642	63.98 ± 11.56	240	65.81 ± 10.7	469	65.25 ± 11.36	67	15.0%	6.2%	11.4%	1.8%				
		Female	55.07 ± 17.53	4983	61.48 ± 15.26	359	55.09 ± 14.72	143	64.29 ± 14.75	251	55.49 ± 15.67	35	6.7%	2.8%	4.8%	0.7%				
		Both	57.99 ± 16.57	8617	63.86 ± 12.81	1001	60.66 ± 13.52	383	65.28 ± 12.28	720	61.9 ± 13.74	102	10.4%	4.3%	7.7%	1.2%	8.61	22.5	11.97	84.5
	BBJ2-42K	Male	67.53 ± 11.55	17346	69.19 ± 9.2	2559	68.5 ± 9.25	1101	69.6 ± 9.06	1750	69.03 ± 8.74	292	12.9%	6.0%	9.2%	1.7%				
		Female	64.99 ± 13.01	17301	67.29 ± 11.42	1078	64.96 ± 10.8	387	68.03 ± 11.48	786	63.91 ± 10.05	95	5.9%	2.2%	4.3%	0.5%				
		Both	66.26 ± 12.37	34647	68.63 ± 9.95	3637	67.58 ± 9.8	1488	69.11 ± 9.9	2536	67.77 ± 9.33	387	9.5%	4.1%	6.8%	1.1%	9.53	23.28	13.66	89.5
	TMM-50K	Male	61.85 ± 10.79	15631	64.14 ± 8.09	2973	63.75 ± 8.38	1684	64.42 ± 7.89	1804	63.84 ± 8.43	515	16.0%	9.7%	10.3%	3.2%				
		Female	59.32 ± 11.56	28602	61.37 ± 8.93	2415	61.08 ± 8.87	1242	61.44 ± 8.99	1499	60.61 ± 8.95	326	7.8%	4.2%	5.0%	1.1%				
		Both	60.22 ± 11.35	44233	62.90 ± 8.59	5388	62.62 ± 8.69	2926	63.07 ± 8.53	3303	62.59 ± 8.77	841	10.9%	6.2%	6.9%	1.9%	8.21	15.12	13.39	52.6
Total			240675		29739		11964		20972		3197	11.0%	4.7%	8.0%	1.3%	8.09	20.12	11.48	75.3	

Table 4. Significant loci associated with PUD or PUD subtypes in discovery discover-stage GWAS

* If the significant locus is novel, the analysis which first identified the loci is shown. BBJ1-180K, the discovery-stage GWAS.

Phenotype	rsID	SNPID	CHR	POS (GRCh37)	Region	Nearest Gene	EA	NEA	EA	BETA	SE	P	Novel or known locus*	Index of novel locus at the stage	Note
PUD	rs10179688	2:141883206:G:A	2	141883206	intronic	LRPIB	A	G	0.513	-0.063	0.012	4.68E-08	BBJ1-180K	1	not replicated or not significant in meta-analysis
PUD	rs546238399	5:40790627:C:CTCT	5	40790627	intronic	PRKAA1	CTCT	C	0.616	0.083	0.012	2.41E-12	BBJ1-180K	2	
PUD	rs55785724	5:176817583:A:T	5	176817583	intronic	SLC34A1	T	A	0.305	-0.067	0.012	1.50E-08	BBJ1-180K	3	
PUD	rs201976370	7:127609781:AAAAAC:A	7	127609781	intronic	SND1	A	AAAAAC	0.093	0.151	0.023	1.79E-11	BBJ1-180K	4	
PUD	rs34635647	8:143756890:A:G	8	143756890	intergenic	PSCA	G	A	0.618	-0.201	0.011	8.22E-69	Known		
PUD	rs8176719	9:136132908:T:TC	9	136132908	exonic	ABO	TC	T	0.454	-0.072	0.011	3.92E-11	Known		
PUD	rs148675590	13:28567808:AAAC:A	13	28567808	intergenic	URAD	A	AAAC	0.196	0.095	0.014	4.54E-11	Known		
PUD	rs7216154	17:39883595:T:C	17	39883595	exonic	HAP1	C	T	0.729	0.080	0.014	3.30E-09	Known		
PUD	rs1047781	19:49206631:A:T	19	49206631	exonic	FUT2	T	A	0.385	0.097	0.014	1.67E-12	Known		
PUD	rs397768861	20:57471283:A:AT	20	57471283	intronic	GNAS	AT	A	0.710	0.071	0.012	1.13E-08	BBJ1-180K	5	
DU	rs141625351	1:155133066:G:GGT	1	155133066	intergenic	KRTCAP2	GGT	G	0.090	0.219	0.033	3.72E-11	Known		
DU	rs1156336	3:16967221:A:T	3	16967221	intronic	PLCL2	T	A	0.482	0.100	0.017	8.10E-09	BBJ1-180K	6	
DU	rs530990807	3:169139475:TA:AA	3	169139475	intronic	MECOM	A	T	0.275	-0.118	0.020	2.32E-09	BBJ1-180K	7	
DU	rs3805495	5:40755568:C:T	5	40755568	intronic	TTC33	T	C	0.579	0.117	0.017	1.85E-11	BBJ1-180K	2	
DU	rs79774308	6:25411022:A:G	6	25411022	intronic	CARMIL1	G	A	0.056	0.252	0.039	7.65E-11	BBJ1-180K	8	
DU	rs61160304	7:127249659:C:T	7	127249659	downstream	PAX4	T	C	0.102	0.219	0.030	5.75E-13	BBJ1-180K	4	
DU	rs4917256	8:143047081:C:T	8	143047081	intergenic	MIR4539	T	C	0.414	-0.102	0.018	2.04E-08	Known		
DU	rs34635647	8:143756890:A:G	8	143756890	intergenic	PSCA	G	A	0.619	-0.410	0.018	1.78E-109	Known		
DU	rs534451473	9:80598965:CAAAAAAAAAA:C	9	80598965	intronic	GNAQ	C	CAAAAAAAAAA	0.394	0.102	0.019	4.98E-08	BBJ1-180K	9	
DU	rs8176719	9:136132908:T:TC	9	136132908	exonic	ABO	TC	T	0.455	-0.104	0.017	2.26E-09	Known		
DU	rs201008788	13:28555156:TCTCTCTCTC:T	13	28555156	intronic	URAD	T	TCTCTCTCTC	0.181	0.162	0.023	1.83E-12	Known		
DU	rs34074411	17:39867248:C:T	17	39867248	intergenic	GAST	T	C	0.619	0.117	0.021	1.62E-08	Known		
DU	rs1047781	19:49206631:A:T	19	49206631	exonic	FUT2	T	A	0.384	0.133	0.022	1.27E-09	Known		
DU	rs6026574	20:57456495:T:A	20	57456495	ncRNA intronic	LOC101927932	A	T	0.660	0.106	0.019	1.36E-08	BBJ1-180K	5	
GU	rs34635647	8:143756890:A:G	8	143756890	intergenic	PSCA	G	A	0.621	-0.128	0.013	2.11E-22	Known		
BU	rs34635647	8:143756890:A:G	8	143756890	intergenic	PSCA	G	A	0.622	-0.431	0.035	4.82E-34	Known		

Table 5. Significant loci associated with PUD or PUD subtypes from sex-stratified and chrX analyses in BBJ1-180K.

Analysis	Phenotype	rsID	SNPID	Region	Nearest Gene	CHR	POS (GRCh37)	EA	NEA	EAF	BETA	SE	P
Male (autosomes)	PUD	rs4685405	3:16981683:G:T	intronic	PLCL2	3	16981683	T	G	0.521	0.075	0.014	3.28013E-08
	PUD	rs146095444	6:131582218:C:T	intronic	AKAP7	6	131582218	T	C	0.022	0.387	0.062	3.67069E-10
	PUD	rs201976370	7:127609781:AAAAAC:A	intronic	SND1	7	127609781	A	AAAAAC	0.093	0.175	0.028	2.46457E-10
	PUD	rs2978977	8:143755720:C:A	intergenic	JRK	8	143755720	A	C	0.484	-0.214	0.014	4.18132E-55
	PUD	rs8176719	9:136132908:T:TC	exonic	ABO	9	136132908	TC	T	0.454	-0.076	0.013	1.99226E-08
	PUD	rs148675590	13:28567808:AAAC:A	intergenic	URAD	13	28567808	A	AAAC	0.197	0.109	0.018	8.48754E-10
	PUD	rs7216154	17:39883595:T:C	exonic	HAP1	17	39883595	C	T	0.729	0.099	0.017	2.28481E-09
	PUD	rs1047781	19:49206631:A:T	exonic	FUT2	19	49206631	T	A	0.386	0.097	0.017	8.84756E-09
	GU	rs2978977	8:143755720:C:A	intergenic	JRK	8	143755720	A	C	0.487	-0.146	0.016	3.45558E-20
	DU	rs141625351	1:155133066:G:GGT	intergenic	KRTCAP2	1	155133066	GGT	G	0.091	0.238	0.039	9.34757E-10
	DU	rs7637568	3:16965216:T:C	intronic	PLCL2	3	16965216	C	T	0.490	0.126	0.021	8.20662E-10
	DU	rs79928271	3:169139475:TA:AA	intronic	MECOM	3	169139475	AA	TA	0.276	-0.132	0.023	1.05614E-08
	DU	rs11956047	5:40812377:G:A	intergenic	RPL37	5	40812377	A	G	0.563	0.115	0.021	3.03487E-08
	DU	rs79774308	6:25411022:A:G	intronic	CARMIL1	6	25411022	G	A	0.056	0.285	0.046	4.61564E-10
	DU	rs61160304	7:127249659:C:T	downstream	PAX4	7	127249659	T	C	0.102	0.236	0.036	4.03393E-11
	DU	rs4917256	8:143047081:C:T	intergenic	MIR4539	8	143047081	T	C	0.413	-0.122	0.021	1.2591E-08
	DU	rs111617116	8:143753693:T:TCCC	intergenic	JRK	8	143753693	TCCC	T	0.615	-0.413	0.022	2.0394E-78
	DU	rs148675590	13:28567808:AAAC:A	intergenic	URAD	13	28567808	A	AAAC	0.196	0.172	0.027	2.06901E-10
	DU	rs6026579	20:57463993:C:T	intronic	GNAS	20	57463993	T	C	0.707	0.136	0.023	3.4868E-09
	BU	rs2976387	8:143759364:G:A	intergenic	PSCA	8	143759364	A	G	0.491	-0.412	0.040	2.96957E-25
Female (autosomes)	PUD	rs2294008	8:143761931:C:T	UTR5	PSCA	8	143761931	T	C	0.631	-0.171	0.019	5.49509E-19
	DU	rs34635647	8:143756890:A:G	intergenic	PSCA	8	143756890	G	A	0.619	-0.421	0.035	8.9455E-34
	BU	rs34635647	8:143756890:A:G	intergenic	PSCA	8	143756890	G	A	0.620	-0.482	0.068	9.83791E-13
ChrX	PUD	rs1402972219	X:108570184:CA:CAA	intergenic	GUCY2F	X	108570184	C	CA	0.454	0.084	0.011	1.382E-15
	GU	rs1205530	X:108540162:C:T	intergenic	GUCY2F	X	108540162	T	C	0.370	0.068	0.011	2.318E-09
	DU	rs1102322	X:108561629:T:G	intergenic	GUCY2F	X	108561629	T	G	0.422	0.118	0.015	6.498E-15

3. 3 Replication and East Asian-specific meta-analysis

In the second stage, replication GWAS was conducted in individuals from three independent studies, namely BBJ1-12K (1,001 cases), BBJ2-42K (3,637 cases), and TMM¹⁵-50K (a population-based study; 5,388 cases; Table 3). To keep consistency, the replication datasets were also imputed using 1KG Phase 3 panel and tested for associations of autosomal variants with the same settings as in the discovery GWAS in BBJ1-180K. Among the nine novel lead variants associated with PUD or its subtypes, four were nominally associated ($P < 0.05$) with PUD or its subtypes in the same direction in at least two replication datasets. Five novel loci were replicated in the population-based dataset of TMM-50K ($P < 0.05$ in the same direction; Table 6).

To further increase statistical power, an East Asian-specific meta-analysis combining the discovery GWAS and three replication GWAS ($N_{\text{case}} = 29,739$; $N_{\text{control}} = 240,675$) was performed. Fixed-effect meta-analyses using the inverse-variance-weighted (IVW) method implemented in METAL software were performed for PUD and its subtypes. The genomic inflation factors (λ_{gc}) and linkage disequilibrium (LD) score regression (LDSC²⁷) intercepts ranged from 1.03 to 1.08 and from 1.01 to 1.02, respectively (Table 7), which indicates no substantial bias. In the second stage (namely EAS-specific meta-analysis), a total of 25 non-overlapping risk loci associated with PUD or any subtype were identified, of which 11 were additional novel loci that were not reported previously or identified in the discovery GWAS in this study (Table 8; Table 9; Figure 6).

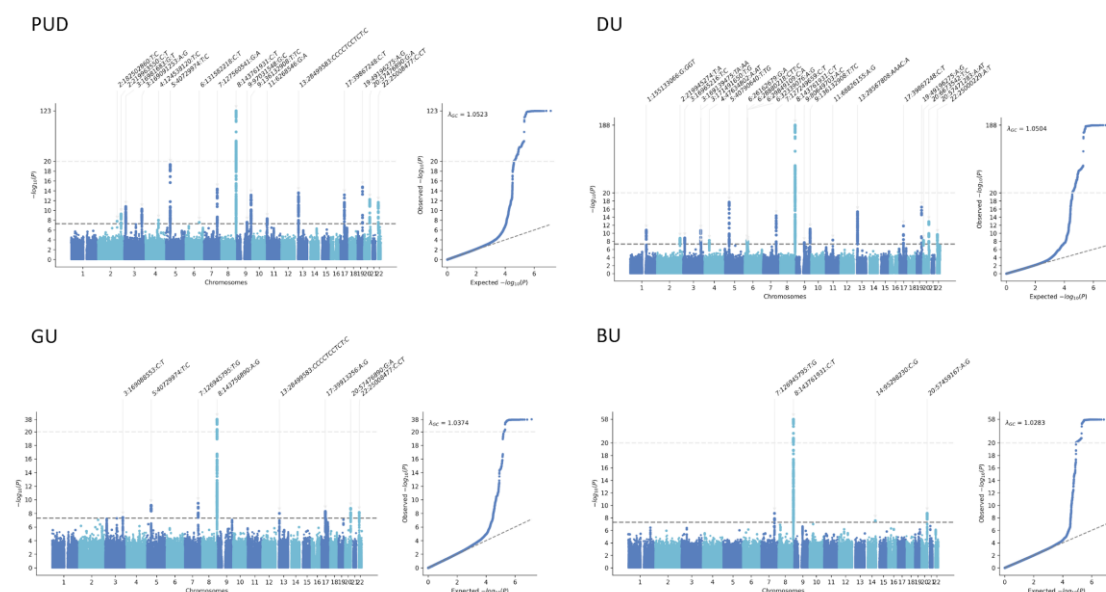


Figure 6. Manhattan plots and Q-Q plots for PUD and its subtypes from the East Asian-specific meta-analysis.

P values were derived from the East Asian-specific meta-analyses combining the four studies. Variants are plotted against GRCh37. For variants above the top light grey dashed line ($-\log_{10}(P) > 20$), P values are rescaled. Significant loci are annotated with the lead variant. The bottom dark grey dashed line indicates the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$).

Table 6. Replication of novel variants identified in discovery stage GWAS in BBJ1-12K, BBJ2-42K, and TMM-50K.

SNPID	CHR	POS (GRCh37)	EA	Imputation Rsq	Study	PUD					DU					GU					BU				
						BETA	SE	P	HetSg	HetP	BETA	SE	P	HetSg	HetP	BETA	SE	P	HetSg	HetP	BETA	SE	P	HetSg	HetP
1:155133066:G:GGT	1	155133066	GGT	0.832 BBJ1-180K	0.094	0.021	4.87E-06				0.219	0.033	3.72E-11			0.043	0.024	7.17E-02			0.174	0.063	5.62E-03		
				0.738 BBJ1-12K	0.006	0.080	9.39E-01				0.106	0.123	3.89E-01			-0.023	0.093	8.01E-01			0.191	0.236	4.19E-01		
				0.740 BBJ2-42K	-0.079	0.043	6.61E-02				-0.026	0.065	6.91E-01			-0.126	0.051	1.25E-02			-0.191	0.125	1.25E-01		
				0.805 TMM-50K	0.081	0.042	4.99E-02				0.192	0.055	4.44E-04			0.033	0.052	5.18E-01			0.274	0.098	5.28E-03		
				Meta	0.063	0.017	1.67E-04	78.5	2.97E-03		0.171	0.025	1.70E-11	74.4	8.48E-03	0.014	0.019	4.83E-01	68.3	2.36E-02	0.145	0.048	2.41E-03	67.5	2.62E-02
2:141883206:G:A	2	141883206	A	0.886 BBJ1-180K	-0.063	0.012	4.68E-08				-0.078	0.018	2.00E-05			-0.054	0.013	5.04E-05			-0.049	0.035	1.65E-01		
				0.790 BBJ1-12K	-0.018	0.054	7.38E-01				-0.031	0.084	7.15E-01			0.014	0.063	8.30E-01			0.160	0.160	3.17E-01		
				0.795 BBJ2-42K	0.044	0.028	1.14E-01				0.039	0.042	3.58E-01			0.062	0.033	5.77E-02			0.144	0.081	7.67E-02		
				0.956 TMM-50K	-0.010	0.021	6.28E-01				0.004	0.028	8.74E-01			0.006	0.027	8.28E-01			0.106	0.051	3.72E-02		
				Meta	-0.039	0.009	2.82E-05	80.2	1.69E-03		-0.042	0.014	2.84E-03	70.8	1.63E-02	-0.029	0.011	9.29E-03	77.5	3.92E-03	0.021	0.027	4.33E-01	69.3	2.05E-02
3:16967221:A:T	3	16967221	T	0.983 BBJ1-180K	0.054	0.011	7.40E-07				0.100	0.017	8.10E-09			0.042	0.013	8.36E-04			0.137	0.033	3.91E-05		
				0.955 BBJ1-12K	0.077	0.049	1.17E-01				0.092	0.076	2.24E-01			0.080	0.057	1.59E-01			0.146	0.144	3.09E-01		
				0.956 BBJ2-42K	0.042	0.026	1.02E-01				0.029	0.039	4.58E-01			0.044	0.030	1.42E-01			0.005	0.074	9.48E-01		
				0.988 TMM-50K	0.053	0.021	1.20E-02				0.060	0.028	2.96E-02			0.051	0.026	5.12E-02			0.066	0.050	1.84E-01		
				Meta	0.053	0.009	2.55E-09	0	9.31E-01		0.082	0.014	1.66E-09	18.4	2.99E-01	0.045	0.010	1.60E-05	0	9.22E-01	0.103	0.026	5.45E-05	12.6	3.30E-01
3:169139475:TA:AA	3	169139475	AA	0.962 BBJ1-180K	-0.041	0.012	8.99E-04				-0.118	0.020	2.32E-09			-0.012	0.014	3.95E-01			-0.109	0.038	3.68E-03		
				0.981 BBJ1-12K	-0.039	0.053	4.68E-01				-0.074	0.083	3.68E-01			-0.035	0.062	5.75E-01			-0.137	0.157	3.83E-01		
				0.981 BBJ2-42K	-0.032	0.028	2.48E-01				-0.073	0.042	8.12E-02			-0.028	0.033	3.87E-01			-0.167	0.081	3.82E-02		
				0.995 TMM-50K	-0.048	0.023	3.95E-02				-0.081	0.031	7.79E-03			-0.037	0.029	2.03E-01			-0.121	0.055	2.73E-02		
				Meta	-0.041	0.010	3.86E-05	0	9.79E-01		-0.101	0.015	2.02E-11	0	6.45E-01	-0.019	0.012	1.04E-01	0	8.59E-01	-0.120	0.028	2.30E-05	0	9.32E-01
5:40790627:C:CTCT	5	40790627	CTCT	0.885 BBJ1-180K	0.083	0.012	2.41E-12				0.125	0.019	3.41E-11			0.069	0.014	3.71E-07			0.141	0.036	9.32E-05		
				0.864 BBJ1-12K	0.080	0.053	1.29E-01				0.226	0.082	5.74E-03			0.035	0.062	5.74E-01			0.300	0.156	5.38E-02		
				0.860 BBJ2-42K	0.120	0.027	1.16E-05				0.159	0.041	1.28E-04			0.099	0.032	2.11E-03			0.134	0.079	9.10E-02		
				0.907 TMM-50K	0.088	0.023	1.04E-04				0.107	0.030	3.24E-04			0.052	0.028	6.23E-02			0.019	0.054	7.29E-01		
				Meta	0.089	0.010	4.56E-20	0	6.70E-01		0.128	0.015	2.05E-18	0	4.73E-01	0.069	0.011	8.98E-10	0	6.73E-01	0.113	0.028	4.24E-05	42.3	1.58E-01
5:176817583:A:T	5	176817583	T	0.980 BBJ1-180K	-0.067	0.012	1.50E-08				-0.070	0.019	2.07E-04			-0.066	0.014	1.51E-06			-0.066	0.036	6.72E-02		
				0.989 BBJ1-12K	0.017	0.053	7.47E-01				0.035	0.082	6.65E-01			0.013	0.062	8.29E-01			0.069	0.156	6.59E-01		
				0.989 BBJ2-42K	0.002	0.027	9.41E-01				0.002	0.041	9.54E-01			-0.009	0.032	7.78E-01			-0.068	0.079	3.87E-01		
				0.996 TMM-50K	-0.010	0.023	6.46E-01				-0.034	0.030	2.52E-01			0.008	0.028	7.88E-01			-0.027	0.054	6.16E-01		
				Meta	-0.046	0.010	2.56E-06	70.7	1.66E-02		-0.049	0.015	8.32E-04	26.8	2.51E-01	-0.044	0.011	8.72E-05	62.2	4.73E-02	-0.052	0.028	6.00E-02	0	8.00E-01
6:25411022:A:G	6	25411022	G	0.946 BBJ1-180K	0.093	0.024	1.11E-04				0.252	0.039	7.65E-11			0.026	0.028	3.52E-01			0.176	0.073	1.64E-02		
				0.956 BBJ1-12K	-0.039	0.105	7.14E-01				0.012	0.163	9.40E-01			-0.052	0.122	6.70E-01			-0.002	0.307	9.94E-01		
				0.958 BBJ2-42K	0.088	0.054	1.03E-01				0.054	0.082	5.06E-01			0.083	0.064	1.91E-01			-0.073	0.157	6.45E-01		
				0.963 TMM-50K	0.031	0.050	5.30E-01				0.064	0.066	3.32E-01			0.020	0.062	7.42E-01			0.102	0.117	3.84E-01		
				Meta	0.078	0.020	7.84E-05	0	4.69E-01		0.176	0.030	6.41E-09	70.1	1.83E-02	0.030	0.023	1.97E-01	0	7.54E-01	0.120	0.057	3.44E-02	0	5.21E-01
7:127249659:C:T	7	127249659	T	0.894 BBJ1-180K	0.123	0.019	8.04E-11				0.219	0.030	5.75E-13			0.094	0.022	1.72E-05			0.258	0.058	9.02E-06		
				0.452 BBJ1-12K	-0.011	0.133	9.34E-01				-0.156	0.205	4.48E-01			0.053	0.156	7.33E-01			-0.123	0.394	7.55E-01		
				0.470 BBJ2-42K	0.222	0.067	9.46E-04				0.212	0.102	3.83E-02			0.239	0.079	2.60E-03			0.293	0.197	1.36E-01		
				0.751 TMM-50K	0.133	0.042	1.43E-03				0.161	0.055	3.46E-03			0.155	0.052	2.85E-03			0.323	0.100	1.19E-03		
				Meta	0.129	0.017	8.52E-15	4.3	3.71E-01		0.200	0.026	4.59E-15	23.5	2.70E-01	0.110	0.019	1.21E-08	26.9	2.51E-01	0.269	0.048	2.33E-08	0	7.20E-01
8:143756890:A:G	8	143756890	G	0.954 BBJ1-180K	-0.201	0.011	8.22E-69				-0.410	0.018	1.78E-109			-0.128	0.013	2.11E-22			-0.431	0.035	4.82E-34		
				0.949 BBJ1-12K	-0.333	0.051	6.00E-11				-0.482	0.080	1.37E-09			-0.257	0.059	1.47E-05			-0.346	0.151	2.17E-02		
				0.930 BBJ2-42K	-0.215	0.026	3.39E-16				-0.440	0.040	1.53E-27			-0.107	0.031	6.23E-04			-0.358	0.077	3.88E-06		
				0.940 TMM-50K	-0.284	0.023	2.73E-36				-0.426	0.030	9.50E-46			-0.215	0.028	2.04E-14			-0.504	0.055	2.96E-20		
				Meta	-0.222	0.009	1.16E-123	80.8	1.35E-03		-0.420	0.014	9.44E-187	0	7.52E-01	-0.143	0.011	4.73E-39	76.7	4.90E-03	-0.437	0.027	9.33E-58	0	3.99E-01
9:80598965:CAAAAAAAAAAA:C	9	80598965	C	0.884 BBJ1-180K	0.054	0.012	4.52E-06				0.102	0.019	4.98E-08			0.044	0.014	1.13E-03			0.161	0.036	7.15E-06		
				0.369 BBJ1-12K	0.002	0.082	9.84E-01				0.058	0.127	6.48E-01			0.024	0.095	8.04E-01			0.356	0.241	1.40E-01		
				0.395 BBJ2-42K	0.052	0.042	2.11E-01				0.037	0.063	5.54E-01			0.080	0.049	1.02E-01			0.177	0.121	1.44E-01		
				0.917 TMM-50K	0.005	0.022	8.23E-01				0.032	0.029	2.76E-01			0.006	0.028	8.18E-01			0.103	0.053	5.04E-02		
				Meta	0.043	0.010	1.69E-05	27.4	2.48E-01		0.079	0.015	2.18E-07	34.9	2.03E-01	0.039	0.012	8.57E-04	0	5.20E-01	0.148	0.029	2.44E-07	0	6.48E-01
9:136132908:T:TC	9	136132908	TC	0.994 BBJ1-180K	-0.072	0.011	3.92E-11				-0.104	0.017	2.26E-09			-0.063	0.013	4.79E-07			-0.126	0.033	1.58E-04		
				0.983 BBJ1-12K	-0.075	0.049	1.25E-01				-0.208	0.076	5.82E-03			-0.015	0.057	7.97E-01			-0.143	0.144	3.21E-01		
				0.992 BBJ2-42K	-0.070	0.025	5.29E-03				-0.113	0.038	3.08E-03			-0.049	0.030	9.91E-02			-0.098	0.074	1.83E-01		
				0.990 TMM-50K	-0.041	0.021	5.05E-02				-0.036	0.028	2.00E-01			-0.036	0.026	1.69E-01			-0.002	0.050	9.61E-01		
				Meta	-0.067	0.009	7.75E-14	0	6.27E-01		-0.092	0.014	8.96E-12	58.3	6.59E-02	-0.056	0.010	9.22E-08	0	6.84E-01	-0.091	0.026	3.61E-04	30.4	2.30E-01
13:28555156:TCTCTCTCTCT:C	13	28555156	T	0.958 BBJ1-180K	0.090	0.014	4.15E-10				0.162	0.023	1.83E-12			0.066	0.017	6.70E-05			0.177	0.044	5.22E-05		
				0.833																					

Table 7. Estimations of confounding bias and SNP-heritability by LD score regression.

Observed-scale heritability was converted to liability-scale. EAS, East Asian-specific meta-analysis of BBJ1-180K, BBJ1-12K, BBJ2-42K, and TMM-50K. N/A, not applicable.

Cohort	Phenotype	Liability-scale heritability	SE (Liability-scale heritability)	P (Liability-scale heritability)	Lambda GC	Mean Chi2	Intercept	SE (Intercept)	Ratio	SE (Ratio)
BBJ1-180K	PUD	0.045	0.008	2.96E-08	1.059	1.079	1.006	0.007	0.070	0.090
BBJ1-180K	GU	0.035	0.007	6.71E-07	1.050	1.055	1.005	0.007	0.083	0.132
BBJ1-180K	DU	0.075	0.020	7.71E-05	1.038	1.059	0.996	0.007	Ratio < 0	N/A
BBJ1-180K	BU	0.028	0.025	1.32E-01	1.014	1.014	1.004	0.006	0.326	0.447
BBJ1-12K	PUD	0.160	0.133	1.15E-01	1.020	1.015	1.004	0.007	0.272	0.465
BBJ1-12K	GU	0.048	0.143	3.68E-01	1.017	1.011	1.008	0.006	0.743	0.585
BBJ1-12K	DU	0.466	0.275	4.50E-02	1.008	1.004	0.989	0.006	Ratio < 0	N/A
BBJ1-12K	BU	N/A	N/A	N/A	1.008	0.994	0.994	0.006	N/A	N/A
BBJ2-42K	PUD	N/A	N/A	N/A	0.984	0.995	1.013	0.007	N/A	N/A
BBJ2-42K	GU	N/A	N/A	N/A	0.987	0.996	1.010	0.007	N/A	N/A
BBJ2-42K	DU	0.099	0.065	6.46E-02	1.005	1.004	0.992	0.006	Ratio < 0	N/A
BBJ2-42K	BU	0.398	0.177	1.22E-02	1.017	1.007	0.988	0.006	Ratio < 0	N/A
TMM-50K	PUD	0.062	0.024	4.60E-03	1.011	1.017	0.994	0.006	Ratio < 0	N/A
TMM-50K	GU	0.069	0.031	1.32E-02	1.014	1.014	0.995	0.007	Ratio < 0	N/A
TMM-50K	DU	0.092	0.039	9.73E-03	1.008	1.016	0.994	0.006	Ratio < 0	N/A
TMM-50K	BU	N/A	N/A	N/A	0.999	0.999	1.004	0.006	N/A	N/A
EAS	PUD	0.055	0.010	6.51E-09	1.083	1.129	1.020	0.007	0.157	0.056
EAS	GU	0.035	0.007	6.26E-08	1.062	1.077	1.018	0.007	0.235	0.085
EAS	DU	0.104	0.025	1.59E-05	1.077	1.118	1.014	0.007	0.114	0.059
EAS	BU	0.061	0.026	9.82E-03	1.032	1.034	1.010	0.007	0.300	0.195

Table 8. Significant loci associated with PUD or PUD subtypes in EAS-specific meta-analysis.

* If the significant loci were novel, the analysis which first identified the loci was shown. BBJ1-180, discovery GWAS. EAS, East Asian-specific meta-analysis. Directions of effect sizes for each lead variant from GWASs were shown in the order of BBJ1-180K, BBJ1-12K, BBJ2-42K, and TMM-50K.

Phenotype	rsID	SNPID	CHR	POS (GRCh37)	Region	Nearest Gene	EA	NEA	EAF	BETA	SE	P	N	DIRECTION	HetISq	HetP	Novel or known locus*	Index of novel locus at the stage
PUD	rs184426772	2:182502860:T:C	2	182502860	intronic	CERKL	C	T	0.0082	0.7957	0.1408	1.605E-08	97523	?+++	0	0.9949	EAS	1
PUD	rs11692085	2:219963550:C:T	2	219963550	intronic	NHEJ1	T	C	0.2706	0.0633	0.0102	5.253E-10	270414	+++	0	0.4209	EAS	2
PUD	rs4685405	3:16981683:G:T	3	16981683	intronic	PLCL2	T	G	0.5241	0.0603	0.009	1.634E-11	270414	++++	0	0.9925	BBJ-180K	
PUD	rs1273886	3:169091253:A:G	3	169091253	intronic	MECOM	G	A	0.1286	0.0889	0.0135	5.223E-11	270414	++++	61.6	0.05012	BBJ-180K	
PUD	rs2553380	4:124538120:T:C	4	124538120	intergenic	LINC01091	C	T	0.648	-0.0602	0.0105	1E-08	270414	----	0	0.576	EAS	3
PUD	rs6860328	5:40729974:T:C	5	40729974	intronic	TTC33	C	T	0.5771	0.0833	0.0091	4.471E-20	270414	++++	0	0.6832	BBJ-180K	
PUD	rs146095444	6:131582218:C:T	6	131582218	intronic	AKAP7	T	C	0.0217	0.2148	0.0387	2.759E-08	270414	++++	44.9	0.142	EAS	4
PUD	rs72607744	7:127560541:G:A	7	127560541	intronic	SND1	A	G	0.0951	0.1477	0.0188	4.449E-15	270414	++++	0	0.4821	BBJ-180K	
PUD	rs2294008	8:143761931:C:T	8	143761931	UTR5	PSCA	T	C	0.6354	-0.2177	0.0092	1.05E-123	270414	----	82.2	0.000752	Known	
PUD	rs12347577	9:97031548:G:C	9	97031548	intronic	ZNF169	C	G	0.3224	-0.0536	0.0096	2.593E-08	270414	----	0	0.6945	EAS	5
PUD	rs8176719	9:136132908:T:TC	9	136132908	exonic	ABO	TC	T	0.4486	-0.0667	0.0089	7.747E-14	270414	----	0	0.6274	Known	
PUD	rs12792379	11:6268546:G:A	11	6268546	intergenic	CNGA4	A	G	0.0896	0.0917	0.0157	4.864E-09	270414	++++	11.5	0.3352	Known	
PUD	rs749161312	13:28499583:CCCCTCCTCT:C	13	28499583	UTR3	PDX1	CCCCTCCTCT	C	0.5944	-0.0757	0.0099	2.541E-14	270414	----	0	0.5938	Known	
PUD	rs34074411	17:39867248:C:T	17	39867248	intergenic	GAST	T	C	0.6342	0.078	0.0104	6.963E-14	270414	++++	0	0.6873	Known	
PUD	rs11665674	19:49196275:A:G	19	49196275	intergenic	FUT2	G	A	0.4486	0.0853	0.0107	1.696E-15	270414	++++	5.3	0.3663	Known	
PUD	rs8114689	20:57476890:G:A	20	57476890	intronic	GNAS	A	G	0.5929	0.0661	0.0092	5.672E-13	270414	++++	0	0.6416	BBJ-180K	
PUD	rs11416248	22:25008477:C:CT	22	25008477	intronic	GGT1	CT	C	0.3207	-0.0762	0.0109	2.279E-12	270414	----	0	0.6325	EAS	6
GU	rs77132614	3:169088553:C:T	3	169088553	intronic	MECOM	T	C	0.0982	0.1013	0.0184	3.969E-08	261647	++++	33.2	0.2133	BBJ-180K	
GU	rs6860328	5:40729974:T:C	5	40729974	intronic	TTC33	C	T	0.5762	0.0655	0.0106	6.387E-10	261647	++++	0	0.8911	BBJ-180K	
GU	rs17866912	7:126945795:T:G	7	126945795	intergenic	GRM8	G	T	0.0983	0.1471	0.0234	3.247E-10	261647	++++	0	0.4995	BBJ-180K	
GU	rs34635647	8:143756890:A:G	8	143756890	intergenic	PSCA	G	A	0.6233	-0.1431	0.0109	4.733E-39	261647	----	76.7	0.0049	Known	
GU	rs749161312	13:28499583:CCCCTCCTCT:C	13	28499583	UTR3	PDX1	CCCCTCCTCT	C	0.595	-0.0665	0.0116	9.968E-09	261647	----	0	0.7332	Known	
GU	rs11869903	17:39913256:A:G	17	39913256	intronic	JUP	G	A	0.5053	0.0613	0.0105	5.13E-09	261647	++++	50.4	0.109	Known	
GU	rs8114689	20:57476890:G:A	20	57476890	intronic	GNAS	A	G	0.5925	0.0644	0.0107	1.832E-09	261647	++++	0	0.7396	BBJ-180K	
GU	rs11416248	22:25008477:C:CT	22	25008477	intronic	GGT1	CT	C	0.3208	-0.0733	0.0127	7.128E-09	261647	----	0	0.6568	EAS	6
DU	rs141625351	1:155133066:G:GGT	1	155133066	intergenic	KRTCAP2	G	GGT	0.9026	-0.1709	0.0254	1.695E-11	252639	+-+	74.4	0.008483	Known	
DU	rs9288539	2:219945274:T:A	2	219945274	intronic	NHEJ1	A	T	0.2762	0.0921	0.0153	1.737E-09	252639	+++	0	0.4463	EAS	2
DU	rs7637568	3:16965216:T:C	3	16965216	intronic	PLCL2	C	T	0.4955	0.0818	0.0135	1.497E-09	252639	++++	0	0.482	BBJ-180K	
DU	rs79928271	3:169139475:TA:AA	3	169139475	intronic	MECOM	A	T	0.2771	-0.1014	0.0151	2.023E-11	252639	----	0	0.6446	BBJ-180K	
DU	rs13086914	3:171491650:T:G	3	171491650	intronic	PLD1	G	T	0.2042	-0.0939	0.0172	4.83E-08	252639	----	0	0.7645	EAS	7
DU	rs34742353	4:47634802:A:AT	4	47634802	intronic	CORIN	A	AT	0.8247	0.1069	0.0184	5.875E-09	252639	++++	0	0.4898	EAS	8
DU	rs373477888	5:40790640:T:TG	5	40790640	intronic	PRKAA1	TG	T	0.6136	0.1279	0.0146	2.045E-18	252639	++++	0	0.4735	BBJ-180K	
DU	rs9467684	6:26162679:G:A	6	26162679	intronic	HIST1H2BD	A	G	0.0597	0.1651	0.0283	5.155E-09	252639	++++	35.8	0.1973	BBJ-180K	
DU	rs61160304	7:127249659:C:T	7	127249659	downstream	PAX4	T	C	0.0985	0.2001	0.0255	4.594E-15	252639	+++	23.5	0.2701	BBJ-180K	
DU	rs2294008	8:143761931:C:T	8	143761931	UTR5	PSCA	T	C	0.6381	-0.4143	0.0141	7.69E-189	252639	----	0	0.7178	Known	
DU	rs10870013	9:80649703:A:C	9	80649703	intergenic	GNAQ	C	A	0.353	0.0831	0.0148	2.105E-08	252639	++++	0	0.5966	BBJ-180K	
DU	rs8176719	9:136132908:T:TC	9	136132908	exonic	ABO	TC	T	0.4474	-0.0924	0.0135	8.956E-12	252639	----	58.3	0.0659	Known	
DU	rs3019776	11:68826155:A:G	11	68826155	intronic	TPCN2	G	A	0.5301	0.093	0.0159	4.984E-09	252639	++++	20.6	0.2864	EAS	9
DU	rs148675590	13:28567808:AAAC:A	13	28567808	intergenic	URAD	A	AAAC	0.191	0.1468	0.0181	4.897E-16	252639	++++	0	0.643	Known	
DU	rs34074411	17:39867248:C:T	17	39867248	intergenic	GAST	T	C	0.6366	0.1107	0.0157	1.58E-12	252639	++++	0	0.7996	Known	
DU	rs11665674	19:49196275:A:G	19	49196275	intergenic	FUT2	G	A	0.4484	0.1363	0.0162	3.381E-17	252639	++++	0	0.771	Known	
DU	rs6117384	20:6673542:T:C	20	6673542	intergenic	LINC01713	C	T	0.7426	0.0925	0.0159	5.697E-09	252639	+++	24.9	0.2619	EAS	10
DU	rs5842244	20:57471283:A:AT	20	57471283	intronic	GNAS	A	AT	0.2879	-0.1126	0.0152	1.318E-13	252639	+-	57.3	0.07097	BBJ-180K	
DU	rs5751904	22:25000229:A:T	22	25000229	intronic	GGT1	T	A	0.319	-0.0988	0.0155	2.089E-10	252639	----	0	0.6916	EAS	6
BU	rs17866912	7:126945795:T:G	7	126945795	intergenic	GRM8	G	T	0.0978	0.3425	0.057	1.907E-09	243872	+++	0	0.7973	BBJ-180K	
BU	rs2294008	8:143761931:C:T	8	143761931	UTR5	PSCA	T	C	0.642	-0.4326	0.0268	1.389E-58	243872	----	0	0.4167	Known	
BU	rs147272036	14:95298230:C:G	14	95298230	intergenic	GSC-DT	G	C	0.0057	1.3272	0.2391	2.85E-08	243872	++++	0	0.4469	EAS	11
BU	rs6026576	20:57459167:A:G	20	57459167	ncRNA intronic	LOC101927932	G	A	0.6613	0.1627	0.0271	1.843E-09	243872	++++	0	0.4207	BBJ-180K	

Table 9. Significant loci associated with PUD or PUD subtypes from genome-wide meta-analyses.

Lead variants from each significant locus identified in the fixed-effect population-specific or the cross-ancestry meta-analysis are shown. Only the most significant lead variant is shown for loci associated with multiple phenotypes or identified in multiple analyses. Variants were annotated with the nearest genes. EAF_{EAS} and EAF_{EUR}, effect allele frequencies in population-specific meta-analysis. For variants not available in the datasets, EAF was obtained from the Genome Aggregation Database or the 1000 Genomes Project.

Phenotypes	rsID	CHR	POS (GRCh37)	Nearest Gene	Effect Allele / Non Effect Allele	OR	95% CI	P	Analysis	EAF _{EAS}	EAF _{EUR}
PUD	rs59781317	1	156016356	<i>UBQLN4</i>	G/A	1.04	(1.03-1.06)	3.51E-08	Cross-ancestry	0.77	0.25
PUD	rs184426772	2	182502860	<i>CERKL</i>	C/T	2.22	(1.68-2.92)	1.61E-08	EAS	0.01	0.00
PUD, DU	rs11692085	2	219963550	<i>NHEJ1</i>	T/C	1.05	(1.04-1.07)	2.58E-12	Cross-ancestry	0.27	0.66
PUD, DU, GU	rs4685405	3	16981683	<i>PLCL2</i>	T/G	1.05	(1.04-1.07)	9.84E-12	Cross-ancestry	0.52	0.18
DU, PUD, GU	rs79928271	3	169139475	<i>MECOM</i>	A/T	0.90	(0.88-0.93)	2.02E-11	EAS	0.28	0.13
DU	rs13086914	3	171491650	<i>PLD1</i>	G/T	0.91	(0.88-0.94)	4.83E-08	EAS	0.20	0.39
DU	rs34742353	4	47634802	<i>CORIN</i>	A/AT	1.11	(1.07-1.15)	5.88E-09	EAS	0.83	0.74
PUD	rs2553380	4	124538120	<i>LINC01091</i>	C/T	0.94	(0.92-0.96)	1.00E-08	EAS	0.65	0.84
PUD, DU, GU	rs3805497	5	40746885	<i>TTC33</i>	T/A	1.08	(1.06-1.09)	5.32E-25	Cross-ancestry	0.58	0.25
PUD	rs1801020	5	176836532	<i>FI2</i>	G/A	0.95	(0.94-0.97)	1.38E-11	Cross-ancestry	0.35	0.75
DU	rs41265804	6	29924159	<i>HLA-A</i>	G/C	0.92	(0.90-0.95)	1.08E-09	Cross-ancestry	0.49	0.17
PUD	rs146095444	6	131582218	<i>AKAP7</i>	T/C	1.24	(1.15-1.34)	2.76E-08	EAS	0.02	0.00
PUD, DU	rs416879	6	160774838	<i>SLC22A3</i>	G/A	0.96	(0.95-0.97)	2.48E-09	Cross-ancestry	0.28	0.48
PUD, DU, GU, BU	rs72607744	7	127560541	<i>SND1</i>	A/G	1.16	(1.12-1.20)	4.45E-15	EAS	0.10	0.00
DU	rs7470279	9	80607789	<i>GNAQ</i>	T/A	1.09	(1.06-1.12)	1.56E-10	Cross-ancestry	0.37	0.15
PUD	rs10992997	9	96701663	<i>BARX1</i>	G/A	1.05	(1.03-1.06)	1.41E-08	Cross-ancestry	0.14	0.42
DU	rs3019776	11	68826155	<i>TPCN2</i>	G/A	1.10	(1.07-1.13)	1.38E-09	Cross-ancestry	0.53	0.97
BU	rs147272036	14	95298230	<i>GSC-DT</i>	G/C	3.77	(2.36-6.02)	2.85E-08	EAS	0.01	0.00
DU	rs511893	16	88990941	<i>CBFA2T3</i>	G/T	0.91	(0.88-0.94)	6.39E-09	Cross-ancestry	0.30	0.34
DU	rs2642030	17	65605075	<i>PITPNC1</i>	G/A	1.07	(1.05-1.10)	2.00E-08	Cross-ancestry	0.56	0.25
DU	rs6117384	20	6673542	<i>LINC01713</i>	C/T	1.09	(1.07-1.12)	6.32E-12	Cross-ancestry	0.74	0.73
PUD, DU, GU, BU	rs6123837	20	57465571	<i>GNAS</i>	A/G	1.05	(1.04-1.07)	2.12E-14	Cross-ancestry	0.58	0.36
PUD, DU	rs12625329	20	62709274	<i>RGS19</i>	A/G	1.04	(1.03-1.06)	8.09E-09	Cross-ancestry	0.66	0.46
PUD, DU, GU	rs11416248	22	25008477	<i>GGT1</i>	CT/C	0.93	(0.91-0.95)	2.28E-12	EAS	0.32	0.22
DU	rs7288137	22	50458020	<i>TTL8</i>	A/G	1.08	(1.06-1.12)	7.73E-09	Cross-ancestry	0.20	0.22
Previously known loci											
PUD, DU	rs1345894981 (rs147048677)	1	155161794	<i>MUC1</i>	T/C	1.14	(1.10-1.19)	2.54E-12	Cross-ancestry	0.02	0.06
DU, PUD, GU, BU	rs2294008	8	143761931	<i>PSCA</i>	T/C	0.66	(0.64-0.68)	7.69E-189	EAS	0.64	0.46
PUD, DU, GU	rs8176719	9	136132908	<i>ABO</i>	TC/T	0.93	(0.92-0.95)	5.76E-24	Cross-ancestry	0.45	0.37
PUD	rs78459074	11	1029905	<i>MUC6</i>	G/A	0.95	(0.93-0.96)	7.60E-09	Cross-ancestry	0.21	0.12
PUD, DU	rs10500661	11	6273744	<i>CCKBR</i>	C/T	1.11	(1.09-1.13)	7.75E-27	Cross-ancestry	0.09	0.20
PUD, DU, GU	rs9581957	13	28557889	<i>URAD</i>	T/C	1.07	(1.06-1.09)	2.20E-20	Cross-ancestry	0.18	0.33
PUD, DU, GU	rs34074411	17	39867248	<i>GAST</i>	T/C	1.07	(1.06-1.09)	1.07E-22	Cross-ancestry	0.63	0.45
DU	rs11665674	19	49196275	<i>FUT2</i>	G/A	1.15	(1.11-1.18)	3.38E-17	EAS	0.45	0.00

3. 4 Cross-ancestry meta-analysis, effect size, and genetic correlation comparison

In the third stage, publicly available European GWASs of PUD and its subtypes using samples from FinnGen and UK Biobank^{12,20,23} were collected (Table 10). After QC and harmonization (Methods), a fixed-effect IVW cross-ancestry meta-analysis (totaling 52,032 PUD cases and 905,344 controls) was carried out, combining the Japanese and European studies. In this stage, six additional loci for PUD and DU reached the genome-wide significance level ($P < 5.0 \times 10^{-8}$; Table 9; Figure 7; Table 11). Furthermore, a cross-ancestry meta-regression utilizing MR-MEGA²⁶ was conducted and identified 23 known and described novel loci mentioned above in the East Asian-specific and cross-ancestry meta-analysis (Table 12).

Across the second and third stages of association analyses (the East Asian-specific and cross-ancestry meta-analyses), a total of 25 non-overlapping novel loci for PUD and its subtypes were identified (Table 9; Figure 9). Two novel loci identified in the discovery stage were not significant in any of the meta-analyses (Table 6).

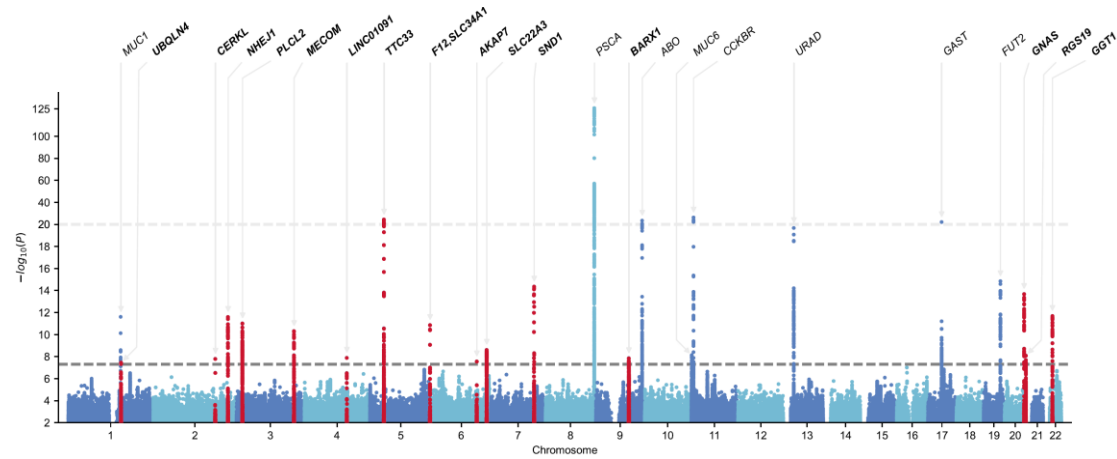


Figure 7. Manhattan plot of the cross-ancestry meta-analysis of PUD.

P values were derived from the cross-ancestry meta-analysis of 52,032 cases and 905,344 controls of EAS or EUR ancestry. Meta-analysis was performed using the inverse-variance weighted method under the fixed-effect model. For variants above the top light grey dashed line ($-\log_{10}(P) > 20$), values are rescaled. Lead variants are annotated with the nearest gene name. Novel loci are highlighted in red. Variants are plotted against GRCh37 (hg19). The bottom dark grey dashed line indicates the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$). Variants with $-\log_{10}(P) < 2$ are omitted.

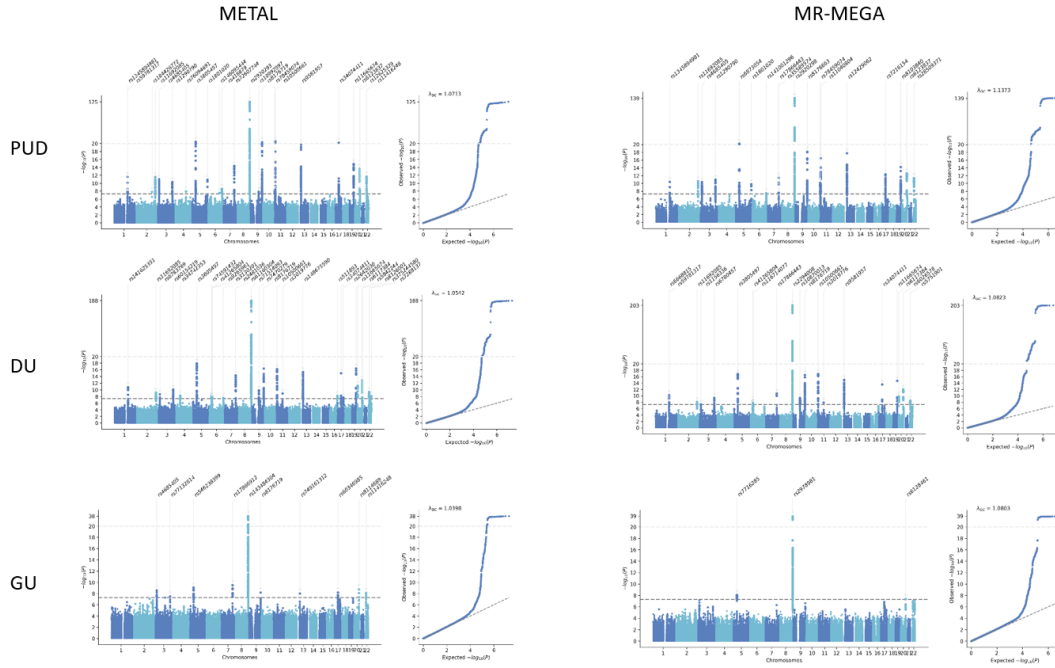


Figure 8. Manhattan plots and Q-Q plots for PUD and its subtypes from the cross-ancestry meta-analysis.

P values were derived from the cross-ancestry meta-analyses using either METAL or MR-MEGA. For summary statistics from MR-MEGA, P values were recalculated from Chi-square statistics. Variants are plotted against GRCh37. For variants above the top light grey dashed line ($-\log_{10}(P) > 20$), P values are rescaled. Significant loci were annotated with the lead variant. The bottom dark grey dashed line indicates the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$).

the difference in the genetic architecture of PUD across ancestries. For PUD subtypes, effect sizes for DU showed a higher correlation ($r = 0.79$) across ancestries compared to that for GU ($r = 0.63$; Table 14). The genetic correlation of GU was relatively low ($\rho_{gi} = 0.45$, $P = 7.3 \times 10^{-3}$), whereas the genetic architecture of DU did not show a significant difference across ancestries ($\rho_{gi} = 0.72$, $P = 9.6 \times 10^{-2}$; Table 14).

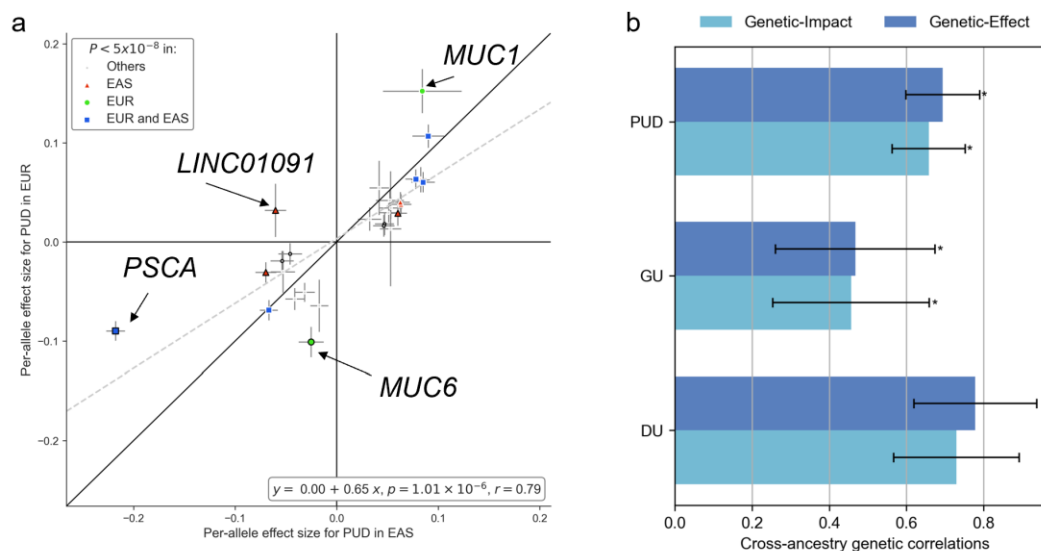


Figure 10. Cross-ancestry effect size comparison and genetic correlation analysis.

a. Per-allele effect size (logarithm of odds ratios) comparison using East Asian-specific and European-specific summary statistics for PUD. Lead variants associated with PUD or any subtype in East Asian-specific, European-specific, or cross-ancestry meta-analysis were selected for comparison. The most significant associations were shown if overlapping variants existed (interval < 500 kb). Variants with nominally significant heterogeneity ($P_{het} < 0.05$) were denoted by the black marker edges. The grey dashed line represents the fitted linear regression line. Pearson's r is shown. b. Cross-ancestry genetic correlation for PUD, GU, and DU estimated by Popcorn. Asterisks indicate estimates that were significantly less than one after Bonferroni correction ($P < 0.05/6$). Error bars represent the standard error.

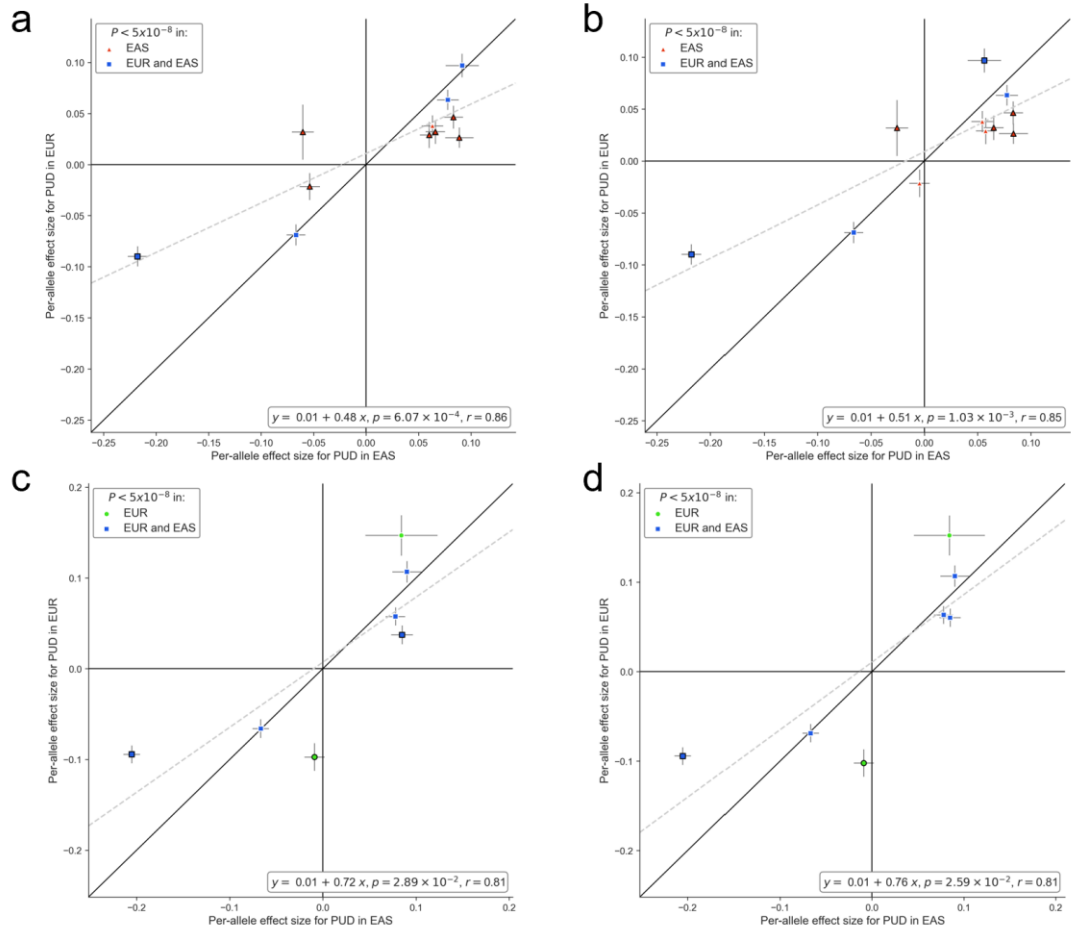


Figure 11. Effect size comparison of lead variants with and without the winner's curse corrections. a, comparison using lead variants of significant loci ascertained in EAS. b, the winner's curse-corrected comparison using lead variants of significant loci ascertained in EAS. c, comparison using lead variants of significant loci ascertained in EUR. d, the winner's curse-corrected comparison using lead variants of significant loci ascertained in EUR.

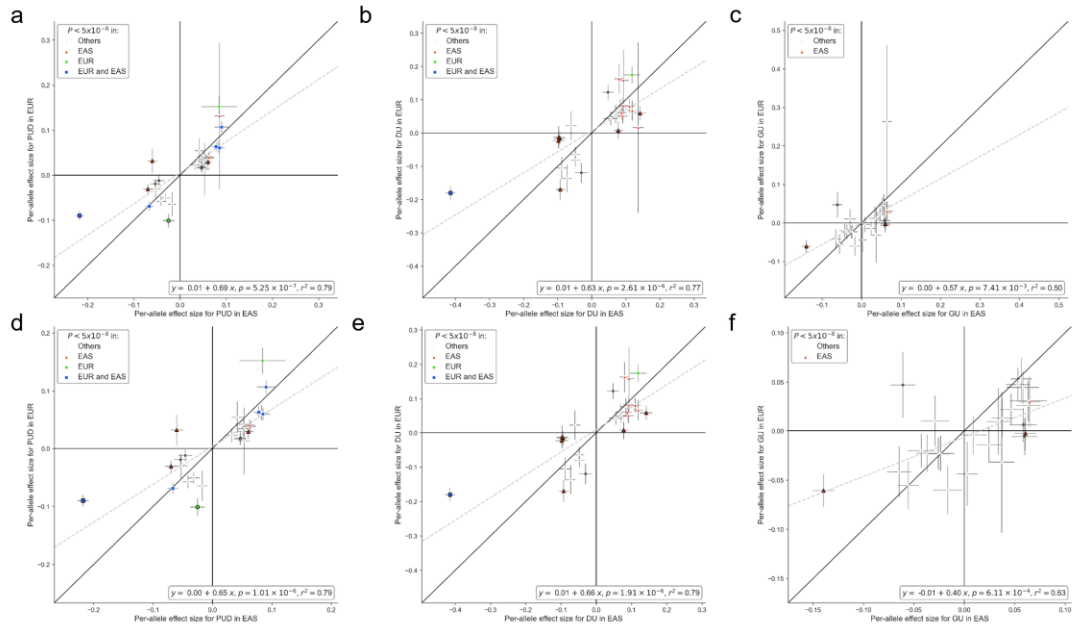


Figure 12. Cross-ancestry effect size comparison of lead variants for PUD and its subtypes.

Per-allele effect size (β) comparison using EAS-specific and EUR-specific summary statistics for PUD and its subtypes. Lead variants associated with PUD or any subtypes in EAS-specific, EUR-specific, or cross-ancestry meta-analysis were selected for comparison. The most significant associations are shown if overlapping variants exist (interval < 500 kb). Variants with significant heterogeneity ($P_{\text{het}} < 0.05$) are denoted by the black marker edges. The grey dashed line represents the fitted linear regression line. **a**, 28 available variants (existing in both datasets). **b**, 27 available variants. **c**, 27 available variants. **d**, 27 available variants with MAF > 0.01. **e**, 26 available variants with MAF > 0.01. **f**, 26 available variants with MAF > 0.01.

Table 10. Summary of publicly available summary statistics used in meta-analyses and effect size comparisons.

Cohort	Phenotype	Phenotype IDs (original)	Number of Cases	Number of Controls	N	Methods	Imputation Rsq/INFO threshold	Genome Build	Number of Variants	Covariates	URL
BBJ1	GC	Gastric Cancers	6,563	195,745	202,308	SAIGE	0.7	GRCh37	8,678,256	sex, age, PC1-5	http://jenger.riken.jp/61/
UKB	PUD	131591, 131593, 131595, 131597	16,666	439,661	456,327	BOLT-LMM	0.3	rsID only	8,546,066	sex, age PC1-20	https://www.nature.com/articles/s41467-021-21280-7#Sec12
UKB	DU	531.3	3,002	401,525	404,527	SAIGE	N.A	GRCh37	28,309,970	sex, age, PC1-4	https://pheweb.org/UKB-SAIGE/about
UKB	GU	531.2	4,109	401,525	405,634	SAIGE	N.A	GRCh37	28,320,186	sex, age, PC1-4	https://pheweb.org/UKB-SAIGE/about
FinnGen	PUD	K11_GASTRODUOLC	5,627	225,008	230,635	SAIGE	0.6	GRCh38	16,355,199	sex, age, PC1-10, genotyping batch	https://r6.finnngen.fi/
FinnGen	DU	K11_DULC	2,151	225,008	227,159	SAIGE	0.6	GRCh38	16,355,191	sex, age, PC1-10, genotyping batch	https://r6.finnngen.fi/
FinnGen	GU	K11_GULC	3,601	225,008	228,609	SAIGE	0.6	GRCh38	16,355,197	sex, age, PC1-10, genotyping batch	https://r6.finnngen.fi/

Table 11. Significant loci associated with PUD or PUD subtypes from inverse-variance-weighted cross-ancestry meta-analysis.

* If the significant loci were novel, the analysis which first identified the loci was shown. BBJ1-180, discovery GWAS. EAS, East Asian-specific meta-analysis. Directions of effect sizes for each lead variant from GWASs were shown in the order of BBJ1-180K, BBJ1-12K, BBJ2-42K, TMM-50K, UKB, FinnGen. Cross-ancestry, cross-ancestry meta-analysis.

Phenotype	rsID	CHR	POS (GRCh37)	Region	Nearest Gene	EA	NEA	EAF	BETA	SE	P	N	DIRECTION	HetISq	HetP	Novel or known locus*	Index of novel locus at the stage
PUD	rs1345894981	1	155161794	exonic	MUC1	T	C	0.0479	0.135	0.019	2.543E-12	726741	+-----?	15.5	0.316	Known	
PUD	rs59781317	1	156016356	intronic	UBQLN4	G	A	0.5106	0.044	0.008	3.511E-08	957376	+++++	17.7	0.2992	Cross-ancestry	1
PUD	rs184426772	2	182502860	intronic	CERKL	C	T	0.0082	0.796	0.141	1.601E-08	97523	?+---?	0.0	0.9949	EAS	
PUD	rs11692085	2	219963550	intronic	NHEJ1	C	C	0.4613	0.051	0.007	2.575E-12	957376	+-----	16.1	0.3103	EAS	
PUD	rs4685405	3	16981683	intronic	PLCL2	T	G	0.4085	0.05	0.007	9.836E-12	957376	+++++	0.0	0.4383	BBJ1-180K	
PUD	rs1290790	3	169091569	intronic	MECOM	T	A	0.3316	0.051	0.008	4.913E-11	957376	+++++	70.3	0.004785	BBJ1-180K	
PUD	rs76094691	4	124537967	intergenic	LINC01091	T	C	0.2797	0.059	0.01	1.29E-08	957376	+++++	0.0	0.4904	EAS	
PUD	rs3805497	5	40746885	intronic	TTC33	T	A	0.4669	0.076	0.007	5.32E-25	726741	++++?	0.0	0.5624	BBJ1-180K	
PUD	rs1801020	5	176836532	UTR5	F12	G	A	0.5118	-0.05	0.007	1.378E-11	957376	+----	4.9	0.3856	BBJ1-180K	
PUD	rs146095444	6	131582218	intronic	AKAP7	T	C	0.0217	0.215	0.039	2.782E-08	270414	++++?	44.9	0.142	EAS	
PUD	rs416879	6	160774838	intronic	SLC22A3	G	A	0.3833	-0.04	0.007	2.479E-09	957376	-----	0.0	0.6018	Cross-ancestry	2
PUD	rs72607744	7	127560541	intronic	SND1	A	G	0.0951	0.148	0.019	4.538E-15	270414	++++?	0.0	0.4818	BBJ1-180K	
PUD	rs2920293	8	143765414	intergenic	PSCA	G	C	0.6079	-0.2	0.008	4.42E-126	501049	----?	89.5	1.05E-07	Known	
PUD	rs10992997	9	96701663	intergenic	BARX1	G	A	0.3226	0.045	0.008	1.411E-08	957376	+++++	0.0	0.7644	EAS	
PUD	rs8176719	9	136132908	exonic	ABO	TC	T	0.4151	-0.07	0.007	5.757E-24	957376	-----	0.0	0.7763	Known	
PUD	rs78459074	11	1029905	intronic	MUC6	G	A	0.176	-0.06	0.01	7.603E-09	957376	-----	73.6	0.001989	Known	
PUD	rs10500661	11	6273744	intergenic	CCKBR	C	T	0.1622	0.101	0.009	7.745E-27	957376	+++++	0.0	0.4339	Known	
PUD	rs9581957	13	28557889	intronic	URAD	T	C	0.2642	0.071	0.008	2.203E-20	957376	+++++	14.0	0.3251	Known	
PUD	rs34074411	17	39867248	intergenic	GAST	T	C	0.5355	0.07	0.007	1.069E-22	957376	+++++	0.0	0.4275	Known	
PUD	rs11665674	19	49196275	intergenic	FUT2	G	A	0.4467	0.086	0.011	1.472E-15	501049	++++?	0.0	0.5176	Known	
PUD	rs6123837	20	57465571	intronic	GNAS	A	G	0.4787	0.052	0.007	2.12E-14	957376	+++++	1.7	0.4053	BBJ1-180K	
PUD	rs12625329	20	62709274	intronic	RGS19	A	G	0.5541	0.04	0.007	8.088E-09	957376	+++++	22.2	0.2671	Cross-ancestry	3
PUD	rs11416248	22	25008477	intronic	GGT1	CT	C	0.3207	-0.08	0.011	2.1E-12	270414	----?	0.0	0.6317	EAS	
DU	rs141625351	1	155133066	intergenic	KRTCAP2	G	GGT	0.9026	-0.17	0.025	1.69E-11	252639	--+??	74.3	0.008521	Known	
DU	rs11692085	2	219963550	intronic	NHEJ1	T	C	0.4039	0.077	0.013	7.09E-10	884325	+-----	3.4	0.3947	EAS	
DU	rs6783769	3	16966423	intronic	PLCL2	T	G	0.4877	0.081	0.014	2.31E-09	252639	++++?	25.3	0.2596	BBJ1-180K	
DU	rs60154219	3	169136565	intronic	MECOM	G	C	0.2677	-0.09	0.014	8.54E-11	479798	----?	15.1	0.3183	BBJ1-180K	
DU	rs34742353	4	47634802	intronic	CORIN	A	AT	0.8248	0.107	0.018	6.16E-09	252639	++++?	0.0	0.4916	EAS	
DU	rs3805497	5	40746885	intronic	TTC33	T	A	0.5211	0.11	0.013	1.12E-18	657166	++++?	9.0	0.355	BBJ1-180K	
DU	rs41265804	6	29924159	intergenic	HLA-A	G	C	0.4565	-0.078	0.0129	1.082E-09	479798	----?	38.4	0.1654	BBJ1-180K	
DU	rs2481036	6	160705005	intergenic	SLC22A2	C	T	0.2041	-0.08	0.014	2.84E-08	884325	+----	26.3	0.237	Cross-ancestry	
DU	rs61160304	7	127249659	downstream	PAX4	T	C	0.0985	0.2	0.026	4.77E-15	252639	+---?	23.6	0.2697	BBJ1-180K	
DU	rs71514093	8	143763259	intronic	PSCA	T	TG	0.3648	0.416	0.0142	1.07E-188	252639	++++?	0.0	0.6594	Known	
DU	rs7470279	9	80607789	intronic	GNAQ	T	A	0.3467	0.089	0.014	1.56E-10	479798	++++?	22.5	0.2711	BBJ1-180K	
DU	rs8176719	9	136132908	exonic	ABO	TC	T	0.449	-0.1	0.012	4.16E-17	479798	----?	67.6	0.01504	Known	
DU	rs10500661	11	6273744	intergenic	CCKBR	C	T	0.1415	0.145	0.017	6.79E-17	884325	+++++	66.4	0.01083	Known	
DU	rs3019776	11	68826155	intronic	TPCN2	G	A	0.5428	0.095	0.016	1.38E-09	479798	++++?	6.1	0.372	EAS	
DU	rs148675590	13	28567808	intergenic	URAD	A	AAAC	0.191	0.147	0.018	4.67E-16	252639	++++?	0.0	0.6429	Known	
DU	rs511893	16	88990941	intronic	CBFA2T3	G	T	0.3079	-0.09	0.016	6.39E-09	479798	+--?	25.4	0.2524	Cross-ancestry	4
DU	rs34074411	17	39867248	intergenic	GAST	T	C	0.5695	0.1	0.012	1.10E-15	884325	+++++	0.0	0.7185	Known	
DU	rs2642030	17	65605075	intronic	PITPNC1	G	A	0.4793	0.068	0.012	2.00E-08	884325	+-----	11.4	0.3428	Cross-ancestry	5
DU	rs11665674	19	49196275	intergenic	FUT2	G	A	0.4466	0.136	0.016	3.98E-17	479798	++++?	0.0	0.8538	Known	
DU	rs6117384	20	6673542	intergenic	LINC01713	C	T	0.7336	0.089	0.013	6.32E-12	884325	+-----	6.4	0.3758	EAS	
DU	rs5842244	20	57471283	intronic	GNAS	A	AT	0.2879	-0.11	0.015	1.28E-13	252639	+--?	57.3	0.07108	BBJ1-180K	
DU	rs8126001	20	62711459	UTR5	OPRL1	T	C	0.6238	0.072	0.013	3.03E-08	657166	++++?	0.0	0.5958	Cross-ancestry	3
DU	rs375244580	22	24992953	intronic	GGT1	A	AT	0.316	-0.09	0.015	4.47E-10	252639	----?	0.0	0.7732	EAS	
DU	rs7288137	22	50458020	intronic	TLL8	A	G	0.2079	0.081	0.014	7.73E-09	884325	+++++	0.0	0.4736	Cross-ancestry	6
GU	rs4685405	3	16981683	intronic	PLCL2	T	G	0.4578	0.056	0.009	2.765E-09	895890	+++++	0.0	0.8909	BBJ1-180K	
GU	rs77132614	3	169088553	intronic	MECOM	T	C	0.0971	0.101	0.018	3.396E-08	490256	++++?	10.9	0.3439	BBJ1-180K	
GU	rs546238399	5	40790627	intronic	PRKAA1	CTCT	C	0.6131	0.069	0.011	8.634E-10	261647	++++?	0.0	0.674	BBJ1-180K	
GU	rs17866912	7	126945795	intergenic	GRM8	G	T	0.0983	0.147	0.023	3.286E-10	261647	++++?	0.0	0.4998	BBJ1-180K	
GU	rs143484304	8	143760366	intergenic	PSCA	T	G	0.6342	-0.140	0.0108	1.329E-38	261647	----?	78.4	0.003076	Known	
GU	rs8176719	9	136132908	exonic	ABO	TC	T	0.451	-0.06	0.01	6.698E-09	490256	----?	0.0	0.8281	Known	
GU	rs749161312	13	28499583	UTR3	PDX1	CCCCTCCTCT	C	0.595	-0.07	0.012	1.011E-08	261647	----?	0.0	0.7334	Known	
GU	rs60346985	17	39920786	intronic	JUP	GGAAA	G	0.4752	0.064	0.011	6.475E-09	261647	++++?	33.7	0.2101	Known	
GU	rs8114689	20	57476890	intronic	GNAS	A	G	0.5925	0.064	0.011	1.83E-09	261647	++++?	0.0	0.7388	BBJ1-180K	
GU	rs11416248	22	25008477	intronic	GGT1	CT	C	0.3208	-0.07	0.013	7.512E-09	261647	----?	0.0	0.6567	EAS	

Table 12. Significant loci associated with PUD or PUD subtypes from cross-ancestry meta-regression using MR-MEGA.

* If the significant loci were novel, the analysis which first identified the loci was shown. BBJ1-180, discovery GWAS. EAS, East Asian-specific meta-analysis. Cross-ancestry, cross-ancestry meta-analysis. Directions of effect sizes for each lead variant from GWASs were shown in the order of BBJ1-180K, BBJ1-12K, BBJ2-42K, TMM-50K, UKB, and FinnGen.

Phenotype	rsID	Region	Nearest Gene	CHR	POS (GRCh37)	EA	NEA	EA	P	Number of cohorts	N	Directions	P (ancestry)	P (residual_het)	Novel or known locus*
PUD	rs1345894981	exonic	MUC1	1	155161794 T		C	0.0426942	3.769E-11	5	726741	+----?	0.283819	0.3308	Known
PUD	rs11692085	intronic	NHEJ1	2	219963550 T		C	0.549306	2.714E-11	6	957376	+-----	0.200643	0.432514	EAS
PUD	rs4685405	intronic	PLCL2	3	16981683 T		G	0.276387	4.708E-11	6	957376	+++++	0.0945369	0.991571	BBJ-180K
PUD	rs1290790	intronic	MECOM	3	169091569 T		A	0.352724	1.095E-11	6	957376	+++++	0.00443309	0.110797	BBJ-180K
PUD	rs7705504	intronic	TTC33	5	40739261 T		C	0.340395	1.79E-23	6	957376	+++++	0.000902994	0.711068	BBJ-180K
PUD	rs1801020	UTR5	F12	5	176836532 G		A	0.631956	1.633E-10	6	957376	+-----	0.241556	0.491093	BBJ-180K
PUD	rs141001296	intergenic	SLC22A2	6	160711815 C		T	0.234535	3.133E-08	6	957376	--+-	0.262322	0.351057	Cross-ancestry
PUD	rs17866443	intergenic	ZNF800	7	127058953 A		C	0.093404	3.565E-12	5	501049	+---?	0.0612415	0.116768	BBJ-180K
PUD	rs35589574	intergenic	MIR129-1	7	127846470 T		C	0.16915	3.247E-08	5	501049	++++?	0.127013	0.845046	BBJ-180K
PUD	rs2976391	intronic	PSCA	8	143762724 A		C	0.391829	3.443E-65	6	957376	+++++	4.93706e-12	0.0253642	Known
PUD	rs8176719	exonic	ABO	9	136132908 TC		T	0.39871	4.127E-22	6	957376	-----	0.687445	0.625536	Known
PUD	rs78459074	intronic	MUC6	11	1029905 G		A	0.149573	6.862E-11	6	957376	-----	0.000210096	0.575656	Known
PUD	rs10500661	intergenic	CCKBR	11	6273744 C		T	0.17081	4.689E-25	6	957376	+++++	0.514495	0.317501	Known
PUD	rs9581957	intronic	URAD	13	28557889 T		C	0.286992	1.426E-19	6	957376	+++++	0.0717945	0.909341	Known
PUD	rs34074411	intergenic	GAST	17	39867248 T		C	0.501953	1.915E-21	6	957376	+++++	0.179464	0.689136	Known
PUD	rs8103840	intronic	FUT1	19	49254955 T		C	0.50636	6.17E-15	6	957376	-----	0.0081837	0.280769	Known
PUD	rs6123837	intronic	GNAS	20	57465571 A		G	0.420828	2.429E-13	6	957376	+++++	0.184652	0.635038	BBJ-180K
PUD	rs28509371	intronic	GGT1	22	24997846 C		T	0.331864	4.263E-12	6	957376	-----	0.0274061	0.674038	EAS
GU	rs7716285	downstream	TTC33	5	40711361 G		A	0.343922	8.3E-09	6	895890	+++++	0.00706057	0.795994	BBJ-180K
GU	rs56370107	intergenic	ARC	8	143707530 T		C	0.545995	1.643E-21	6	895890	----+	1.08972e-10	0.812821	Known
GU	rs6128461	intronic	GNAS	20	57477090 C		T	0.622828	4.143E-08	6	895890	+++++	0.000210996	0.708476	BBJ-180K
DU	rs6668815	intergenic	DPM3	1	155124985 G		A	0.594054	5.126E-11	5	657166	----?	1.38495e-06	0.832165	Known
DU	rs59781317	intronic	UBQLN4	1	156016356 G		A	0.405675	3.085E-08	6	884325	+++++	0.0171213	0.927226	Cross-ancestry
DU	rs11692085	intronic	NHEJ1	2	219963550 T		C	0.549339	7.583E-09	6	884325	+++++	0.259108	0.479727	EAS
DU	rs1156336	intronic	PLCL2	3	16967221 T		A	0.403007	4.485E-08	6	884325	+++++	0.000574679	0.307051	BBJ-180K
DU	rs6780457	intronic	MECOM	3	169137059 T		C	0.194167	4.726E-10	6	884325	-----	0.104386	0.662083	BBJ-180K
DU	rs3805497	intronic	TTC33	5	40746885 T		A	0.375735	1.506E-17	5	657166	++++?	0.164966	0.672744	BBJ-180K
DU	rs1611713	intergenic	HCP5B	6	29829251 C		A	0.285888	3.182E-08	6	884325	-----	0.016337	0.392146	BBJ-180K
DU	rs17866443	intergenic	ZNF800	7	127058953 A		C	0.0910201	1.434E-11	5	479798	+---?	0.0261802	0.0089082	BBJ-180K
DU	rs2976391	intronic	PSCA	8	143762724 A		C	0.38812	9.751E-95	6	884325	+++++	2.8473e-11	0.80245	Known
DU	rs10870013	intergenic	GNAQ	9	80649703 C		A	0.237324	4.579E-10	6	884325	+++++	0.000391716	0.568393	BBJ-180K
DU	rs8176719	exonic	ABO	9	136132908 TC		T	0.452759	3.31E-17	5	479798	----?	0.0103141	0.203402	Known
DU	rs10500661	intergenic	CCKBR	11	6273744 C		T	0.168496	1.464E-17	6	884325	+++++	0.00278733	0.37232	Known
DU	rs3019776	intronic	TPCN2	11	68826155 G		A	0.737039	1.333E-08	5	479798	++++?	0.241624	0.491935	Known
DU	rs9581957	intronic	URAD	13	28557889 T		C	0.288996	1.025E-15	6	884325	+++++	0.00653199	0.336365	Known
DU	rs34074411	intergenic	GAST	17	39867248 T		C	0.504127	2.902E-14	6	884325	+++++	0.391945	0.799692	Known
DU	rs11665674	intergenic	FUT2	19	49196275 G		A	0.23773	2.047E-15	5	479798	++++?	0.704115	0.725234	Known
DU	rs6117384	intergenic	LINC01713	20	6673542 C		T	0.726362	1.635E-10	6	884325	+-----	0.519042	0.258407	EAS
DU	rs6026578	ncRNA intronic	LOC101927932	20	57463472 G		C	0.651721	9.769E-13	6	884325	+++++	0.0215192	0.0690219	BBJ-180K
DU	rs5751901	intronic	GGT1	22	24992266 C		T	0.333422	3.239E-09	6	884325	-----	0.0325959	0.658388	EAS

Table 13. Effect comparison of PUD signals between East Asians and Europeans.

5 lead variants were not available for effect comparison. *LD proxies will be used If the lead variants were not available in the summary statistics.

rsID	CHR	POS (GRCh37)	Region	Nearest Gene	EA	NEA	EAS				EUR				Q	HetP	HetP<0.05	Δ BETA
							EAF	BETA	SE	P	EAF	BETA	SE	P				
rs1345894981	1	155161794	exonic	MUC1	T	C	0.0169	0.0843	0.0387	2.94E-02	0.0582	0.1522	0.0223	8.79E-12	2.31	1.28E-01	False	-0.0679
rs59781317	1	156016356	intronic	UBQLN4	G	A	0.773	0.0533	0.0113	2.21E-06	0.2545	0.0342	0.0111	2.12E-03	1.45	2.28E-01	False	0.0191
rs11692085	2	219963550	intronic	NHEJ1	T	C	0.2706	0.0633	0.0102	5.25E-10	0.6559	0.0379	0.0103	2.38E-04	3.07	7.97E-02	False	0.0254
rs4685405	3	16981683	intronic	PLCL2	T	G	0.5241	0.0603	0.009	1.63E-11	0.1775	0.0289	0.0127	2.27E-02	4.07	4.37E-02	True	0.0314
rs13086914	3	171491650	intronic	PLD1	G	T	0.2044	-0.0537	0.0114	2.28E-06	0.3896	-0.0193	0.01	5.41E-02	5.15	2.33E-02	True	-0.0344
rs28424168	4	47634802	intronic	CORIN	A	C	0.1796	-0.0458	0.0116	7.45E-05	0.2819	-0.0121	0.0108	2.64E-01	4.52	3.35E-02	True	-0.0337
rs2553380	4	124538120	intergenic	LINC01091	C	T	0.648	-0.0602	0.0105	1.00E-08	0.837	0.0318	0.0269	2.37E-01	10.15	1.44E-03	True	-0.092
rs3805497	5	40746885	intronic	TTC33	T	A	0.5782	0.0828	0.0091	5.64E-20	0.247	0.0627	0.0127	7.93E-07	1.66	1.98E-01	False	0.0201
rs1801020	5	176836532	UTR5	F12	G	A	0.3462	-0.0412	0.0093	9.25E-06	0.7453	-0.0577	0.011	1.74E-07	1.31	2.52E-01	False	0.0165
rs9467684	6	26162679	intronic	HIST1H2BD	A	G	0.0605	0.0688	0.0184	1.89E-04	0.0149	-0.0277	0.0402	4.91E-01	4.76	2.91E-02	True	0.0965
rs416879	6	160774838	intronic	SLC22A3	G	A	0.2822	-0.0315	0.0099	1.41E-03	0.4809	-0.0509	0.0097	1.68E-07	1.96	1.62E-01	False	0.0194
rs2294008	8	143761931	UTR5	PSCA	T	C	0.6354	-0.2177	0.0092	1.05E-123	0.4568	-0.0899	0.0098	4.37E-20	90.40	1.95E-21	True	-0.1278
rs7470279	9	80607789	intronic	GNAQ	T	A	0.3688	0.0419	0.0097	1.47E-05	0.154	0.0544	0.0273	4.63E-02	0.19	6.66E-01	False	-0.0125
rs10992997	9	96701663	intergenic	BARX1	G	A	0.1377	0.0524	0.0137	1.30E-04	0.4188	0.0418	0.0099	2.32E-05	0.39	5.31E-01	False	0.0106
rs8176719	9	136132908	exonic	ABO	TC	T	0.4486	-0.0667	0.0089	7.75E-14	0.3711	-0.0689	0.0102	1.35E-11	0.03	8.71E-01	False	0.0022
rs78459074	11	1029905	intronic	MUC6	G	A	0.2104	-0.0251	0.0124	4.24E-02	0.1245	-0.101	0.0152	2.95E-11	14.97	1.09E-04	True	0.0759
rs10500661	11	6273744	intergenic	CCKBR	C	T	0.0901	0.0902	0.0156	7.48E-09	0.203	0.1067	0.0118	1.16E-19	0.71	3.99E-01	False	-0.0165
rs3019776	11	68826155	intronic	TPCN2	G	A	0.5294	0.053	0.0105	4.42E-07	0.968	0.0132	0.058	8.20E-01	0.46	5.00E-01	False	0.0398
rs9581957	13	28557889	intronic	URAD	T	C	0.1843	0.0851	0.0116	2.55E-13	0.3266	0.0602	0.0103	4.53E-09	2.58	1.08E-01	False	0.0249
rs511893	16	88990941	intronic	CBFA2T3	G	T	0.2973	-0.0529	0.012	1.11E-05	0.335	-0.0302	0.021	1.50E-01	0.88	3.48E-01	False	-0.0227
rs34074411	17	39867248	intergenic	GAST	T	C	0.6342	0.078	0.0104	6.96E-14	0.4471	0.0633	0.0099	1.41E-10	1.05	3.06E-01	False	0.0147
rs2642030	17	65605075	intronic	PITPNC1	G	A	0.562	0.0471	0.0093	3.72E-07	0.2497	0.0181	0.0112	1.06E-01	3.97	4.64E-02	True	0.029
rs11665674	19	49196275	intergenic	FUT2	G	A	0.4486	0.0853	0.0107	1.70E-15	0.0037	0.1315	0.1615	4.16E-01	0.08	7.75E-01	False	-0.0462
rs6117384	20	6673542	intergenic	LINC01713	C	T	0.7416	0.0465	0.0105	8.95E-06	0.7258	0.0162	0.0109	1.38E-01	4.01	4.53E-02	True	0.0303
rs6123837	20	57465571	intronic	GNAS	A	G	0.5795	0.0627	0.0093	1.50E-11	0.3576	0.04	0.0102	8.42E-05	2.70	1.00E-01	False	0.0227
rs12625329	20	62709274	intronic	RGS19	A	G	0.6556	0.0512	0.0099	2.38E-07	0.4565	0.0294	0.0097	2.54E-03	2.47	1.16E-01	False	0.0218
rs3859862	22	25008477	intronic	GGT1	G	A	0.3173	-0.0697	0.0102	1.02E-11	0.3409	-0.0308	0.0103	2.76E-03	7.20	7.29E-03	True	-0.0389
rs7288137	22	50458020	intronic	TTL8	A	G	0.2017	0.0325	0.0114	4.41E-03	0.2201	0.0229	0.0117	4.98E-02	0.35	5.57E-01	False	0.0096

Table 14. Cross-ancestry genetic correlation of PUD and PUD subtypes.

Summary statistics for HapMap 3 SNPs from EAS- or EUR-specific meta-analysis were used.

Phenotype	Genetic correlation	Estimates	SE	Z	P
PUD	genetic-impact correlation (pgi)	0.658	0.09	3.61	3.05E-04
	genetic-effect correlation (pge)	0.694	0.10	3.18	1.46E-03
GU	genetic-impact correlation (pgi)	0.457	0.20	2.68	7.30E-03
	genetic-effect correlation (pge)	0.467	0.21	2.58	9.87E-03
DU	genetic-impact correlation (pgi)	0.729	0.16	1.66	9.65E-02
	genetic-effect correlation (pge)	0.778	0.16	1.39	1.63E-01

3.5 Conditional and fine-mapping analysis

To further explore the secondary signals at the identified loci, a stepwise conditional analysis using GCTA-COJO³⁰ with an in-sample LD reference for EAS was carried out in East Asians. Four additional independent signals reaching genome-wide significance ($P < 5.0 \times 10^{-8}$) for PUD and three independent signals at the *PSCA* locus for DU were identified. *PSCA* locus had the largest number of independent associations (three for PUD, four for DU, and two for GU and BU) (Table 15). It is noteworthy that near *CDX2* and *GAST* loci, two of the previously reported loci in European individuals¹², independent signals at *PDX1* (Figure 13a) and *JUP2* (Figure 13b) loci were detected, respectively.

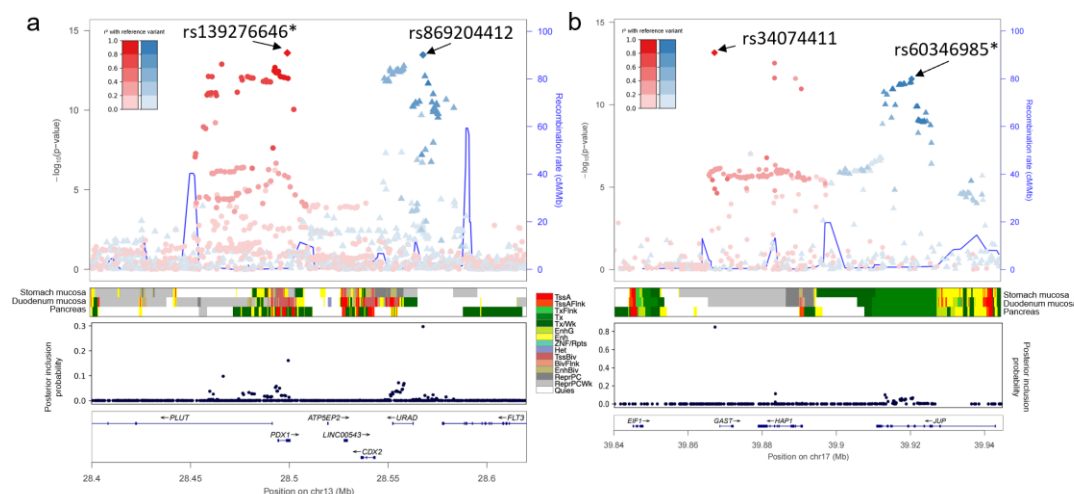


Figure 13. EAS-specific secondary signals at *PDX1* and *JUP*.

a. Regional plot at *PDX1*-*CDX2* locus for PUD association from East Asian-specific meta-analysis. Variants are colored to match the lead SNPs in the highest LD, and the extent of LD with the lead variant is shown by a color gradient (red or blue). Posterior inclusion probability (PIP) was derived from fine-mapping analysis. Variants are plotted against GRCh37 (hg19). Chromatin states (core 15-state model) are shown for three related tissue types, namely stomach mucosa, duodenal mucosa, and pancreas. TssA, active transcription start site (TSS); TssAFlnk, flanking active TSS; TxFlnk, transcription at gene 5' and 3' ends; Tx, strong transcription; TxWk, weak transcription; EnhG, genic enhancers; Enh, enhancers; ZNF/Rpts, ZNF genes & repeats; Het, heterochromatin; TssBiv, bivalent/poised TSS; BivFlnk, flanking bivalent TSS/Enh; EnhBiv, bivalent enhancer; ReprPC, repressed PolyComb; ReprPCWk, weak repressed PolyComb; Quies, quiescent/low. b. Regional plot at *GAST*-*JUP* locus for PUD association from East Asian-specific meta-analysis.

To fine-map the causal variants, a fine-mapping analysis using SuSiE³¹ was conducted. In this specific analysis, the aim is to search for nonsynonymous variants in 95% credible sets to link the disease-associated loci to potential alteration of protein functions. A total of 10 nonsynonymous variants at six non-overlapping loci were identified, six of which were in novel loci for PUD and its subtypes (Table 16). Of those, rs2233580 (*PAX4*; p.R200H; CADD⁷⁴ score = 29.8) was also associated with type 2 diabetes. The variant was common (MAF > 0.05) in 1KG EAS but almost monozygotic in non-EAS populations. rs4745 (p.D159V; PIP = 0.05 for DU; CADD score = 15.2) in *EFNA1* was common in EAS and EUR and was associated with GC. This was the lead variant of cis-splicing quantitative trait loci (sQTL) for *EFNA1* in the stomach and was in high LD with rs407203721⁷⁵ (lead sQTL variant for *MUC1* in the stomach; $r^2 = 0.74$ in 1KG EAS). In addition to the missense variants, we found rs4390169 in the

credible set (upstream of *EFNA1*; PIP = 0.06 for DU; in high LD with rs4745, $r^2 = 0.99$ in 1KG EAS and EUR) to be the lead variant of cis-protein QTL (pQTL) in plasma for EFNA1⁷⁶.

Previous epidemiological studies suggest PUD is associated with Blood group O. In the credible sets of previously reported *ABO*³ and *FUT2*¹² loci for PUD, rs8176719 (lead variant at *ABO* locus) and rs1047781 (in the credible sets at *FUT2* locus; PIP = 0.63) were found. rs8176719 deletion resulted in the O allele, whereas rs1047781 (p.I140F) was an EAS-specific common variant (MAF = 0.439 in 1KG EAS), and its A-allele determined the FUT2 secretor status. To investigate the correlation of the ABO blood group and FUT2 secretor status with PUD, a logistic regression analysis was performed. The results showed that blood group O (OR = 1.14, $P = 6.0 \times 10^{-14}$) and non-secretor status (OR = 1.17, $P = 2.9 \times 10^{-11}$) were significantly correlated with a higher risk of PUD, which was consistent for all PUD subtypes (Table 17). To further investigate the potential interactions between blood group O and non-secretor status, logistic regression analysis including an interaction term was performed. However, significant interactions ($P < 0.05/6$) were not detected between blood group O and non-secretor status (Table 18)

Table 15. Independent variants from COJO analysis in East Asians.

Phenotype	rsID	Region	Nearest Gene	CHR	POS (GRCh37)	EA	EAF in original analysis	BETA in original analysis	SE in original analysis	P in original analysis	Estimated effective sample size	EAF in reference panel	BETA in joint analysis	SE in joint analysis	P in joint analysis	LD correlation between the SNP i and SNP i + 1
PUD	rs11676741	intronic	NHEJ1	2	219949913 A		0.7275	-0.0629	0.0101	4.73E-10	315131	0.722582	-0.0629	0.0101006	4.74E-10	0
	rs9824172	intronic	PLCL2	3	16974123 T		0.4734	-0.0597	0.0089	1.97E-11	322728	0.470854	-0.0597	0.00890061	1.98E-11	0
	rs1273886	intronic	MECOM	3	169091253 A		0.8714	-0.0889	0.0135	4.54E-11	312033	0.878926	-0.0889	0.0135009	4.56E-11	0
	rs2553380	intergenic	LINC01091	4	124538120 T		0.352	0.0602	0.0105	9.85E-09	253416	0.316734	0.0602	0.0105007	9.87E-09	0
	rs546238399	intronic	PRKAA1	5	40790627 CTCT		0.614	0.0885	0.0096	3.01E-20	291717	0.647706	0.0885	0.00960138	3.04E-20	0
	rs28947805	intergenic	ZNF800	7	127046841 A		0.0243	0.1956	0.037	1.25E-07	196336	0.0109317	0.204878	0.0370206	3.13E-08	-0.0322604
	rs72607744	intronic	SND1	7	127560541 A		0.0951	0.1477	0.0188	3.95E-15	209490	0.0590115	0.150951	0.0188119	1.02E-15	0
	rs2294010	intronic	PSCA	8	143762430 A		0.3678	0.2186	0.0092	8.48E-125	323286	0.371	0.17322	0.00981116	9.25E-70	0.392452
	rs200645096	intergenic	LY6K	8	143776719 A		0.6908	0.2202	0.0116	2.37E-80	221398	0.812327	0.146686	0.0122751	6.51E-33	-0.0360819
	rs2060474	intergenic	SLURP2	8	143842101 A		0.1468	-0.1467	0.0158	1.62E-20	203757	0.0700299	-0.101677	0.0159352	1.76E-10	0
	rs12347577	intronic	ZNF169	9	97031548 C		0.3224	-0.0536	0.0096	2.36E-08	316543	0.323027	-0.0536	0.00960046	2.36E-08	0
	rs8176719	exonic	ABO	9	136132908 T		0.5514	0.0667	0.0089	6.66E-14	325241	0.548659	0.0667	0.00890075	6.69E-14	0
	rs12792379	intergenic	CNGA4	11	6268546 A		0.0896	0.0917	0.0157	5.20E-09	316957	0.0875969	0.0917	0.0157008	5.21E-09	0
	rs749161312	UTR3	PDX1	13	28499583 CCCCTCC		0.5944	-0.0757	0.0099	2.07E-14	269678	0.675544	-0.06243	0.01014	7.42E-10	-0.223661
	rs148675590	intergenic	URAD	13	28567808 A		0.1924	0.0901	0.0119	3.69E-14	289605	0.167079	0.0739038	0.0121884	1.33E-09	0
	rs34074411	intergenic	GAST	17	39867248 T		0.6342	0.078	0.0104	6.38E-14	253954	0.534197	0.0694074	0.0105096	4.00E-11	-0.159514
	rs11869903	intronic	JUP	17	39913256 A		0.4946	-0.0605	0.009	1.79E-11	314739	0.491604	-0.051894	0.00909447	1.16E-08	0
	rs11665674	intergenic	FUT2	19	49196275 A		0.5514	-0.0853	0.0107	1.56E-15	224994	0.558077	-0.0853	0.0107015	1.58E-15	0
	rs6128461	intronic	GNAS	20	57477090 T		0.4028	-0.0656	0.0091	5.65E-13	319907	0.394731	-0.0656	0.00910072	5.67E-13	0
	rs11416248	intronic	GGT1	22	25008477 CT		0.3207	-0.0762	0.0109	2.73E-12	246200	0.258706	-0.0762	0.0109011	2.75E-12	0
DU	rs141625351	intergenic	KRTCAP2	1	155133066 G		0.9026	-0.1709	0.0254	1.72E-11	239214	0.925208	-0.1709	0.0254024	1.72E-11	0
	rs9288539	intronic	NHEJ1	2	219945274 A		0.2762	0.0921	0.0153	1.75E-09	289942	0.282432	0.0921	0.0153009	1.75E-09	0
	rs7637568	intronic	PLCL2	3	16965216 T		0.5045	-0.0818	0.0135	1.37E-09	297828	0.502253	-0.0818	0.0135008	1.37E-09	0
	rs79928271	intronic	MECOM	3	169139475 AA		0.2771	-0.1014	0.0151	1.88E-11	297068	0.276808	-0.1014	0.0151011	1.88E-11	0
	rs34742353	intronic	CORIN	4	47634802 A		0.8247	0.1069	0.0184	6.26E-09	277220	0.85213	0.1069	0.0184011	6.27E-09	0
	rs546238399	intronic	PRKAA1	5	40790627 CTCT		0.6136	0.1279	0.0146	1.95E-18	268435	0.647706	0.1279	0.0146021	1.97E-18	0
	rs9467684	intronic	HIST1H2BD	6	26162679 A		0.0597	0.1651	0.0283	5.41E-09	301806	0.060431	0.1651	0.0283015	5.42E-09	0
	rs61160304	downstream	PAX4	7	127249659 T		0.0985	0.2001	0.0255	4.26E-15	234961	0.0946443	0.2001	0.0255033	4.29E-15	0
	rs2294008	UTR5	PSCA	8	143761931 T		0.6381	-0.4143	0.0141	9.06E-190	294717	0.628871	-0.321875	0.0150797	4.35E-101	-0.392458
	rs200645096	intergenic	LY6K	8	143776719 A		0.6897	0.4185	0.0177	1.36E-123	201838	0.812327	0.276366	0.0187504	3.61E-49	-0.0360819
	rs2060474	intergenic	SLURP2	8	143842101 A		0.1471	-0.2668	0.0241	1.74E-28	186106	0.0700299	-0.185606	0.024318	2.30E-14	-0.0182926
	rs564980087	upstream	LY6D	8	143868011 T		0.8722	-0.2607	0.0296	1.28E-18	138874	0.952892	-0.185501	0.0297254	4.36E-10	0
	rs10870013	intergenic	GNAQ	9	80649703 A		0.647	-0.0831	0.0148	1.97E-08	271230	0.642438	-0.0831	0.0148008	1.97E-08	0
	rs8176719	exonic	ABO	9	136132908 T		0.5526	0.0924	0.0135	7.68E-12	301127	0.548659	0.0924	0.013501	7.71E-12	0
	rs3019776	intronic	TPCN2	11	68826155 A		0.4699	-0.093	0.0159	4.94E-09	215459	0.451742	-0.093	0.0159012	4.96E-09	0
	rs148675590	intergenic	URAD	13	28567808 A		0.191	0.1468	0.0181	5.04E-16	268007	0.167079	0.1468	0.0181022	5.08E-16	0
	rs34074411	intergenic	GAST	17	39867248 T		0.6366	0.1107	0.0157	1.78E-12	237930	0.534197	0.1107	0.0157016	1.79E-12	0
	rs11665674	intergenic	FUT2	19	49196275 A		0.5516	-0.1363	0.0162	3.98E-17	208990	0.558077	-0.1363	0.0162027	4.02E-17	0
	rs6038586	intergenic	LINC01713	20	6673320 A		0.2561	-0.0922	0.0158	5.36E-09	285299	0.250793	-0.0922	0.0158009	5.38E-09	0
	rs5842244	intronic	GNAS	20	57471283 A		0.2879	-0.1126	0.0152	1.28E-13	286443	0.277994	-0.1126	0.0152014	1.29E-13	0
	rs5751904	intronic	GGT1	22	25000229 A		0.681	0.0988	0.0155	1.84E-10	259969	0.753184	0.0988	0.0155012	1.85E-10	0
GU	rs77132614	intronic	MECOM	3	169088553 T		0.0982	0.1013	0.0184	3.68E-08	280370	0.0637258	0.1013	0.018401	3.69E-08	0
	rs6860328	intronic	TTC33	5	40729974 T		0.4238	-0.0655	0.0106	6.44E-10	306362	0.418235	-0.0655	0.0106006	6.46E-10	0
	rs17866912	intergenic	GRM8	7	126945795 T		0.9017	-0.1471	0.0234	3.25E-10	173177	0.963182	-0.1471	0.0234026	3.27E-10	0
	rs34635647	intergenic	PSCA	8	143756890 A		0.3767	0.1431	0.0109	2.26E-39	301190	0.377182	0.119675	0.0115583	4.01E-25	0.392687
	rs200645096	intergenic	LY6K	8	143776719 A		0.689	0.1366	0.0135	4.57E-24	215170	0.812327	0.0874065	0.0143146	1.02E-09	0
	rs749161312	UTR3	PDX1	13	28499583 CCCCTCC		0.595	-0.0665	0.0116	9.88E-09	259234	0.675544	-0.0665	0.0116007	9.90E-09	0
	rs11869903	intronic	JUP	17	39913256 A		0.4947	-0.0613	0.0105	5.28E-09	305012	0.491604	-0.0613	0.0105006	5.29E-09	0
	rs6128461	intronic	GNAS	20	57477090 T		0.4031	-0.0639	0.0106	1.66E-09	310927	0.394731	-0.0639	0.0106006	1.66E-09	0
BU	rs5760499	intronic	GGT1	22	25009875 A		0.3052	-0.0774	0.0134	7.64E-09	220758	0.201823	-0.0774	0.013401	7.66E-09	0
	rs17866912	intergenic	GRM8	7	126945795 T		0.9022	-0.3425	0.057	1.87E-09	161651	0.963123	-0.3425	0.0570062	1.88E-09	0
	rs2294008	UTR5	PSCA	8	143761931 T		0.642	-0.4326	0.0268	1.30E-58	280526	0.628871	-0.361479	0.0283788	3.65E-37	-0.392458
	rs200645096	intergenic	LY6K	8	143776719 A		0.6873	0.4156	0.0332	5.95E-36	195509	0.812327	0.268894	0.0351535	2.02E-14	0
	rs6026576	ncRNA intronic	LOC101927932	20	57459167 A		0.3387	-0.1627	0.0271	1.93E-09	281745	0.331714	-0.1627	0.0271017	1.93E-09	0

Table 16. Nonsynonymous variants in credible sets from fine-mapping analysis using SuSiE in East Asians.

Locus Index	Phenotype	rsID	CHR	POS (GRCh37)	EA	NEA	MAF	BETA	SE	Posterior Inclusion Probability	Credible Set Number	Region	Gene	Consequence
1	DU	rs4745	1	155106227	A	T	0.0917	-0.1577	0.0237	0.052039283	1	exonic	EFNA1	nonsynonymous SNV
2	DU	rs2233580	7	127253550	C	T	0.0896	-0.1957	0.0255	0.127790773	1	exonic	PAX4	nonsynonymous SNV
2	GU	rs2233580	7	127253550	C	T	0.0901	-0.1101	0.0193	0.019518757	1	exonic	PAX4	nonsynonymous SNV
2	PUD	rs2233580	7	127253550	C	T	0.0904	-0.1263	0.0166	0.040158162	1	exonic	PAX4	nonsynonymous SNV
2	BU	rs2233580	7	127253550	C	T	0.0891	-0.2747	0.0483	0.151833684	1	exonic	PAX4	nonsynonymous SNV
2	DU	rs3824004	7	127253551	G	T	0.0285	-0.1634	0.0411	0.010215591	2	exonic	PAX4	nonsynonymous SNV
3	PUD	rs12236219	9	97062981	C	T	0.2993	0.05	0.0097	0.00407495	1	exonic	ZNF169	nonsynonymous SNV
3	PUD	rs10993163	9	97082699	G	A	0.2964	0.0505	0.0098	0.004043119	1	exonic	NUTM2F	nonsynonymous SNV
3	PUD	rs202099818	9	97087823	G	C	0.2967	0.0505	0.0101	0.001911238	1	exonic	NUTM2F	nonsynonymous SNV
3	PUD	rs75315722	9	97087857	G	A	0.2927	0.0511	0.0101	0.002549444	1	exonic	NUTM2F	nonsynonymous SNV
4	DU	rs8176719	9	136132908	T	TC	0.4474	0.0924	0.0135	0.162182726	1	exonic	ABO	Frameshift variant
4	PUD	rs8176719	9	136132908	T	TC	0.4486	0.0667	0.0089	0.164886271	1	exonic	ABO	Frameshift variant
5	DU	rs9579139	13	28552425	G	T	0.1841	-0.1413	0.0177	0.080543301	1	exonic	URAD	nonsynonymous SNV
5	PUD	rs9579139	13	28552425	G	T	0.1848	-0.0839	0.0116	0.033847896	2	exonic	URAD	nonsynonymous SNV
6	DU	rs1047781	19	49206631	A	T	0.3816	-0.1342	0.0163	0.457644885	1	exonic	FUT2	nonsynonymous SNV
6	PUD	rs1047781	19	49206631	A	T	0.3826	-0.0857	0.0108	0.632785925	1	exonic	FUT2	nonsynonymous SNV

Table 17. Logistic regression analysis for association of ABO blood group and FUT2 secretor status with peptic ulcers.

Blood group and secretor status		PUD			DU			GU			BU		
		OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P
Blood group	O vs A,B,AB	1.14	(1.12, 1.16)	1.65E-15	1.21	(1.18, 1.24)	9.77E-14	1.12	(1.1, 1.14)	1.76E-09	1.27	(1.21, 1.33)	1.27E-06
	A vs B,AB,O	0.96	(0.94, 0.97)	6.55E-03	0.92	(0.9, 0.95)	1.45E-03	0.96	(0.94, 0.98)	3.18E-02	0.86	(0.82, 0.9)	1.61E-03
	B vs A,AB,O	0.95	(0.93, 0.97)	4.74E-03	0.91	(0.89, 0.94)	3.09E-03	0.96	(0.94, 0.98)	6.69E-02	0.91	(0.86, 0.97)	1.22E-01
	AB vs A,B,O	0.92	(0.9, 0.95)	2.97E-03	0.95	(0.91, 0.99)	2.45E-01	0.92	(0.89, 0.95)	8.25E-03	1.02	(0.94, 1.1)	8.20E-01
Secretor status	Non-Secretor vs Secretor	1.16	(1.13, 1.19)	1.25E-11	1.28	(1.25, 1.33)	1.45E-13	1.12	(1.09, 1.14)	1.07E-05	1.31	(1.21, 1.38)	4.27E-05
Blood group and secretor status		PUD			DU			GU			BU		
		OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P
Blood group	O vs A,B,AB	1.14	(1.12, 1.15)	6.00E-14	1.21	(1.18, 1.24)	5.63E-13	1.12	(1.1, 1.14)	1.36E-08	1.28	(1.22, 1.35)	6.91E-07
	A vs B,AB,O	0.96	(0.94, 0.98)	1.09E-02	0.92	(0.9, 0.94)	1.17E-03	0.96	(0.95, 0.98)	5.12E-02	0.85	(0.81, 0.89)	9.19E-04
	B vs A,AB,O	0.95	(0.93, 0.97)	8.89E-03	0.92	(0.89, 0.95)	5.95E-03	0.96	(0.94, 0.98)	8.35E-02	0.91	(0.86, 0.97)	1.23E-01
	AB vs A,B,O	0.93	(0.9, 0.95)	5.07E-03	0.96	(0.92, 1.)	3.12E-01	0.92	(0.89, 0.95)	1.07E-02	1.02	(0.94, 1.1)	8.27E-01
Secretor status	Non-Secretor vs Secretor	1.17	(1.14, 1.19)	2.99E-11	1.29	(1.24, 1.33)	7.25E-13	1.12	(1.09, 1.15)	1.16E-05	1.31	(1.23, 1.4)	5.50E-05

Table 18. Logistic regression analysis for ABO-FUT2 interactions.

	Risk allele or factor		PUD			DU			GU			BU		
			OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P
Including related individuals	Risk Allele dosage	rs8176719	1.10	(1.08, 1.13)	1.18E-07	1.15	(1.12, 1.19)	2.08E-06	1.10	(1.07, 1.12)	1.80E-05	1.24	(1.17, 1.31)	2.78E-04
		rs1047781	1.15	(1.12, 1.17)	9.25E-08	1.21	(1.16, 1.26)	4.32E-06	1.14	(1.10, 1.17)	1.41E-05	1.30	(1.20, 1.40)	8.94E-04
		rs1047781 x rs8176719 interaction	0.97	(0.95, 0.98)	6.46E-02	0.95	(0.92, 0.98)	1.00E-01	0.96	(0.94, 0.98)	8.67E-02	0.90	(0.85, 0.95)	6.80E-02
	Risk factors	Non Secretor status	1.16	(1.12, 1.19)	1.16E-07	1.31	(1.25, 1.36)	1.85E-10	1.11	(1.08, 1.15)	6.87E-04	1.40	(1.29, 1.51)	2.27E-05
		Blood group O	1.14	(1.12, 1.16)	5.51E-13	1.22	(1.19, 1.25)	1.56E-12	1.12	(1.09, 1.14)	5.04E-08	1.31	(1.24, 1.38)	4.68E-07
		O x Non-Secretor interaction	1.02	(0.97, 1.07)	6.89E-01	0.96	(0.89, 1.03)	5.42E-01	1.02	(0.96, 1.08)	7.28E-01	0.81	(0.70, 0.93)	1.37E-01
	Risk allele or factor		PUD			DU			GU			BU		
			OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P	OR	95% C.I.	P
Including only unrelated individuals	Risk Allele dosage	rs8176719	1.10	(1.08, 1.12)	1.01E-06	1.15	(1.11, 1.18)	7.74E-06	1.09	(1.07, 1.12)	4.74E-05	1.26	(1.19, 1.34)	1.21E-04
		rs1047781	1.14	(1.11, 1.17)	8.10E-07	1.19	(1.14, 1.24)	2.41E-05	1.13	(1.10, 1.16)	3.81E-05	1.31	(1.21, 1.42)	8.79E-04
		rs1047781 x rs8176719 interaction	0.97	(0.95, 0.99)	9.91E-02	0.96	(0.93, 0.99)	1.64E-01	0.96	(0.94, 0.98)	8.46E-02	0.89	(0.83, 0.94)	4.28E-02
	Risk factors	Non Secretor status	1.16	(1.12, 1.19)	2.33E-07	1.30	(1.24, 1.35)	1.43E-09	1.12	(1.08, 1.15)	5.86E-04	1.40	(1.29, 1.52)	3.11E-05
		Blood group O	1.13	(1.11, 1.15)	1.32E-11	1.21	(1.18, 1.25)	1.39E-11	1.11	(1.09, 1.14)	2.41E-07	1.32	(1.25, 1.40)	2.80E-07
		O x Non-Secretor interaction	1.02	(0.97, 1.07)	6.55E-01	0.97	(0.90, 1.05)	7.02E-01	1.01	(0.96, 1.07)	7.99E-01	0.81	(0.70, 0.94)	1.44E-01

3. 6 Overlap of eQTL and pQTL

To detect the functionally relevant genes, the GTEx v8 datasets⁷⁵ were searched for overlap of lead cis expression quantitative trait locus (cis-eQTL) with PUD signals or their LD-proxies (LD $r^2 > 0.6$ in EAS or EUR)⁵⁴. The most significant eQTL hits for a gene within each tissue type were interpreted (Figure 14). Overlaps were detected for novel variants with eQTLs associated with *IHH*, *PLCL2*, *PTGER4*, *ZNF322*, *HIATL1*, *FAM211B*, and *GGT1* in the stomach.

Moreover, five recent large-scale pQTL datasets^{54–58} from serum or plasma were searched for overlap of cis- or trans-pQTL with PUD signals or their LD-proxies. In the analysis, overlaps with 88 unique significant pQTL associations were observed, most of which (93.1%) were trans-pQTL and involved the lead SNP at the ABO locus (Table 19). The cis-pQTL alleles in LD with PUD risk alleles were associated with increased levels of *EFNA1* and *OBP2B*, and decreased levels of *NHEJ1*, *ABO*, and *GGT1* (the cis-pQTLs overlapped with the cis-eQTLs mentioned above for *EFNA1*, *OBP2B*, *NHEJ1*, and *GGT1*). For trans-pQTLs in LD with the lead variants, links with multiple proteins were observed, including F8, F10, PROS1 (blood coagulation-related), and trefoil factor family peptides (which play important roles in response to gastrointestinal mucosal injury).

Additional analysis of the overlaps suggested plausible proteins and pathways (Supplementary Note; Table 19). By searching for pQTL associations, the PUD risk alleles of lead variants at *ABO* and *GGT1* were linked with a reduced level of coagulation factor VIII (F8) and increased levels of factor X (F10) and PROS1 (cofactor to activated protein C in the degradation of factor VIII). The risk allele of a lead SNP (rs1801020; *F12*) identified in the cross-ancestry meta-analysis is associated with decreased plasma levels of factor XII⁷⁷. Factor XII, VIII, X, and PROS1 are all involved in the intrinsic pathway of blood coagulation^{78,79}. These suggest that blood coagulation may be involved in peptic ulcer bleeding and healing. It is also likely that these signals were identified due to selection bias since PUD patients with severe symptoms are more likely to be detected than those without such severity. Additionally, in the pQTL study⁷⁶ that observed the proteomic associations with ABO blood groups and FUT2 secretor status, 31 proteins were found to be significantly associated with secretor status and non-O blood groups in the same direction (Methods). Notably, five proteins (CBLIF, CRNN, DSG2, REG1A, and REG1B) showed directionally concordant associations with secretor status and all three non-O blood groups (A, B, and AB) (Table 20).

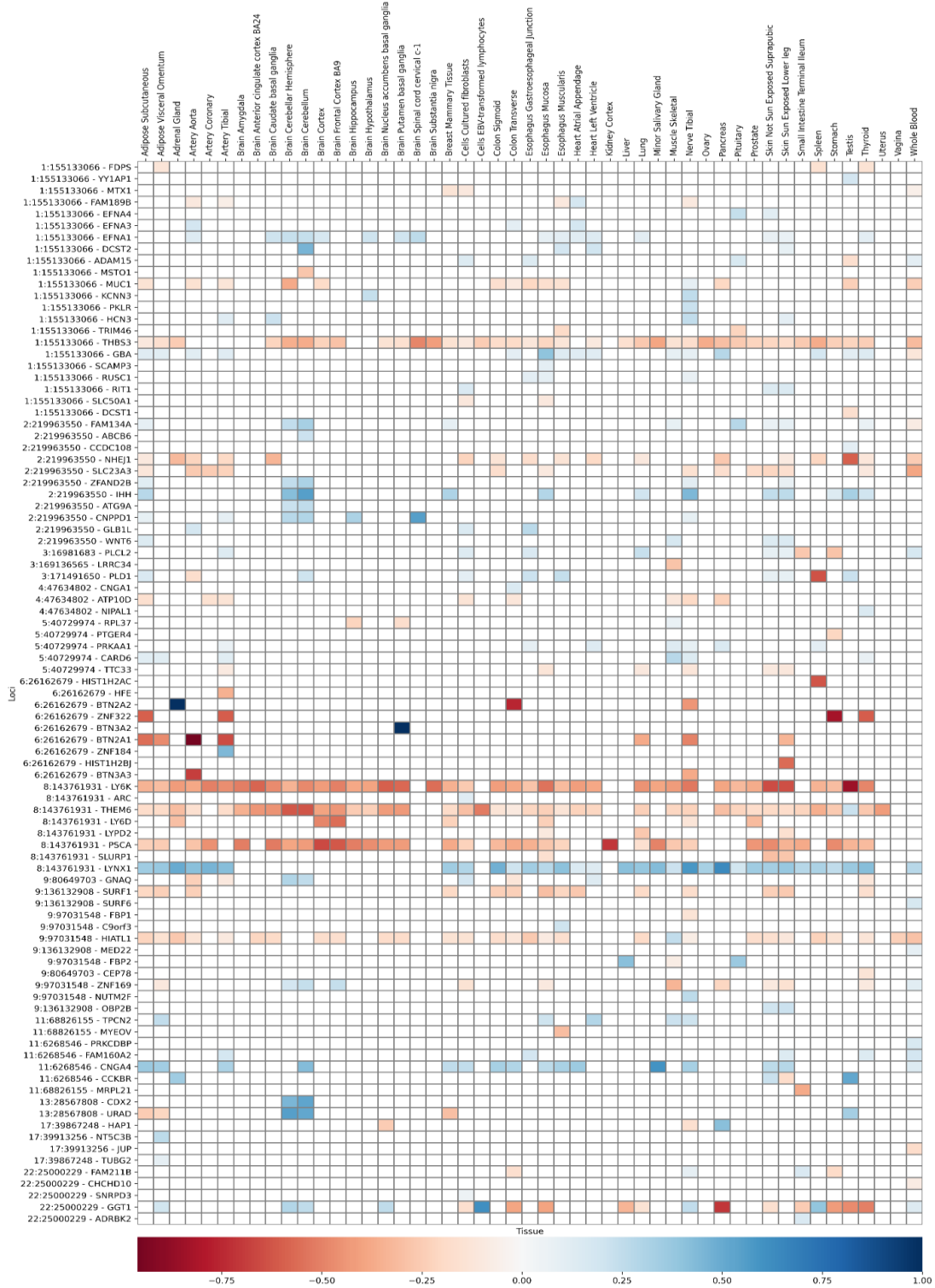


Figure 14. Overlap between PUD signals and significant cis-eQTL variants of the GTEx database. Overlap of significant cis-eQTL variants (FDR < 5%) of the GTEx version 8 datasets⁷⁵ with lead variants in novel loci or its LD proxy ($r^2 > 0.6$ in 1KG EAS or EUR populations¹⁶) in 41 tissue types. Square colors represent the normalized beta values (slope of the linear regression in eQTL mapping) of the eQTL allele that is in LD with the PUD risk allele. For each transcript, only the most significant eQTL association is shown. Columns represent the tissue types, and rows show the genome coordinates of PUD lead variants (GRCh37) and target transcripts of the overlapping eQTL association.

Table 19. Overlap of lead variants associated with PUD and PUD subtypes in EAS with pQTL signals.

Study	Sample	Target(Gene Name)	UniProt ID	pQTL (rsID)	CHR:POS (pQTL)	Effect Allele (pQTL)	BETA (pQTL)	P (pQTL)	Trans/Cis (pQTL)	Lead variant (PUD)	PUD risk allele of the lead variant	Alleles in LD between pQTL and lead variant	pQTL Allele in LD with PUD risk allele	Aligned_beta (pQTL Allele in LD with PUD risk allele)	LD R2 between the lead variant and reported pQTL	Population for LD R2 calculation
Ferkingstad, E.(2021)	plasma	KL	Q9UEF7	rs2974937	1:155168849	C	-0.05	5.50E-10	trans	1:155133066:GGGT	GT	→T,GT=C	C	-0.05	0.62	EAS
Ferkingstad, E.(2021)	plasma	CHGA	P10645	rs12801747	11:6286326	C	0.08	2.75E-15	trans	11:6268546:GA	A	G→T,A=C	C	0.08	0.81	EAS
Ferkingstad, E.(2021)	plasma	NTS	P30990	rs148675590,rs57333138,rs71328567808	13:28567808	A	0.06	4.47E-11	trans	13:28567808:AAAC>A	A	AAC→AAC,→	A	0.06	1.00	EAS
Ferkingstad, E.(2021)	plasma	NHEJ1	Q9H9Q4	rs2272019	2:220002534	C	0.09	2.24E-25	cis	2:219963550:C>T	T	C=C,T=A	A	-0.09	0.89	EAS
Ferkingstad, E.(2021)	plasma	F10	P00742	rs2017869	22:24997309	C	-0.06	6.46E-14	trans	22:25000229:A>T	A	A=G,T=C	G	0.06	0.98	EAS
Ferkingstad, E.(2021)	plasma	PROS1	P07225	rs1807629	22:25002323	C	-0.05	1.55E-09	trans	22:25000229:A>T	A	A=T,T=C	T	0.05	0.96	EUR
Ferkingstad, E.(2021)	plasma	CBLF	P27352	rs2920281	8:143760444	T	-0.09	1.23E-30	trans	8:143761931:C>T	C	C=C,T=T	C	0.09	0.94	EUR
Ferkingstad, E.(2021)	plasma	TIE1	P35590	rs950529388	9:136137065	G	0.22	4.47E-111	trans	9:136132908:T:TC	T	→A,C=G	A	-0.22	0.95	EAS
Ferkingstad, E.(2021)	plasma	REG1A	P05451	rs950529388	9:136137065	G	0.14	9.55E-49	trans	9:136132908:T:TC	T	→A,C=G	A	-0.14	0.95	EAS
Ferkingstad, E.(2021)	plasma	F8	P00451	rs950529388	9:136137065	G	0.45	0.00E+00	trans	9:136132908:T:TC	T	→A,C=G	A	-0.45	0.95	EAS
Ferkingstad, E.(2021)	plasma	REG3G	Q6UW15	rs950529388	9:136137065	G	0.09	1.02E-21	trans	9:136132908:T:TC	T	→A,C=G	A	-0.09	0.95	EAS
Ferkingstad, E.(2021)	plasma	REG1B	P48304	rs950529388	9:136137065	G	0.14	1.29E-47	trans	9:136132908:T:TC	T	→A,C=G	A	-0.14	0.95	EAS
Ferkingstad, E.(2021)	plasma	GOLM1	Q8NBJ4	rs950529388	9:136137065	G	0.52	0.00E+00	trans	9:136132908:T:TC	T	→A,C=G	A	-0.52	0.95	EAS
Ferkingstad, E.(2021)	plasma	MBL2	P11226	rs950529388	9:136137065	G	0.23	2.51E-188	trans	9:136132908:T:TC	T	→A,C=G	A	-0.23	0.95	EAS
Ferkingstad, E.(2021)	plasma	CD209	Q9NNX6	rs950529388	9:136137065	G	0.88	0.00E+00	trans	9:136132908:T:TC	T	→A,C=G	A	-0.88	0.95	EAS
Ferkingstad, E.(2021)	plasma	PNLIPRP1	P54315	rs950529388	9:136137065	G	0.08	4.57E-18	trans	9:136132908:T:TC	T	→A,C=G	A	-0.08	0.95	EAS
Ferkingstad, E.(2021)	plasma	CASC4	Q6P4E1	rs950529388	9:136137065	G	0.22	2.00E-133	trans	9:136132908:T:TC	T	→A,C=G	A	-0.22	0.95	EAS
Ferkingstad, E.(2021)	plasma	IMPAD1	Q9NXG2	rs950529388	9:136137065	G	0.10	1.32E-24	trans	9:136132908:T:TC	T	→A,C=G	A	-0.10	0.95	EAS
Ferkingstad, E.(2021)	plasma	ABO	P16442	rs950529388	9:136137065	G	1.34	0.00E+00	cis	9:136132908:T:TC	T	→A,C=G	A	-1.34	0.95	EAS
Ferkingstad, E.(2021)	plasma	FAM3B	P58499	rs950529388	9:136137065	G	0.09	1.41E-12	trans	9:136132908:T:TC	T	→A,C=G	A	-0.09	0.95	EAS
Ferkingstad, E.(2021)	plasma	FAM3D	Q9QBQ1	rs950529388	9:136137065	G	0.14	7.24E-21	trans	9:136132908:T:TC	T	→A,C=G	A	-0.14	0.95	EAS
Ferkingstad, E.(2021)	plasma	QSOX2	Q6ZRP7	rs950529388	9:136137065	G	0.23	8.32E-53	trans	9:136132908:T:TC	T	→A,C=G	A	-0.23	0.95	EAS
Ferkingstad, E.(2021)	plasma	GOLM1	Q8NBJ4	rs950529388	9:136137065	G	0.25	4.17E-55	trans	9:136132908:T:TC	T	→A,C=G	A	-0.25	0.95	EAS
Ferkingstad, E.(2021)	plasma	CASC4	Q6P4E1	rs975381715	9:136137106	A	0.42	0.00E+00	trans	9:136132908:T:TC	T	→G,C=A	G	-0.42	0.98	EAS
Ferkingstad, E.(2021)	plasma	PYY	P10082	rs975381715	9:136137106	A	-0.08	8.71E-18	trans	9:136132908:T:TC	T	→G,C=A	G	0.08	0.98	EAS
Ferkingstad, E.(2021)	plasma	SELE	P16581	rs975381715	9:136137106	A	-0.14	1.26E-32	trans	9:136132908:T:TC	T	→G,C=A	G	0.14	0.98	EAS
Ferkingstad, E.(2021)	plasma	HMOX1	P09601	rs950088295	9:136145471	G	0.06	1.20E-11	trans	9:136132908:T:TC	T	→A,C=G	A	-0.06	0.93	EAS
Ferkingstad, E.(2021)	plasma	OBP2B	Q9NPH6	rs920065566	9:136149229	C	-0.07	6.31E-16	cis	9:136132908:T:TC	T	→T,C=C	T	0.07	0.95	EAS
Ferkingstad, E.(2021)	plasma	CCL28	Q9NRJ3	rs920065566	9:136149229	C	-0.14	1.70E-51	trans	9:136132908:T:TC	T	→T,C=C	T	0.14	0.95	EAS
Ferkingstad, E.(2021)	plasma	PNPLA2	Q96AD5	rs920065566	9:136149229	C	-0.06	6.46E-12	trans	9:136132908:T:TC	T	→T,C=C	T	0.06	0.95	EAS
Ferkingstad, E.(2021)	plasma	LGALS8	O00214	rs920065566	9:136149229	C	-0.08	1.95E-10	trans	9:136132908:T:TC	T	→T,C=C	T	0.08	0.95	EAS
Gudjonsson, A.(2022)	serum	F8	P00451	rs136137065	9:136137065	G	0.38	9.99E-79	trans	9:136132908:T:TC	T	→A,C=G	A	-0.38	0.95	EAS
Gudjonsson, A.(2022)	serum	CASC4	Q6P4E1	rs687621	9:136137065	G	0.27	1.15E-36	trans	9:136132908:T:TC	T	→A,C=G	A	-0.27	0.95	EAS
Gudjonsson, A.(2022)	serum	LRRC32	Q14392	rs687621	9:136137065	G	0.25	1.09E-34	trans	9:136132908:T:TC	T	→A,C=G	A	-0.25	0.95	EAS
Gudjonsson, A.(2022)	serum	B3GNT2	Q9NY97	rs514659	9:136142203	C	0.19	7.27E-20	trans	9:136132908:T:TC	T	→A,C=C	A	-0.19	0.95	EAS
Gudjonsson, A.(2022)	serum	CD209	Q9NNX6	rs514659	9:136142203	C	0.93	0.00E+00	trans	9:136132908:T:TC	T	→A,C=C	A	-0.93	0.95	EAS
Gudjonsson, A.(2022)	serum	VWF	P04275	rs514659	9:136142203	C	0.31	6.69E-51	trans	9:136132908:T:TC	T	→A,C=C	A	-0.31	0.95	EAS
Gudjonsson, A.(2022)	serum	CHST15	Q7LFX5	rs9411378	9:136145425	A	-0.31	1.07E-38	trans	9:136132908:T:TC	T	→C,C=A	C	0.31	0.64	EUR
Gudjonsson, A.(2022)	serum	GNS	P15586	rs9411378	9:136145425	A	-0.21	3.91E-19	trans	9:136132908:T:TC	T	→C,C=A	C	0.21	0.64	EUR
Gudjonsson, A.(2022)	serum	SHANK3	Q9BYB0	rs2769071	9:136145974	G	0.34	9.18E-61	trans	9:136132908:T:TC	T	→A,C=G	A	-0.34	0.95	EAS
Gudjonsson, A.(2022)	serum	ABO	P16442	rs505922	9:136149229	C	1.47	0.00E+00	cis	9:136132908:T:TC	T	→T,C=C	T	-1.47	0.95	EAS
Gudjonsson, A.(2022)	serum	MBL2	P11226	rs505922	9:136149229	C	0.23	6.80E-28	trans	9:136132908:T:TC	T	→T,C=C	T	-0.23	0.95	EAS
Gudjonsson, A.(2022)	serum	PNPLA2	Q96AD5	rs505922	9:136149229	C	-0.19	3.35E-18	trans	9:136132908:T:TC	T	→T,C=C	T	0.19	0.95	EAS
Maik, P.(2021)	plasma	TFPI	P04155	rs4971093	1:155144300	A	-0.11	3.99E-16	trans	1:155133066:GGGT	GT	→A,GT=G	G	0.11	0.62	EAS
Maik, P.(2021)	plasma	CD209	Q9NNX6	rs597974	9:136144297	A	-0.83	0.00E+00	trans	9:136132908:T:TC	T	→A,C=G	A	-0.83	0.94	EAS
Maik, P.(2021)	plasma	CBLF	P27352	rs9411377	9:136145404	A	-0.18	3.83E-35	trans	9:136132908:T:TC	T	→C,C=A	C	0.18	0.87	EAS
Maik, P.(2021)	plasma	F8	P00451	rs9411377	9:136145404	A	0.63	0.00E+00	trans	9:136132908:T:TC	T	→C,C=A	C	-0.63	0.87	EAS
Maik, P.(2021)	plasma	VWF	P04275	rs9411377	9:136145404	A	0.24	5.03E-58	trans	9:136132908:T:TC	T	→C,C=A	C	-0.24	0.87	EAS
Maik, P.(2021)	plasma	DHFR	P00374	rs9411377	9:136145404	A	0.11	4.14E-15	trans	9:136132908:T:TC	T	→C,C=A	C	-0.11	0.87	EAS
Maik, P.(2021)	plasma	MBL2	P11226	rs529565	9:136149500	T	-0.22	1.01E-52	trans	9:136132908:T:TC	T	→T,C=C	T	-0.22	0.94	EAS
Maik, P.(2021)	plasma	S100A16	Q96FQ6	rs529565	9:136149500	T	-0.20	4.81E-51	trans	9:136132908:T:TC	T	→T,C=C	T	-0.20	0.94	EAS
Maik, P.(2021)	plasma	REG1A	P05451	rs950529388	9:136137065	A	-0.18	2.21E-35	trans	9:136132908:T:TC	T	→A,C=G	A	-0.18	0.95	EAS
Maik, P.(2021)	plasma	REG1B	P48304	rs950529388	9:136137065	A	-0.17	1.13E-33	trans	9:136132908:T:TC	T	→A,C=G	A	-0.17	0.95	EAS
Maik, P.(2021)	plasma	SVEP1	Q4LDE5	rs950529388	9:136137065	A	-0.11	6.96E-15	trans	9:136132908:T:TC	T	→A,C=G	A	-0.11	0.95	EAS
Maik, P.(2021)	plasma	CHST11	Q9NPF2	rs950529388	9:136137065	A	-0.24	7.12E-68	trans	9:136132908:T:TC	T	→A,C=G	A	-0.24	0.95	EAS
Maik, P.(2021)	plasma	TFRC	P02786	rs1327163796	9:136143136	D	-0.15	3.72E-22	trans	9:136132908:T:TC	T	→,C=T	D	-0.15	0.82	EAS
Maik, P.(2021)	plasma	CLIC5	Q9NZA1	rs1588647366	9:136145907	D	-0.10	5.56E-15	trans	9:136132908:T:TC	T	→,C=A	D	-0.10	0.95	EAS
Maik, P.(2021)	plasma	REG3G	Q6UW15	rs514659	9:136142203	A	-0.12	5.02E-18	trans	9:136132908:T:TC	T	→A,C=C	A	-0.12	0.95	EAS
Maik, P.(2021)	plasma	ABO	P16442	rs576125	9:136144309	A	1.23	0.00E+00	cis	9:136132908:T:TC	T	→G,C=A	G	-1.23	0.94	EAS
Maik, P.(2021)	plasma	LRRC32	Q14392	rs576125	9:136144309	A	0.25	2.18E-83	trans	9:136132908:T:TC	T	→G,C=A	G	-0.25	0.94	EAS
Maik, P.(2021)	plasma	NFASC	O94856	rs1834886880	9:136145993	D	-0.11	3.03E-14	trans	9:136132908:T:TC	T	→AGAAGGGAAAT	AGAAGGGAAATTAA	0.11	0.95	EAS
Maik, P.(2021)	plasma	A4GALT	Q9NPC4	rs13292932	9:136145419	A	0.11	4.46E-14	trans	9:136132908:T:TC	T	→C,C=A	C	-0.11	0.87	EAS
Maik, P.(2021)	plasma	CASC4	Q6P4E1	rs189262702	9:136146068	D	-0.32	1.22E-126	trans	9:136132908:T:TC	T	→,C=T	T	-0.32	0.93	EAS
Maik, P.(2021)	plasma	NHEJ1	Q9H9Q4	rs6713887	2:220026369	T	-0.18	4.00E-41	cis	2:219963550:C>T	T	C=G,T=T	T	-0.18	0.89	EAS

Sun, B.B. (2018)	plasma	CHSTB	Q9NPF2	rs687621	9:136137065	G	0.18	1.30E-11	trans	9:136132908.T:TC	T	->A,C<-G	A		-0.18	0.95	EAS
Sun, B.B. (2018)	plasma	DSG2	Q14126	rs687621	9:136137065	G	0.22	3.60E-17	trans	9:136132908.T:TC	T	->A,C<-G	A		-0.22	0.95	EAS
Sun, B.B. (2018)	plasma	F8	P00451	rs9411377	9:136145404	A	0.51	2.80E-80	trans	9:136132908.T:TC	T	->C,C<=A	C		-0.51	0.87	EAS
Sun, B.B. (2018)	plasma	TECK	O15444	rs9411378	9:136145425	A	-0.21	4.40E-13	trans	9:136132908.T:TC	T	->C,C<=A	C		0.21	0.64	EUR
Sun, B.B. (2018)	plasma	DYR	P00374	rs676457	9:136146227	T	0.19	2.50E-12	trans	9:136132908.T:TC	T	->A,C<=T	A		-0.19	0.95	EAS
Sun, B.B. (2018)	plasma	MBL2	P11226	rs139840563	9:136146448	T	0.24	4.80E-19	trans	9:136132908.T:TC	T	->AAGAC,C<=	AAGAC		-0.24	0.95	EAS
Sun, B.B. (2018)	plasma	ABO	P16442	rs505922	9:136149229	C	1.30	0.00E+00	cis	9:136132908.T:TC	T	->T,C<=C	T		-1.30	0.95	EAS
Sun, B.B. (2018)	plasma	CD209	Q9NNX6	rs505922	9:136149229	C	0.83	1.30E-301	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.83	0.95	EAS
Sun, B.B. (2022)	plasma	EFNA1	P20827	rs4390169	1:155106054	G	0.31	0.00E+00	cis	1:155133066.G:GGT	GT	->A,GT<=G	G		0.31	0.81	EAS
Sun, B.B. (2022)	plasma	HAO1	Q9UJM8	rs12904	1:155106697	A	-0.06	2.05E-15	trans	1:155133066.G:GGT	GT	->A,GT<=G	G		0.06	0.81	EAS
Sun, B.B. (2022)	plasma	ALPP	P05187	rs12743084	1:155161168	G	0.06	7.10E-21	trans	1:155133066.G:GGT	GT	->G,GT<=C	C		-0.06	0.62	EAS
Sun, B.B. (2022)	plasma	GGT1	P19440	rs2006227	22:24995756	A	0.25	2.88E-272	cis	22:25000229.A:T	A	A<=C,T=A	C		-0.25	0.99	EAS
Sun, B.B. (2022)	plasma	CBLIF	P27352	rs34635647	8:143756890	G	-0.05	1.51E-13	trans	8:143761931.C:T	C	C<=A,T=G	A		0.05	0.97	EAS
Sun, B.B. (2022)	plasma	TFF2	Q03403	rs2978981	8:143759137	T	-0.05	1.33E-12	trans	8:143761931.C:T	C	C<=C,T=T	C		0.05	0.99	EUR
Sun, B.B. (2022)	plasma	GHRL	Q9UBU3	rs11786721	8:143760179	C	0.07	1.30E-21	trans	8:143761931.C:T	C	C<=T,T=C	T		-0.07	0.91	EUR
Sun, B.B. (2022)	plasma	PLA2G10	O15496	rs2976388	8:143760256	A	-0.26	1.73E-280	trans	8:143761931.C:T	C	C<=G,T=A	G		0.26	0.94	EUR
Sun, B.B. (2022)	plasma	ADGRG2	Q8IZP9	rs8176719	9:136132908	TC	0.05	3.65E-12	trans	9:136132908.T:TC	T	->C,C<=C	T		-0.05	1.00	EAS
Sun, B.B. (2022)	plasma	LAYN	Q6UX15	rs8176719	9:136132908	TC	0.07	2.03E-22	trans	9:136132908.T:TC	T	->C,C<=C	T		-0.07	1.00	EAS
Sun, B.B. (2022)	plasma	SCARA5	Q6ZMJ2	rs8176719	9:136132908	TC	0.10	4.70E-48	trans	9:136132908.T:TC	T	->C,C<=C	T		-0.10	1.00	EAS
Sun, B.B. (2022)	plasma	THBS4	P35443	rs8176719	9:136132908	TC	0.05	7.73E-12	trans	9:136132908.T:TC	T	->C,C<=C	T		-0.05	1.00	EAS
Sun, B.B. (2022)	plasma	RGMA	Q96B86	rs657152	9:136139265	A	0.05	2.29E-11	trans	9:136132908.T:TC	T	->C,C<=A	C		-0.05	0.97	EUR
Sun, B.B. (2022)	plasma	NOS3	P29474	-	9:136141646	C	0.07	4.66E-17	trans	9:136132908.T:TC	T	->AA,C<=	AA		-0.07	0.95	EAS
Sun, B.B. (2022)	plasma	CTSS	P25774	rs545971	9:136143372	T	-0.07	6.42E-16	trans	9:136132908.T:TC	T	->C,C<=T	C		0.07	0.95	EAS
Sun, B.B. (2022)	plasma	PTPRS	Q13332	rs545971	9:136143372	T	0.06	1.31E-13	trans	9:136132908.T:TC	T	->C,C<=T	C		-0.06	0.95	EAS
Sun, B.B. (2022)	plasma	CRIP2	P52943	rs495203	9:136145240	T	0.07	1.44E-22	trans	9:136132908.T:TC	T	->C,C<=T	C		-0.07	0.95	EAS
Sun, B.B. (2022)	plasma	CKMT1A, CKMT1B	P12532	rs13292932	9:136145419	A	0.07	3.17E-16	trans	9:136132908.T:TC	T	->C,C<=A	C		-0.07	0.87	EAS
Sun, B.B. (2022)	plasma	SOD2	P04179	rs9411378	9:136145425	A	0.06	1.51E-11	trans	9:136132908.T:TC	T	->C,C<=A	C		-0.06	0.64	EUR
Sun, B.B. (2022)	plasma	TIMD4	Q96H15	rs9411378	9:136145425	A	0.10	4.71E-32	trans	9:136132908.T:TC	T	->C,C<=A	C		-0.10	0.64	EUR
Sun, B.B. (2022)	plasma	CA4	P22748	rs582118	9:136145471	G	0.16	9.47E-98	trans	9:136132908.T:TC	T	->A,C<=G	A		-0.16	0.93	EAS
Sun, B.B. (2022)	plasma	CRNN	Q9UBG3	rs582118	9:136145471	G	0.06	4.31E-16	trans	9:136132908.T:TC	T	->A,C<=G	A		-0.06	0.93	EAS
Sun, B.B. (2022)	plasma	CD28	P10747	rs34357864	9:136145907	GA	0.08	1.13E-22	trans	9:136132908.T:TC	T	->C,C<=A	T		-0.08	0.95	EAS
Sun, B.B. (2022)	plasma	FCG23	Q9GZV9	rs34357864	9:136145907	GA	0.09	3.73E-29	trans	9:136132908.T:TC	T	->C,C<=A	T		-0.09	0.95	EAS
Sun, B.B. (2022)	plasma	FCGR2A	P12318	rs782819119	9:136145993	A	0.05	2.58E-14	trans	9:136132908.T:TC	T	->AGAAGGGGAAATAGAAGGGGAAATTAA			-0.05	0.95	EAS
Sun, B.B. (2022)	plasma	GUCA2A	Q02747	rs782819119	9:136145993	A	0.07	6.00E-17	trans	9:136132908.T:TC	T	->AGAAGGGGAAATAGAAGGGGAAATTAA			-0.07	0.95	EAS
Sun, B.B. (2022)	plasma	ULBP2	Q9RZM5	rs677355	9:136146046	A	0.14	9.00E-83	trans	9:136132908.T:TC	T	->G,C<=A	G		-0.14	0.94	EAS
Sun, B.B. (2022)	plasma	BST2	Q10589	rs505922	9:136149229	C	0.10	1.67E-41	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.10	0.95	EAS
Sun, B.B. (2022)	plasma	CD209	Q9NNX6	rs505922	9:136149229	C	0.82	0.00E+00	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.82	0.95	EAS
Sun, B.B. (2022)	plasma	CLEC4G	Q6UXB4	rs505922	9:136149229	C	0.29	6.98E-305	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.29	0.95	EAS
Sun, B.B. (2022)	plasma	ENTPD6	O75354	rs505922	9:136149229	C	0.14	6.83E-88	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.14	0.95	EAS
Sun, B.B. (2022)	plasma	FCGR2B	P31994	rs505922	9:136149229	C	-0.18	9.27E-117	trans	9:136132908.T:TC	T	->T,C<=C	T		0.18	0.95	EAS
Sun, B.B. (2022)	plasma	FGFR2	P21802	rs505922	9:136149229	C	0.05	2.23E-11	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.05	0.95	EAS
Sun, B.B. (2022)	plasma	GCNT1	Q02742	rs505922	9:136149229	C	0.07	1.42E-21	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.07	0.95	EAS
Sun, B.B. (2022)	plasma	GP2	P55259	rs505922	9:136149229	C	0.55	0.00E+00	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.55	0.95	EAS
Sun, B.B. (2022)	plasma	LGALS4	P56470	rs505922	9:136149229	C	-0.52	0.00E+00	trans	9:136132908.T:TC	T	->T,C<=C	T		0.52	0.95	EAS
Sun, B.B. (2022)	plasma	LGALS7, LGALS7B	P47929	rs505922	9:136149229	C	0.31	0.00E+00	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.31	0.95	EAS
Sun, B.B. (2022)	plasma	LY9	Q9HBG7	rs505922	9:136149229	C	0.06	3.72E-14	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.06	0.95	EAS
Sun, B.B. (2022)	plasma	VAMP5	O95183	rs505922	9:136149229	C	0.06	4.82E-15	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.06	0.95	EAS
Sun, B.B. (2022)	plasma	VWF	P04275	rs505922	9:136149229	C	0.31	0.00E+00	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.31	0.95	EAS
Sun, B.B. (2022)	plasma	IDS	P22304	rs529565	9:136149500	C	-0.06	7.57E-13	trans	9:136132908.T:TC	T	->T,C<=C	T		0.06	0.94	EAS
Sun, B.B. (2022)	plasma	REG1A	P05451	rs529565	9:136149500	C	0.14	4.27E-76	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.14	0.94	EAS
Sun, B.B. (2022)	plasma	REG1B	P48304	rs529565	9:136149500	C	0.16	9.19E-98	trans	9:136132908.T:TC	T	->T,C<=C	T		-0.16	0.94	EAS
Sun, B.B. (2022)	plasma	CNTN5	O94779	rs10993170	9:97108540	C	0.23	2.38E-51	trans	9:97031548.GC	G	G<=G,C=C	G		-0.23	0.66	EAS

Table 20. Significant proteomic associations with ABO blood and FUT2 secretor status in concordant directions.

UKBPPP ProteinID:Protein Panel	Coefficients for the tested variables							Directions
	A	AB	B	secretor	secretor*A	secretor*AB	secretor*B	
AMIGO2:Q86SJ2:OID21370:v1:Oncology	-0.148			-0.058				??-???
AMY2A:P04746:OID20333:v1:Cardiometabolic	-0.198	-0.155		-0.087				-?--???
AMY2B:P19961:OID20340:v1:Cardiometabolic	-0.199	-0.147		-0.091				-?--???
CA9:Q16790:OID21417:v1:Oncology	0.07		0.087	0.069	0.14			++?++??
CBLIF:P27352:OID20179:v1:Cardiometabolic	-0.211	-0.264	-0.223	-0.079				----???
CEACAM1:P13688:OID21528:v1:Oncology	-0.237	-0.129		-0.062				-?--???
CKMT1A CKMT1B:P12532:OID20721:v1:Inflammation	0.072	0.171		0.158				+?+?+??
CRNN:Q9UBG3:OID21427:v1:Oncology	0.074	0.123	0.101	0.075				++++???
CTRB1:P17538:OID21150:v1:Neurology	-0.095			-0.064				??-???
CX3CL1:P78423:OID20976:v1:Neurology	-0.107			-0.117				??-???
DSC2:Q02487:OID21079:v1:Neurology			0.135	0.184				?+?+?+?
DSG2:Q14126:OID21113:v1:Neurology	0.198	0.516	0.553	0.077				++++???
FABP2:P12104:OID20200:v1:Cardiometabolic	0.076			0.085				+??+???
FGF19:Q95750:OID20715:v1:Inflammation			0.224	0.418			0.304	?+?+?+?
GALNT3:Q14435:OID20471:v1:Inflammation	0.413	0.281		0.333	0.579	0.448		+?++++?
GCG:P01275:OID21263:v1:Oncology	-0.07			-0.097				??-???
GCNT1:Q02742:OID21379:v1:Oncology	0.116	0.126		0.146	0.202			+?+++??
IFNGR2:P38484:OID20941:v1:Neurology			0.132	0.119	-0.133			?+?+~??
KEL:P23276:OID20881:v1:Neurology	-0.36	-0.239		-0.09				-?--???
L1CAM:P32004:OID21484:v1:Oncology	-0.437	-0.262		-0.072				-?--???
LGALS3:P17931:OID20284:v1:Cardiometabolic	-0.061			-0.139				??-???
MANSC1:Q9H8J5:OID21492:v1:Oncology	-0.087			-0.079				??-???
NECTIN4:Q96NY8:OID21490:v1:Oncology			0.089	0.063				?+?+?+?
REG1A:P05451:OID20415:v1:Cardiometabolic	0.199	0.257	0.203	0.36	0.144			+++++??
REG1B:P48304:OID20266:v1:Cardiometabolic	0.225	0.28	0.206	0.338	0.145			+++++??
SCG2:P13521:OID21400:v1:Oncology	-0.065			-0.147				??-???
SDC1:P18827:OID20169:v1:Cardiometabolic	-0.103			-0.132	-0.122			??-???
SIGLEC9:Q9Y336:OID21390:v1:Oncology	-0.174	-0.148		-0.056				-?--???
SPINK5:Q9NQ38:OID21148:v1:Neurology	-0.048			-0.341				??-???
SPON1:Q9HCB6:OID20759:v1:Inflammation		0.127	0.185	0.137			0.21	?+++?+?
TFF2:Q03403:OID20787:v1:Inflammation	0.051			0.104				+??+???

3. 7 Genetic correlation and pleiotropic effects

Cross-trait LD score regression²⁸ was conducted to evaluate the genetic correlation across PUD-related traits (Figure 15a). DU and GU showed significantly high genetic correlations ($r_g = 0.79$, FDR < 5%) with each other, as expected. Although not statistically significant (FDR < 5%), GU showed a positive genetic correlation with GC ($r_g = 0.17$), whereas DU was negatively correlated ($r_g = -0.14$; Figure 15; Table 21). The genetic correlation of PUD with dietary habits⁸⁰ and 215 complex traits in BBJ¹³ was also investigated (Methods); no significant genetic correlation was observed between PUD and other complex traits in EAS (FDR < 5%; Figure 16; Figure 17; Figure 18).

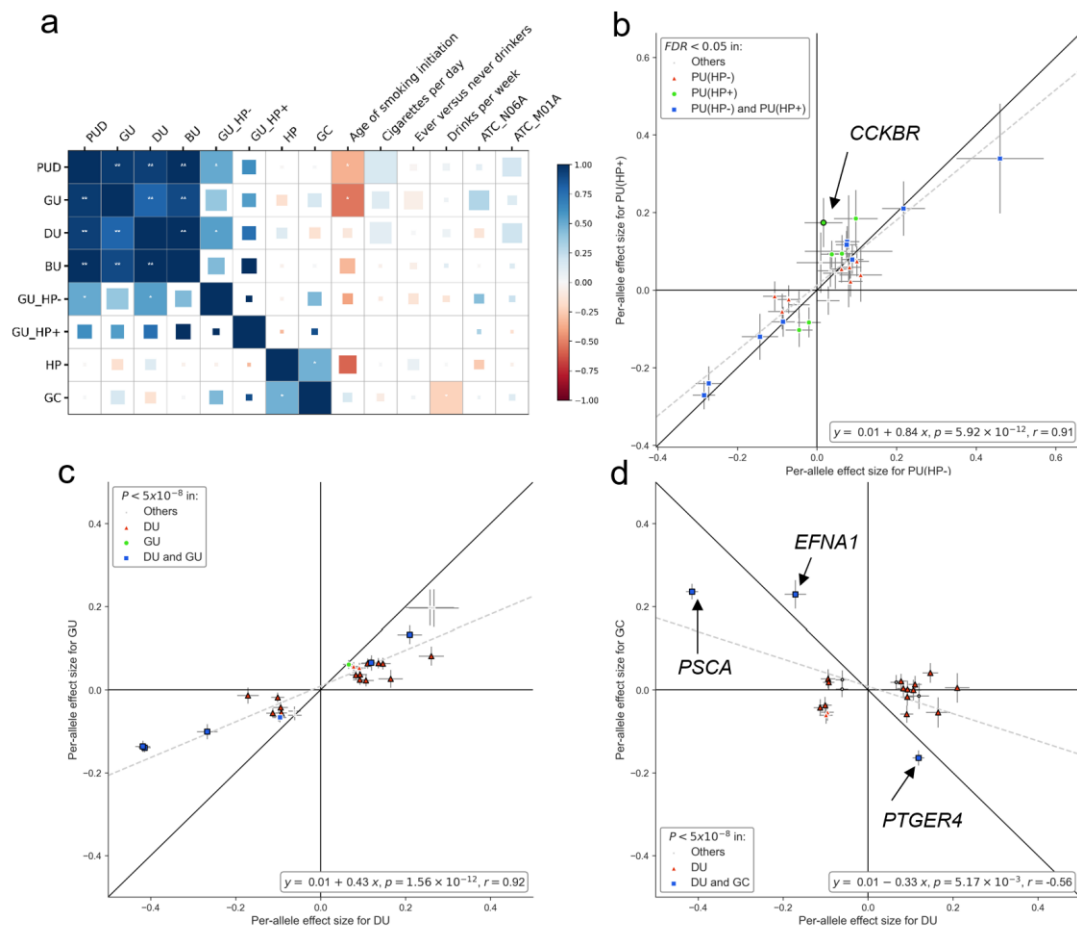


Figure 15. Effect size comparison of distinct variants and genetic correlations across PUD-related traits in the East Asian population.

a. Genetic correlation among PUD, PUD-related phenotypes, and risk factors. *, $P < 0.05$; **, FDR < 5%. The square size in each cell is proportional to $-\log_{10}(P)$. HP, *H. pylori* infection status determined by anti-*H. pylori* IgG level; GU_HP+, *H. pylori*-positive GU; GU_HP-, *H. pylori*-negative GU. b. Effect size comparison for PUD using summary statistics from *H. pylori* stratified analysis. PUD(HP+), *H. pylori*-positive PUD; PUD(HP-), *H. pylori*-negative PUD. c. Per-allele effect size (logarithm of odds ratios) comparison using EAS-specific summary statistics for DU and GU. Lead variants and secondary signals associated with PUD or any subtype in the EAS population were selected for comparison. The most significant associations were shown if overlapping variants existed (interval < 500 kb). Only variants with MAF > 0.01 are shown. Black marker edges denote variants with significant heterogeneity ($P_{het} < 0.05$). The grey dashed line represents the fitted linear regression line. Pearson's r is shown. d.

Effect size comparison between DU and gastric cancer (GC). Effect sizes for DU were obtained from the EAS-specific meta-analysis. GC summary statistics were obtained from previous GWAS conducted in BBJ1.

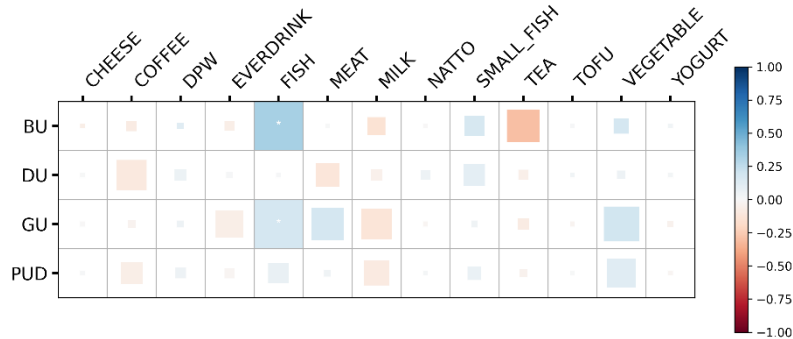


Figure 16. Genetic correlation heatmap of PUD and dietary habits in East Asian population.

Genetic correlation among PUD and dietary habits in BBJ1-180K⁸⁰(Methods). None of the genetic correlations reached the significance threshold (FDR < 5%) after corrections. The size of the colored square in each cell is proportional to the $-\log_{10}(P)$ value of its corresponding correlation. *, P < 0.05.

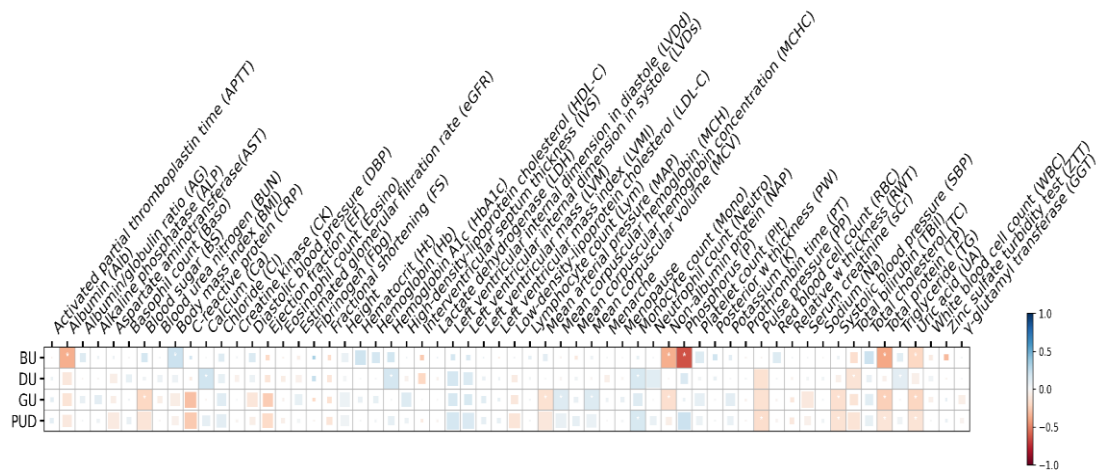


Figure 17. Genetic correlation heatmap of PUD and quantitative traits in East Asian population.

Genetic correlation among PUD and quantitative traits in BBJ1-180K⁸¹ (Methods). None of the genetic correlations reached the significance threshold (FDR < 5%) after corrections. The size of the colored square in each cell is proportional to the $-\log_{10}(P)$ value of its corresponding correlation. *, P < 0.05.

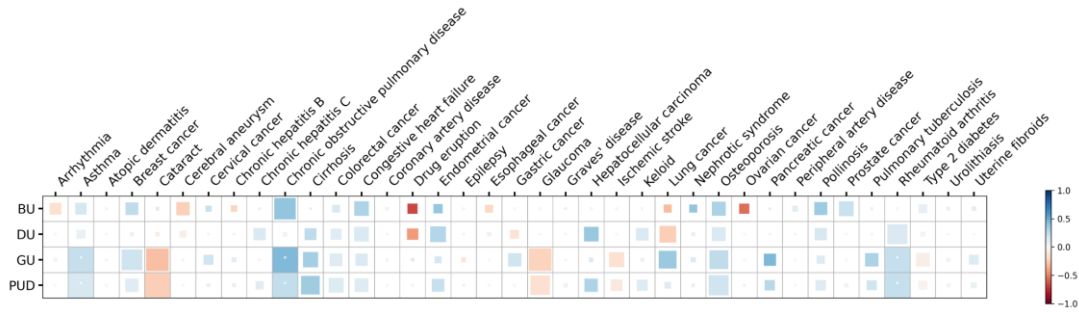


Figure 18. Genetic correlation heatmap of PUD and binary traits in East Asian population.

Genetic correlation among PUD and binary traits in BBJ1-180K⁴³ (Methods). None of the genetic correlations reached the significance threshold ($FDR < 5\%$) after corrections. The size of the colored square in each cell is proportional to the $-\log_{10}(P)$ value of its corresponding correlation. *, $P < 0.05$.

To investigate the pleiotropic effects of distinct variants, a PheWAS lookup was performed using previous large-scale GWAS in a Japanese population¹³. Among the 27 available lead variants associated with PUD and its subtypes, 16 reached the genome-wide significance threshold for at least one trait ($P < 5.0 \times 10^{-8}$). From them, 12 variants were associated with at least two traits after Bonferroni correction ($P < 8.6 \times 10^{-6}$; Figure 19; Figure 20; Figure 21). Both type 2 diabetes (two at *SND1-PAX4* locus and one at *GAST* locus) and GC (*EFNA1*, *PTGER4*, and *PSCA* loci) shared three significant variants after Bonferroni correction ($P < 8.6 \times 10^{-6}$) with PUD or its subtypes (Figure 22).

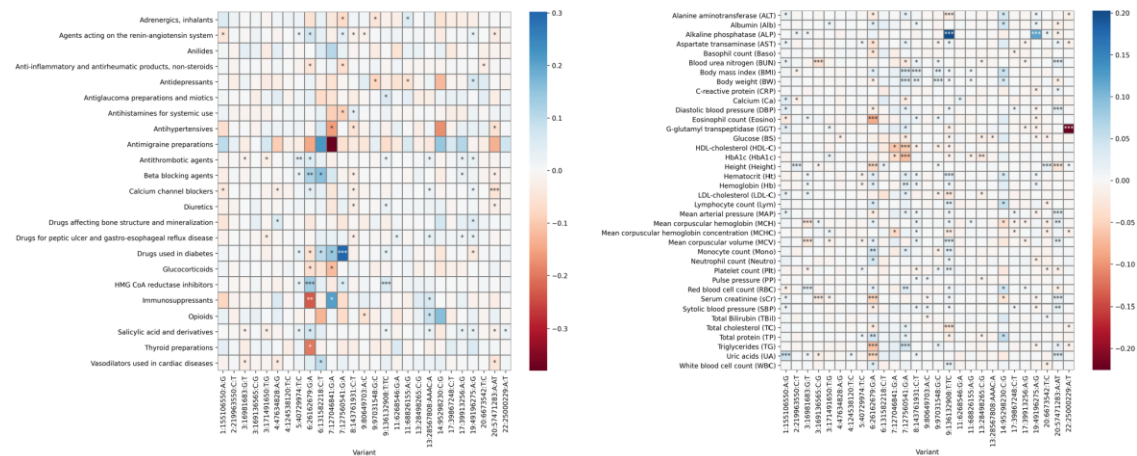


Figure 19. PheWAS heatmap of PUD risk variants with ATC codes and quantitative traits.

Lead variants and independent secondary signals associated with PUD or any subtypes in the EAS population were selected for PheWAS lookup. a, ATC codes. b, quantitative traits. The most significant associations were selected if overlapping variants exist (interval < 500 kb). Summary statistics for ATC codes and quantitative traits were obtained from previous GWASs in BBJ1¹³. Per-allele effect sizes ($\log(OR)$) of the PUD risk alleles are shown. *, $P < 0.05$ (Nominal significance). **, $P < 8.6 \times 10^{-6}$ (Bonferroni correction). ***, $P < 5.0 \times 10^{-8}$ (Genome-wide significance).

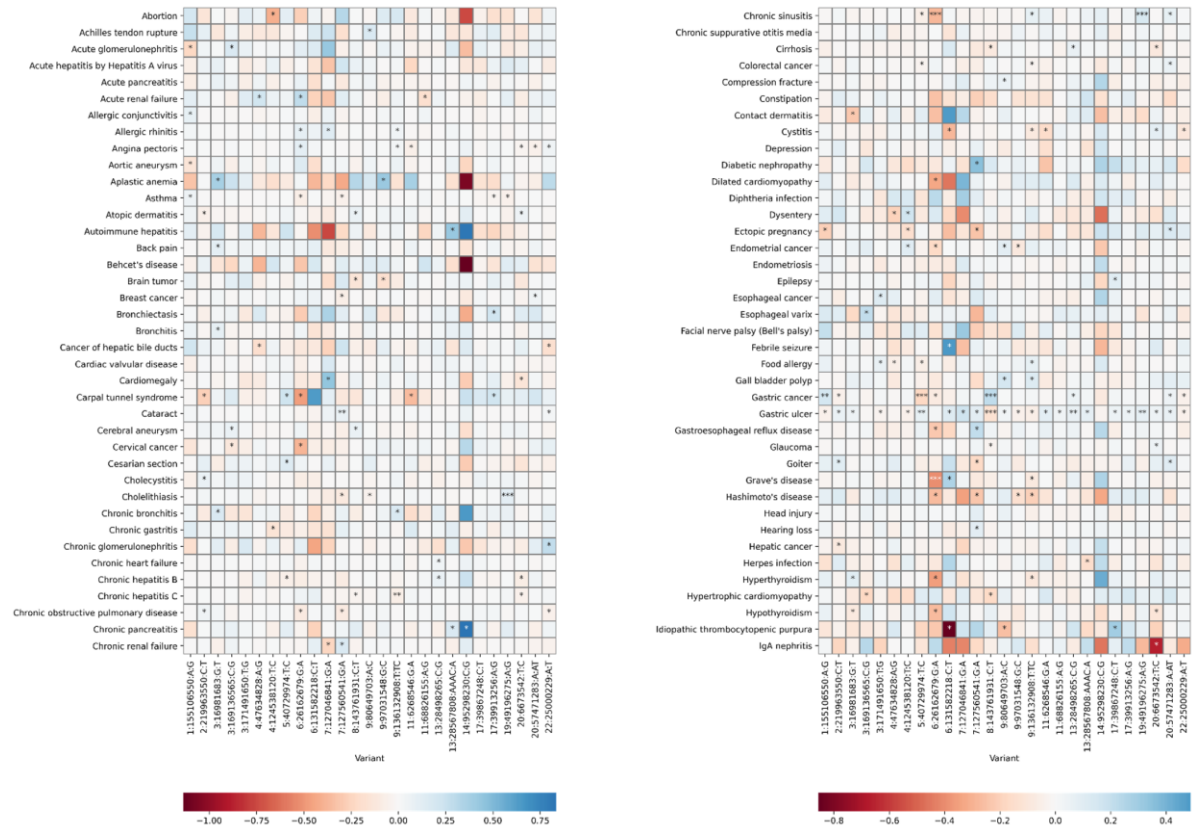


Figure 20. PheWAS heatmap of PUD risk variants with binary traits (Part 1/2).

Lead variants and independent secondary signals associated with PUD or any subtypes in the EAS population were selected for PheWAS lookup. The most significant associations were selected if overlapping variants exist (interval < 500 kb). Summary statistics for binary traits were obtained from previous GWASs in BBJ1-180K¹³. Per-allele effect sizes (log(OR)) of the PUD risk alleles are shown. *, $P < 0.05$ (Nominal significance). **, $P < 8.6 \times 10^{-6}$ (Bonferroni correction). ***, $P < 5.0 \times 10^{-8}$ (Genome-wide significance).



Figure 21. PheWAS heatmap of PUD risk variants with binary traits (Part 2/2).

Lead variants and independent secondary signals associated with PUD or any subtypes in the EAS population were selected for PheWAS lookup. The most significant associations were selected if overlapping variants exist (interval < 500 kb). Summary statistics for binary traits were obtained from previous GWASs in BBJ1-180K¹³. Per-allele effect sizes (log(OR)) of the PUD risk alleles are shown. *, $P < 0.05$ (Nominal significance). **, $P < 8.6 \times 10^{-6}$ (Bonferroni correction). ***, $P < 5.0 \times 10^{-8}$ (Genome-wide significance).

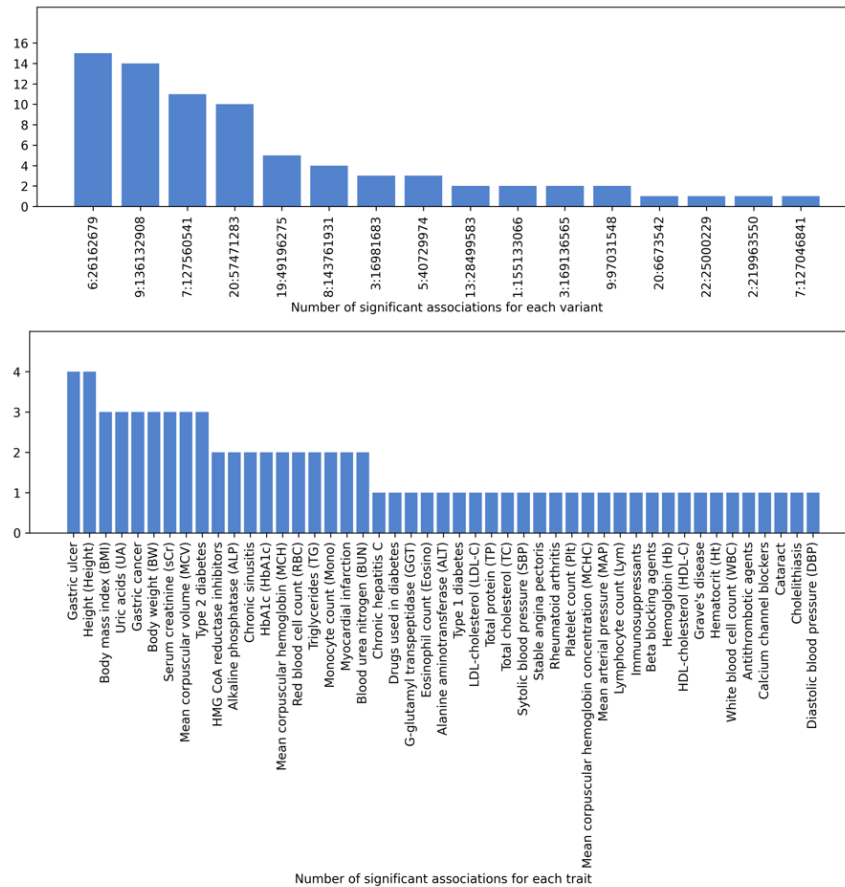


Figure 22. Summary of significant associations identified in PheWAS.

Genome-wide significant associations identified in PheWAS lookup are summarized for each variant (top) or each trait (bottom).

Table 21. Genetic correlation among PUD, PUD-related phenotypes, and its risk factors.

Trait 1	Trait 2	r_g	SE	Z	P
PUD	GU	0.959	0.021	44.850	0.00E+00
PUD	DU	0.942	0.033	28.530	0.00E+00
PUD	BU	1.026	0.169	6.065	1.32E-09
PUD	GU HP-	0.515	0.236	2.185	2.89E-02
PUD	GU HP+	0.618	0.560	1.103	2.70E-01
PUD	HP	-0.008	0.144	-0.058	9.54E-01
PUD	GC	0.011	0.155	0.071	9.43E-01
PUD	Age of smoking initiation	-0.354	0.151	-2.353	1.86E-02
PUD	Cigarettes per day	0.157	0.080	1.954	5.07E-02
PUD	Ever versus never drinkers	-0.022	0.036	-0.620	5.36E-01
PUD	Drinks per week	0.054	0.085	0.637	5.24E-01
PUD	ATC_N06A	0.096	0.164	0.584	5.59E-01
PUD	ATC_M01A	0.181	0.134	1.351	1.77E-01
GU	PUD	0.959	0.021	44.850	0.00E+00
GU	DU	0.798	0.087	9.147	5.84E-20
GU	BU	0.899	0.152	5.915	3.31E-09
GU	GU HP-	0.386	0.262	1.470	1.41E-01
GU	GU HP+	0.556	0.529	1.051	2.93E-01
GU	HP	-0.164	0.179	-0.917	3.59E-01
GU	GC	0.170	0.158	1.080	2.80E-01
GU	Age of smoking initiation	-0.537	0.204	-2.633	8.47E-03
GU	Cigarettes per day	0.131	0.104	1.261	2.07E-01
GU	Ever versus never drinkers	-0.066	0.050	-1.328	1.84E-01
GU	Drinks per week	0.052	0.118	0.438	6.62E-01
GU	ATC_N06A	0.305	0.217	1.401	1.61E-01
GU	ATC_M01A	0.138	0.168	0.822	4.11E-01
DU	PUD	0.942	0.033	28.530	0.00E+00
DU	GU	0.798	0.087	9.147	5.84E-20
DU	BU	1.036	0.134	7.749	9.25E-15
DU	GU HP-	0.553	0.240	2.310	2.09E-02
DU	GU HP+	0.751	0.696	1.079	2.81E-01
DU	HP	0.131	0.146	0.897	3.70E-01
DU	GC	-0.149	0.161	-0.926	3.54E-01
DU	Age of smoking initiation	-0.114	0.139	-0.821	4.12E-01
DU	Cigarettes per day	0.109	0.073	1.487	1.37E-01
DU	Ever versus never drinkers	0.016	0.037	0.420	6.74E-01
DU	Drinks per week	0.062	0.089	0.694	4.88E-01
DU	ATC_N06A	-0.131	0.159	-0.821	4.12E-01
DU	ATC_M01A	0.206	0.148	1.391	1.64E-01
BU	PUD	1.026	0.169	6.065	1.32E-09
BU	GU	0.899	0.152	5.915	3.31E-09
BU	DU	1.036	0.134	7.749	9.25E-15
BU	GU HP-	0.440	0.364	1.208	2.27E-01
BU	GU HP+	1.092	0.966	1.130	2.58E-01
BU	HP	-0.052	0.276	-0.190	8.49E-01
BU	GC	0.014	0.262	0.054	9.57E-01
BU	Age of smoking initiation	-0.331	0.285	-1.164	2.45E-01
BU	Cigarettes per day	-0.054	0.155	-0.347	7.29E-01
BU	Ever versus never drinkers	-0.063	0.103	-0.611	5.41E-01
BU	Drinks per week	0.120	0.279	0.429	6.68E-01
BU	ATC_N06A	-0.012	0.291	-0.040	9.68E-01
BU	ATC_M01A	0.139	0.284	0.490	6.24E-01
GU HP-	PUD	0.515	0.236	2.185	2.89E-02
GU HP-	GU	0.386	0.262	1.470	1.41E-01
GU HP-	DU	0.553	0.240	2.310	2.09E-02
GU HP-	BU	0.440	0.364	1.208	2.27E-01
GU HP-	GU HP+	1.649	2.584	0.638	5.24E-01
GU HP-	HP	-0.049	0.308	-0.160	8.73E-01
GU HP-	GC	0.445	0.414	1.075	2.82E-01
GU HP-	Age of smoking initiation	-0.405	0.462	-0.877	3.81E-01
GU HP-	Cigarettes per day	-0.101	0.230	-0.439	6.61E-01
GU HP-	Ever versus never drinkers	-0.136	0.158	-0.858	3.91E-01
GU HP-	Drinks per week	-0.139	0.223	-0.622	5.34E-01
GU HP-	ATC_N06A	0.426	0.507	0.841	4.01E-01
GU HP-	ATC_M01A	0.222	0.387	0.574	5.66E-01
GU HP+	PUD	0.618	0.560	1.103	2.70E-01
GU HP+	GU	0.556	0.529	1.051	2.93E-01
GU HP+	DU	0.751	0.696	1.079	2.81E-01
GU HP+	BU	1.092	0.966	1.130	2.58E-01
GU HP+	GU HP-	1.649	2.584	0.638	5.24E-01
GU HP+	HP	-0.322	0.805	-0.400	6.90E-01
GU HP+	GC	0.886	1.488	0.596	5.52E-01
GU HP+	ATC_M01A	-0.161	0.577	-0.280	7.80E-01
GU HP+	ATC_N06A	0.313	0.688	0.455	6.49E-01
HP	PUD	-0.008	0.144	-0.058	9.54E-01
HP	GU	-0.164	0.179	-0.917	3.59E-01
HP	DU	0.131	0.146	0.897	3.70E-01
HP	BU	-0.052	0.276	-0.190	8.49E-01
HP	GU HP-	-0.049	0.308	-0.160	8.73E-01
HP	GU HP+	-0.322	0.805	-0.400	6.90E-01
HP	GC	0.478	0.226	2.115	3.44E-02
HP	Age of smoking initiation	-0.582	0.450	-1.295	1.95E-01
HP	Cigarettes per day	0.014	0.154	0.093	9.26E-01
HP	Ever versus never drinkers	0.072	0.078	0.919	3.58E-01
HP	Drinks per week	0.007	0.133	0.055	9.56E-01
HP	ATC_N06A	-0.261	0.311	-0.839	4.02E-01
HP	ATC_M01A	-0.057	0.247	-0.229	8.19E-01
GC	PUD	0.011	0.155	0.071	9.43E-01
GC	GU	0.170	0.158	1.080	2.80E-01
GC	DU	-0.149	0.161	-0.926	3.54E-01
GC	BU	0.014	0.262	0.054	9.57E-01
GC	GU HP-	0.445	0.414	1.075	2.82E-01
GC	GU HP+	0.886	1.488	0.596	5.52E-01
GC	HP	0.478	0.226	2.115	3.44E-02
GC	Age of smoking initiation	0.163	0.197	0.825	4.09E-01
GC	Cigarettes per day	-0.079	0.114	-0.688	4.92E-01
GC	Ever versus never drinkers	0.044	0.117	0.375	7.08E-01
GC	Drinks per week	-0.230	0.108	-2.134	3.28E-02
GC	ATC_N06A	0.062	0.227	0.271	7.86E-01
GC	ATC_M01A	0.125	0.214	0.586	5.58E-01

3. 8 H.pylori stratified analysis

Given the importance of H.pylori infection in PUD etiology, the differences in genetic architectures between HP-induced and HP-unrelated peptic ulcers were examined by conducting HP-stratified association tests for PUD in HP-positive and HP-negative individuals from TMM-50K (Methods; Table 22). For the distinct PUD signals identified in the EAS population (29 variants with MAF > 0.01), per-allele effect sizes for PUD between HP-negative and HP-positive status were highly correlated (Figure 15b; slope = 0.84, SE = 0.07, $r = 0.91$; Figure 23). one lead SNP (rs12792379), specifically associated with HP-positive PUD, at *CCKBR* ($OR_{HP\text{-}positive} = 1.18$, CI = 1.05 – 1.34; $OR_{HP\text{-}negative} = 1.01$, CI= 0.92 - 1.11), and one HP-negative GU locus rs12347577 near *ZNF169* (Cochran's Q test, $P_{het} < 0.05$) were identified. On the other hand, the lead variants in the most significant locus at *PSCA* did not show significant heterogeneity in effect sizes ($P < 0.05$). Colocalization analysis suggested HP-positive PUD and HP-negative PUD shared the causal variant in *PSCA* ($PP4 > 0.8$ for PUD; **Table 23**).

Table 22. H.pylori infection status of participants in TMM-50K.

H.pylori infection status was determined by serum anti-H.pylori IgG level. HP-positive status, anti-H.pylori IgG level ≥ 10 U/ml; HP-negative status, anti-H.pylori IgG level < 10 U/ml.

HP status	Control		PUD		DU		GU		BU	
	Age (mean \pm SD)	N	Age (mean \pm SD)	N	Age (mean \pm SD)	N	Age (mean \pm SD)	N	Age (mean \pm SD)	N
HP-positive (≥ 10 U/ml)	63.25 \pm 8.56	16963	63.36 \pm 7.98	1885	63.14 \pm 8.08	1078	63.58 \pm 7.87	1084	63.35 \pm 7.99	277
HP-negative (< 10 U/ml)	58.27 \pm 12.45	26432	62.66 \pm 8.90	3372	62.34 \pm 9.04	1773	62.84 \pm 8.83	2140	62.28 \pm 9.10	541
Total	60.22 \pm 11.35	43395	62.91 \pm 8.59	5257	62.64 \pm 8.70	2851	63.08 \pm 8.53	3224	62.65 \pm 8.75	818

Table 23. Colocalization analysis for HP-stratified analysis.

Trait 1: GWAS of specified phenotype in HP-negative (HP-) individuals. Trait 2: GWAS of specified phenotype in HP-positive (HP+) individuals. H0: neither trait has a genetic association in the region. H1: only trait 1 has a genetic association in the region. H2: only trait 2 has a genetic association in the region. H3: both traits are associated but with different causal variants. H4: both traits are associated and share a single causal variant.

Analysis	Phenotype	Number of variants	Locus (SNPID for lead variant)	PP.H0.abf	PP.H1.abf	PP.H2.abf	PP.H3.abf	PP.H4.abf	Signal for trait 1	Signal for trait 2
coloc	PUD	2234	1:155133066:G:GGT	0.5924	0.2629	0.0915	0.0406	0.0127	/	/
coloc	GU	2231	1:155133066:G:GGT	0.5666	0.2248	0.1320	0.0523	0.0243	/	/
coloc	DU	2229	1:155133066:G:GGT	0.6212	0.1568	0.1582	0.0399	0.0240	/	/
coloc	BU	2223	1:155133066:G:GGT	0.6083	0.1784	0.1464	0.0429	0.0239	/	/
coloc	PUD	3077	2:219949913:A:T	0.5780	0.1634	0.1816	0.0513	0.0256	/	/
coloc	GU	3067	2:219949913:A:T	0.4927	0.3206	0.1035	0.0673	0.0160	/	/
coloc	DU	3073	2:219949913:A:T	0.3152	0.0895	0.4286	0.1217	0.0450	/	/
coloc	BU	3065	2:219949913:A:T	0.5145	0.1878	0.1856	0.0677	0.0445	/	/
coloc	PUD	4432	3:169091253:A:G	0.2667	0.2872	0.1866	0.2009	0.0587	/	/
coloc	GU	4425	3:169091253:A:G	0.3511	0.2444	0.2284	0.1589	0.0173	/	/
coloc	DU	4426	3:169091253:A:G	0.4246	0.1775	0.2644	0.1105	0.0230	/	/
coloc	BU	4419	3:169091253:A:G	0.3662	0.3351	0.1436	0.1313	0.0239	/	/
coloc	PUD	4723	3:16974123:T:G	0.5097	0.2035	0.1942	0.0775	0.0151	/	/
coloc	GU	4714	3:16974123:T:G	0.4780	0.2185	0.1963	0.0897	0.0176	/	/
coloc	DU	4714	3:16974123:T:G	0.5240	0.1803	0.2086	0.0718	0.0153	/	/
coloc	BU	4703	3:16974123:T:G	0.4559	0.1770	0.2493	0.0968	0.0209	/	/
coloc	PUD	4302	4:124538120:T:C	0.6159	0.1458	0.1834	0.0434	0.0114	/	/
coloc	GU	4299	4:124538120:T:C	0.6137	0.1455	0.1850	0.0438	0.0120	/	/
coloc	DU	4297	4:124538120:T:C	0.5714	0.1670	0.1914	0.0559	0.0142	/	/
coloc	BU	4292	4:124538120:T:C	0.5537	0.1749	0.1945	0.0614	0.0155	/	/
coloc	PUD	4645	5:40790627:C:CTCT	0.3431	0.1788	0.2801	0.1458	0.0522	/	/
coloc	GU	4637	5:40790627:C:CTCT	0.5232	0.1946	0.1936	0.0720	0.0166	/	/
coloc	DU	4635	5:40790627:C:CTCT	0.4135	0.1509	0.3064	0.1118	0.0174	/	/
coloc	BU	4629	5:40790627:C:CTCT	0.4919	0.1938	0.2127	0.0838	0.0178	/	/
coloc	PUD	3410	7:127046841:G:A	0.4416	0.3419	0.1015	0.0786	0.0364	/	/
coloc	GU	3406	7:127046841:G:A	0.5561	0.2346	0.1369	0.0578	0.0146	/	/
coloc	DU	3401	7:127046841:G:A	0.2624	0.4738	0.0692	0.1250	0.0696	/	/
coloc	BU	3397	7:127046841:G:A	0.5147	0.2306	0.1604	0.0719	0.0224	/	/
coloc	PUD	3218	7:127560541:G:A	0.4143	0.3657	0.1006	0.0887	0.0307	/	/
coloc	GU	3214	7:127560541:G:A	0.5338	0.2592	0.1285	0.0624	0.0160	/	/
coloc	DU	3211	7:127560541:G:A	0.2331	0.5014	0.0629	0.1351	0.0675	/	/
coloc	BU	3207	7:127560541:G:A	0.3502	0.1679	0.3112	0.1492	0.0215	/	/
coloc	PUD	5695	8:143761931:C:T	0.0000	0.0000	0.0000	0.0448	0.9552	/	/
coloc	GU	5685	8:143761931:C:T	0.0000	0.0716	0.0000	0.2029	0.7255	/	/
coloc	DU	5686	8:143761931:C:T	0.0000	0.0000	0.0000	0.0380	0.9620	/	/
coloc	BU	5674	8:143761931:C:T	0.0000	0.0299	0.0000	0.0827	0.8874	/	/
SuSiE-coloc	PUD	11729	8:143761931:C:T	0.0000	0.0000	0.0000	0.0874	0.9126	8:143761931:C:T	8:143771714:C:A
SuSiE-coloc	DU	11723	8:143761931:C:T	0.0000	0.0000	0.0000	0.0752	0.9248	8:143761931:C:T	8:143771712:A:C
coloc	PUD	4681	9:97031548:G:C	0.0770	0.6439	0.0257	0.2149	0.0385	/	/
coloc	GU	4669	9:97031548:G:C	0.3810	0.3127	0.1584	0.1300	0.0181	/	/
coloc	DU	4673	9:97031548:G:C	0.0383	0.6844	0.0128	0.2293	0.0351	/	/
coloc	BU	4654	9:97031548:G:C	0.1866	0.5212	0.0695	0.1942	0.0284	/	/
coloc	PUD	4875	9:136132908:T:TC	0.5672	0.1703	0.1916	0.0575	0.0133	/	/
coloc	GU	4864	9:136132908:T:TC	0.5361	0.1648	0.2176	0.0669	0.0146	/	/
coloc	DU	4863	9:136132908:T:TC	0.3566	0.3231	0.1560	0.1413	0.0229	/	/
coloc	BU	4851	9:136132908:T:TC	0.4301	0.2028	0.2344	0.1105	0.0222	/	/
coloc	PUD	4746	11:6268546:G:A	0.4393	0.1205	0.2916	0.0799	0.0686	/	/
coloc	GU	4740	11:6268546:G:A	0.5430	0.1831	0.1938	0.0653	0.0147	/	/
coloc	DU	4740	11:6268546:G:A	0.4181	0.1775	0.2661	0.1130	0.0254	/	/
coloc	BU	4730	11:6268546:G:A	0.4544	0.1841	0.2431	0.0985	0.0198	/	/
coloc	PUD	4124	13:28499583:CCCCTCCTCT:C	0.5282	0.2274	0.1589	0.0684	0.0171	/	/
coloc	GU	4116	13:28499583:CCCCTCCTCT:C	0.6210	0.1618	0.1628	0.0424	0.0120	/	/
coloc	DU	4119	13:28499583:CCCCTCCTCT:C	0.4248	0.1354	0.3091	0.0985	0.0322	/	/
coloc	BU	4106	13:28499583:CCCCTCCTCT:C	0.5496	0.1690	0.2032	0.0625	0.0158	/	/
coloc	PUD	4076	13:28567808:AAAC:A	0.5310	0.2258	0.1587	0.0675	0.0170	/	/
coloc	GU	4067	13:28567808:AAAC:A	0.6219	0.1621	0.1617	0.0422	0.0121	/	/
coloc	DU	4070	13:28567808:AAAC:A	0.4251	0.1339	0.3108	0.0979	0.0323	/	/
coloc	BU	4057	13:28567808:AAAC:A	0.5530	0.1683	0.2017	0.0614	0.0157	/	/
coloc	PUD	3383	17:39867248:C:T	0.6175	0.1507	0.1743	0.0425	0.0149	/	/
coloc	GU	3377	17:39867248:C:T	0.5920	0.1540	0.1893	0.0492	0.0155	/	/
coloc	DU	3375	17:39867248:C:T	0.4998	0.1905	0.2077	0.0791	0.0229	/	/
coloc	BU	3372	17:39867248:C:T	0.5253	0.2003	0.1832	0.0698	0.0214	/	/
coloc	PUD	3258	17:39913256:A:G	0.6185	0.1531	0.1710	0.0423	0.0152	/	/
coloc	GU	3252	17:39913256:A:G	0.5938	0.1545	0.1874	0.0487	0.0156	/	/
coloc	DU	3250	17:39913256:A:G	0.5064	0.1927	0.2018	0.0767	0.0223	/	/
coloc	BU	3247	17:39913256:A:G	0.5353	0.1946	0.1829	0.0665	0.0207	/	/
coloc	PUD	4296	19:49196275:A:G	0.5660	0.1899	0.1716	0.0576	0.0149	/	/
coloc	GU	4291	19:49196275:A:G	0.5891	0.1552	0.1919	0.0505	0.0132	/	/
coloc	DU	4289	19:49196275:A:G	0.3237	0.4061	0.1061	0.1331	0.0310	/	/
coloc	BU	4284	19:49196275:A:G	0.4974	0.2215	0.1814	0.0808	0.0189	/	/
coloc	PUD	3327	20:57477090:T:C	0.5506	0.2094	0.1534	0.0583	0.0282	/	/
coloc	GU	3319	20:57477090:T:C	0.5482	0.1368	0.2387	0.0595	0.0168	/	/
coloc	DU	3319	20:57477090:T:C	0.4693	0.3081	0.1081	0.0710	0.0435	/	/
coloc	BU	3311	20:57477090:T:C	0.5928	0.1716	0.1709	0.0495	0.0153	/	/
coloc	PUD	3937	22:25008477:C:CT	0.5624	0.1669	0.1947	0.0578	0.0183	/	/
coloc	GU	3932	22:25008477:C:CT	0.5557	0.1457	0.2251	0.0590	0.0144	/	/
coloc	DU	3931	22:25008477:C:CT	0.5228	0.2034	0.1823	0.0709	0.0206	/	/
coloc	BU	3927	22:25008477:C:CT	0.3248	0.1117	0.3962	0.1362	0.0310	/	/

3.9 Genetic analyses revealed the heterogeneity of GU

To further explore the similarities and differences in genetic architecture between GU and DU, the effect sizes of distinct signals identified in East Asians (lead variants and independent secondary variants) for GU and DU (Figure 15c; Table 24) were compared. Notably, the effect sizes for GU showed a strong correlation with those for DU (29 variants with $MAF > 0.01$; $r = 0.92$), which was concordant with the high genetic correlations described above. However, the effect sizes for GU were systematically smaller than those for DU (intercept = 0.01, slope = 0.43, and $SE_{slope} = 0.03$), with 19 variants ($19/29 = 65.5\%$) showing significant heterogeneity ($P_{het} < 0.05$) in Cochran's Q test (Figure 23; Table 24). To further verify the findings and avoid potential biases for the comparisons, multiple comparisons were performed including (1) the effect sizes of distinct signals of GWAS in TMM-50K, FinnGen²³, and UKB²⁰, (2) the effect sizes generated by excluding BU samples in the association tests in BBJ1-180K (i.e., no common case in the comparison), and (3) TMM-50K-derived statistics with BBJ1-180K-derived statistics (i.e., no common control in the comparison). In any of these comparisons, GU showed a high correlation (for variants with $MAF > 0.01$, $r = 0.75\text{--}0.90$) with DU, although with smaller effect sizes than DU (Figure 24; Figure 25; Figure 26). Furthermore, SBasesS⁴⁹ was utilized to estimate the SNP-heritability and polygenicity (defined as the proportion of SNPs with nonzero effects) using Hapmap3⁸² SNPs from EAS-specific summary statistics (Methods). Approximately 0.22% of the variants were estimated to have nonzero effects for PUD. Compared to DU, GU showed moderately higher polygenicity but lower heritability ($Pi_{GU} = 0.24\%$, $Pi_{DU} = 0.10\%$; Figure 27; Table 25). The results demonstrated that GU and DU showed a high genetic correlation with most risk loci shared and suggested higher heterogeneity of GU⁸³.

Finally, using the summary statistics derived from TMM-50K, polygenic risk score (PRS) models were generated with PRS-CS⁴⁵ comprising 1,029,637 variants and tested the PRS in BBJ1-180K for associations with PUD or PUD subtypes to investigate the genetic overlap among PUD and PUD subtypes. Compared to HP-positive PRS, HP-negative PRS generally showed stronger associations with PUD or PUD subtypes. The strongest association was between DU PRS and DU ($OR = 1.22$, confidence interval (CI) = 1.20–1.25, $\Delta R^2 = 0.94\%$). DU PRS ($OR = 1.08$, CI = 1.06–1.10, $\Delta R^2 = 0.13\%$) showed a stronger association with GU than GU PRS ($OR = 1.04$, CI = 1.03–1.06, $\Delta R^2 = 0.05\%$; Table 26). The results further validated that GU shares risk loci with DU while having higher heterogeneity than the latter⁸³.

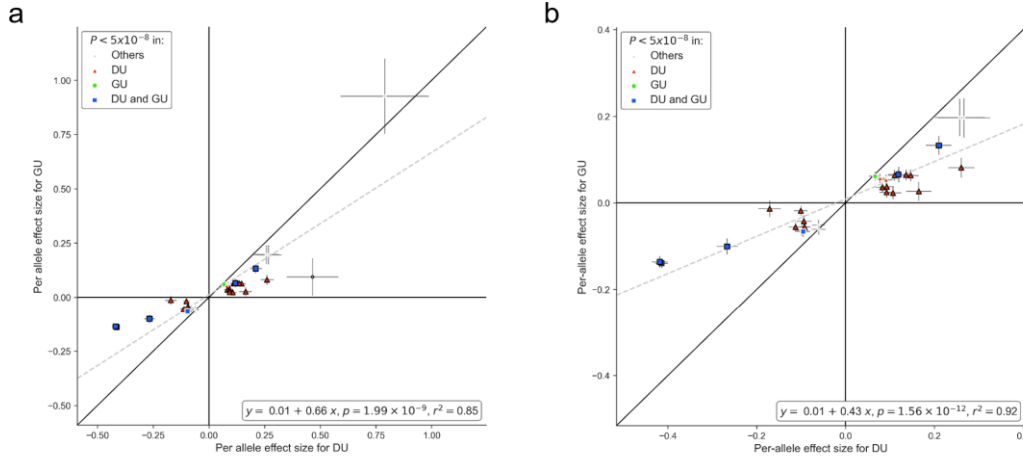


Figure 23. Effect size comparison of distinct signals for DU and GU in East Asian ancestry individuals.

Per-allele effect size ($\log(OR)$) comparison using EAS-specific summary statistics for DU and GU. Lead variants and independent secondary signals associated with PUD or any subtypes in the EAS population were selected for comparison. The most significant associations were shown if overlapping variants exist (interval < 500 kb). Variants with nominal significant heterogeneity ($P_{het} < 0.05$) were denoted by black marker edges. The grey dashed line represents the fitted linear regression line, whose parameters are shown in the bottom right; p, p-value obtained from t-test for slope. Marker colors denote the GWAS in which the significant variants are identified; Others, PUD or BU. **a**, all 31 available variants (existing in both datasets) are shown. **b**, 29 variants with MAF > 0.01 are shown.

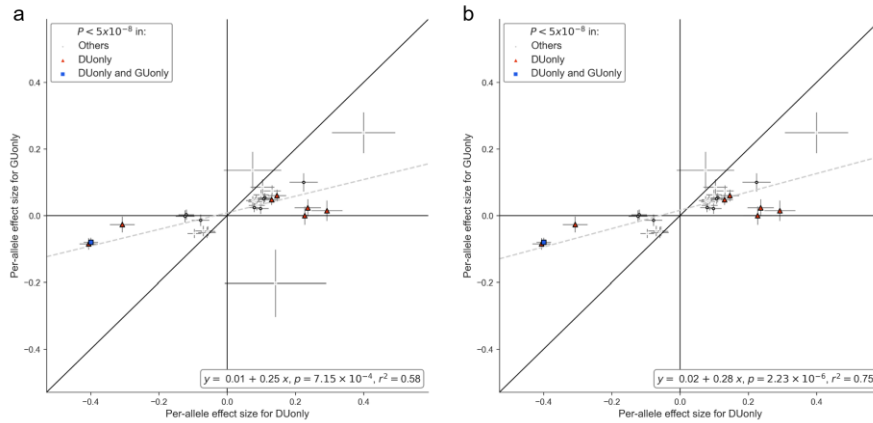


Figure 24. Effect size comparison of distinct signals for DUonly and GUonly in BBJ1-180K.

Per-allele effect size ($\log(OR)$) comparison using summary statistics for DUonly and GUonly in BBJ1-180K. Lead variants and independent secondary signals associated with PUD or any subtypes in the EAS population were selected for comparison. The most significant associations are shown if overlapping variants exist (interval < 500 kb). Variants with significant heterogeneity ($P_{het} < 0.05$) are denoted by black marker edges. The grey dashed line represents the fitted linear regression line, whose parameters were shown in the bottom right; p, p-value obtained from t-test for slope. Marker colors denote the GWAS in which the significant variants were identified; Others, PUD or BU. **a**, 30 available variants (existing in both datasets) are shown. **b**, 29 available variants with MAF > 0.01 are shown.

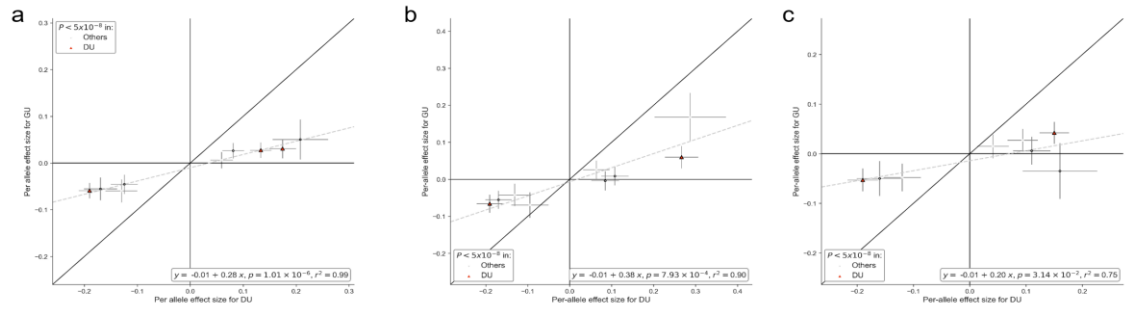


Figure 25. Effect size comparison of distinct signals for DU and GU in European ancestry individuals.

Per-allele effect size (log(OR)) comparison using summary statistics for DU and GU derived from EUR-specific meta-analysis (a), GWASs in UKB (b), and GWASs in FinnGen (c). Details were described in **Methods**. Lead variants associated with PUD or any subtypes in the EUR population were selected for comparison. The most significant associations were shown if overlapping variants exist (interval < 500 kb). Variants with significant heterogeneity ($P_{\text{het}} < 0.05$) are denoted by black marker edges. The grey dashed line represents the fitted linear regression line, whose parameters are shown in the bottom right; p, p-value obtained from t-test for slope. Marker colors denote the GWAS in which the significant variants were identified; Others, PUD. **a, b**, 9 available variants (existing in both datasets) are shown. **c**, 8 available variants are shown.

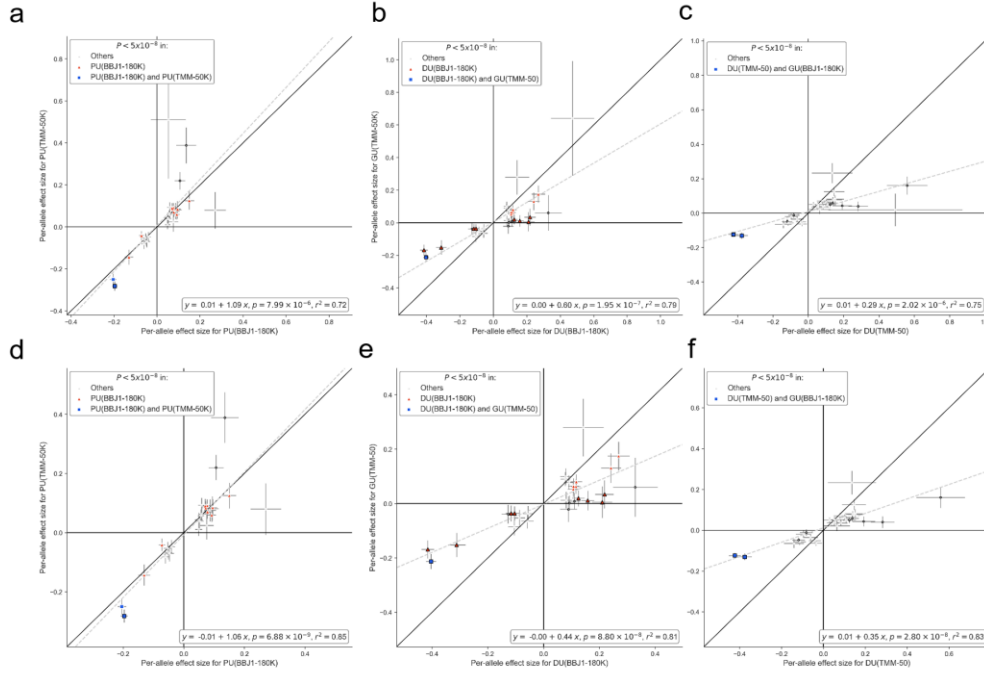


Figure 26. Cross-cohort effect size comparison of distinct signals for DU and GU in East Asian ancestry individuals.

Per-allele effect size (log(OR)) comparison using summary statistics for PUD (**a,d**), DU (**b,e**), and GU (**c,f**) derived from GWASs in BBJ1-180K and TMM-50K. Lead variants and independent secondary signals associated with PUD or any subtypes in the EAS population were selected for comparison. The most significant associations were shown if overlapping variants exist (interval < 500 kb). For PUD, TMM-50K-derived statistics were compared with BBJ1-180K-derived statistics(**a,d**). For DU and GU, TMM-50K-derived GU statistics were compared with BBJ1-180K-derived DU statistics (**b,e**), and TMM-50K-derived DU statistics were compared with BBJ1-180K-derived GU statistics (**c,f**). Variants with significant heterogeneity ($P_{het} < 0.05$) are denoted by black marker edges. The grey dashed line represents the fitted linear regression line, whose parameters are shown in the bottom right; p, p-value obtained from t-test for slope. Marker colors denote the GWAS in which the significant variants were identified; Others, PUD. **a**, 28, **b**, 27, **c**, 27 available variants (existing in both datasets) are shown. **d**, 27, **e**, 26, **f**, 26 available variants with MAF>0.01 are shown.

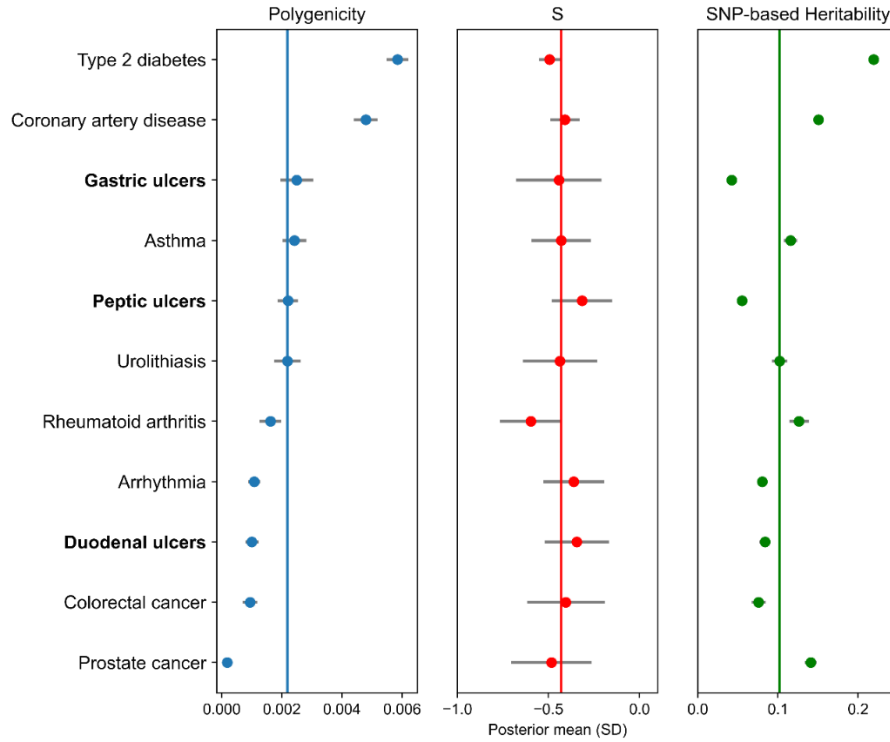


Figure 27. Polygenicity estimation in East Asians using SBayesS.

Summary statistics for binary phenotypes from BBJ1⁴³ and meta-analysis of PUD and its subtypes in East Asians were used for SBayesS analysis. Only phenotypes that converged in the SBayesS model are shown. Phenotypes are ranked based on polygenicity estimates. SNP-based heritability estimates are on the liability scale. The vertical lines indicate the median estimates of the 11 phenotypes. Error bars show the standard deviations (SD) of the posterior means.

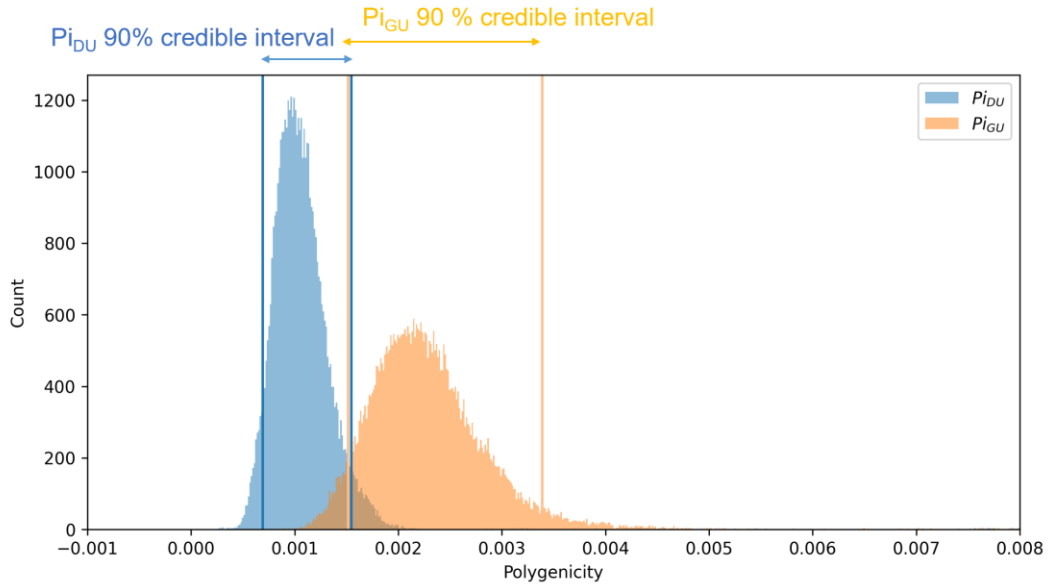


Figure 28. Posterior distribution of polygenicity estimates for GU and DU in EAS using SBayesS.

The histogram shows the posterior distributions of the polygenicity for GU (Pi_{GU}) and DU (Pi_{DU}) estimated by SBayesS. Blue lines, the upper and lower bounds for the 90% credible interval for Pi_{DU} . Yellow lines, the upper and lower bounds for the 90% credible interval for Pi_{GU} .

Table 24. Effect size comparison of PUD signals between DU and GU in East Asians.

SNPID	EA	NEA	DU			GU			Q	HetP	HetP<0.05
			BETA	SE	P	BETA	SE	P			
1:155133066:G:GGT	G	GGT	-0.1709	0.0254	1.70E-11	-0.0136	0.0194	4.83E-01	24.222	8.58E-07	True
2:182502860:T:C	C	T	0.7893	0.1982	6.82E-05	0.9262	0.1733	9.03E-08	0.27038	6.03E-01	False
2:219963550:C:T	T	C	0.0912	0.0155	3.70E-09	0.0532	0.0119	7.74E-06	3.78149	5.18E-02	False
3:16981683:G:T	T	G	0.0778	0.0136	9.77E-09	0.0565	0.0105	6.37E-08	1.53684	2.15E-01	False
3:169136565:C:G	G	C	-0.1008	0.0151	2.54E-11	-0.0183	0.0116	1.16E-01	18.7722	1.47E-05	True
3:171491650:T:G	G	T	-0.0939	0.0172	4.83E-08	-0.0425	0.0133	1.35E-03	5.58873	1.81E-02	True
4:47634802:A:AT	A	AT	0.1069	0.0184	5.88E-09	0.0228	0.0142	1.08E-01	13.0929	2.96E-04	True
4:124538120:T:C	C	T	-0.0606	0.0158	1.19E-04	-0.0609	0.0123	7.47E-07	0.00022	9.88E-01	False
5:40729974:T:C	C	T	0.1189	0.0138	5.55E-18	0.0655	0.0106	6.39E-10	9.41731	2.15E-03	True
6:26162679:G:A	A	G	0.1651	0.0283	5.16E-09	0.0266	0.0215	2.15E-01	15.1862	9.74E-05	True
6:131582218:C:T	T	C	0.2666	0.0586	5.35E-06	0.1961	0.0454	1.54E-05	0.90448	3.42E-01	False
7:127046841:G:A	A	G	0.2573	0.0567	5.62E-06	0.1973	0.0433	5.09E-06	0.7073	4.00E-01	False
7:127560541:G:A	A	G	0.2101	0.0287	2.65E-13	0.1326	0.022	1.68E-09	4.59302	3.21E-02	True
8:143761931:C:T	T	C	-0.4143	0.0141	7.69E-189	-0.1396	0.0107	1.37E-38	240.856	2.56E-54	True
8:143776719:A:C	C	A	-0.4185	0.0177	6.64E-124	-0.1366	0.0135	4.07E-24	160.366	9.41E-37	True
8:143842101:G:A	A	G	-0.2668	0.0241	1.54E-28	-0.1009	0.0184	4.33E-08	29.9366	4.46E-08	True
8:143868011:T:C	C	T	0.2607	0.0296	1.29E-18	0.0814	0.023	3.94E-04	22.8789	1.73E-06	True
9:80649703:A:C	C	A	0.0831	0.0148	2.11E-08	0.0358	0.0114	1.72E-03	6.41057	1.13E-02	True
9:97031548:G:C	C	G	-0.0611	0.0146	2.91E-05	-0.0513	0.0112	4.96E-06	0.28364	5.94E-01	False
9:136132908:T:TC	TC	T	-0.0924	0.0135	8.96E-12	-0.0556	0.0104	9.22E-08	4.6632	3.08E-02	True
11:6268546:G:A	A	G	0.1198	0.0238	5.01E-07	0.0649	0.0183	4.01E-04	3.34396	6.75E-02	False
11:68826155:A:G	G	A	0.093	0.0159	4.98E-09	0.0372	0.0123	2.43E-03	7.70512	5.51E-03	True
13:28499583:CCCCTCCTCT:C	CCCCTCCTCT	C	-0.0955	0.0151	2.30E-10	-0.0665	0.0116	9.97E-09	2.31955	1.28E-01	False
13:28567808:AAAC:A	A	AAAC	0.1468	0.0181	4.90E-16	0.0633	0.0139	5.07E-06	13.3871	2.53E-04	True
14:95298230:C:G	G	C	0.4655	0.1158	5.86E-05	0.0943	0.086	2.73E-01	6.6227	1.01E-02	True
17:39867248:C:T	T	C	0.1107	0.0157	1.58E-12	0.0637	0.0122	1.69E-07	5.58774	1.81E-02	True
17:39913256:A:G	G	A	0.0664	0.0136	1.07E-06	0.0613	0.0105	5.13E-09	0.08811	7.67E-01	False
19:49196275:A:G	G	A	0.1363	0.0162	3.38E-17	0.0644	0.0125	2.85E-07	12.3471	4.42E-04	True
20:6673542:T:C	C	T	0.0925	0.0159	5.70E-09	0.0245	0.0122	4.45E-02	11.5125	6.91E-04	True
20:57471283:A:AT	A	AT	-0.1126	0.0152	1.32E-13	-0.056	0.0117	1.76E-06	8.70698	3.17E-03	True
22:25000229:A:T	T	A	-0.0988	0.0155	2.09E-10	-0.0638	0.012	1.12E-07	3.18803	7.42E-02	False

Table 25. Polygenicity estimation using SBayesS in East Asians.

Observed-scale heritability was converted to liability-scaled heritability based on the prevalence estimated from population-based TMM cohort.

Phenotype	Pi	Pi_sd	S	S_sd	hsq	hsq_sd	Pi_R	S_R	hsq_R
Prostate cancer	0.000188	0.000039	-0.48356	0.220839	0.141007	0.007582	1.015	1.019	1.001
Colorectal cancer	0.000943	0.00024	-0.40414	0.21285	0.075844	0.008617	1.006	1.006	1.005
Duodenal ulcers (DU)	0.001006	0.000224	-0.34389	0.175802	0.083928	0.006978	1.007	1.007	1.004
Arrhythmia	0.001087	0.000201	-0.36174	0.167148	0.080281	0.005321	1	1.003	1
Rheumatoid arthritis	0.001622	0.000363	-0.59612	0.170432	0.126278	0.012269	1.002	1.002	1
Urolithiasis	0.002189	0.000446	-0.43735	0.204089	0.102012	0.00948	1.005	1.013	1.004
Peptic ulcers (PUD)	0.002206	0.000341	-0.31507	0.165144	0.055262	0.003621	1.003	1.003	1.001
Asthma	0.002423	0.000402	-0.42997	0.164024	0.115913	0.008455	1.004	1.008	1.002
Gastric ulcers (GU)	0.002495	0.000544	-0.44268	0.235138	0.042341	0.004272	1.012	1.011	1.01
Coronary artery disease	0.004798	0.000395	-0.40937	0.080503	0.150682	0.004648	1.001	1	1
Type 2 diabetes	0.005846	0.000358	-0.49422	0.056417	0.219323	0.004496	1	1.001	1

Table 26. Associations of polygenic risk scores with PUD and PUD subtypes in BBJ1-180K.

HP+ : GWAS conducted in individuals with HP-positive status. HP- : GWAS conducted in individuals with HP-negative status.

Phenotype	TMM-50K derived PRS model	OR	P	95% C.I.	AUC (base model)	AUC (PRS included)	Improvement in AUC (Δ AUC)	Liability-scale R2 (base model)	Liability-scale R2 (PRS included)	Improvement in R2 (Δ R2)
PUD	PUD(HP+) PRS	1.05	3.01E-12	(1.04 , 1.07)	0.6179	0.6194	0.0014	0.0463	0.0471	0.0007
	PUD(HP-) PRS	1.09	1.13E-29	(1.07 , 1.11)	0.6179	0.6215	0.0036	0.0463	0.0483	0.0020
	PU PRS	1.12	8.85E-48	(1.10 , 1.13)	0.6179	0.6235	0.0056	0.0463	0.0496	0.0032
	HP PRS	1.01	2.19E-01	(0.99 , 1.02)	0.6179	0.6180	0.0000	0.0463	0.0464	0.0000
	GU PRS	1.07	2.51E-18	(1.05 , 1.09)	0.6179	0.6200	0.0020	0.0463	0.0475	0.0012
	GU(HP+) PRS	1.03	2.14E-04	(1.01 , 1.04)	0.6179	0.6183	0.0003	0.0463	0.0466	0.0002
	GU(HP-) PRS	1.06	4.20E-15	(1.05 , 1.08)	0.6179	0.6197	0.0018	0.0463	0.0473	0.0009
	DU PRS	1.12	2.96E-47	(1.10 , 1.13)	0.6179	0.6236	0.0057	0.0463	0.0495	0.0032
	DU(HP+) PRS	1.07	5.28E-17	(1.05 , 1.08)	0.6179	0.6201	0.0021	0.0463	0.0474	0.0011
	DU(HP-) PRS	1.10	3.30E-33	(1.08 , 1.11)	0.6179	0.6221	0.0042	0.0463	0.0485	0.0022
	BU PRS	1.08	2.70E-21	(1.06 , 1.09)	0.6179	0.6206	0.0026	0.0463	0.0477	0.0014
	BU(HP+) PRS	1.02	2.31E-03	(1.01 , 1.04)	0.6179	0.6182	0.0002	0.0463	0.0465	0.0001
	BU(HP-) PRS	1.06	6.41E-15	(1.05 , 1.08)	0.6179	0.6199	0.0019	0.0463	0.0473	0.0009
GU	PUD(HP+) PRS	1.04	3.45E-05	(1.02 , 1.06)	0.6235	0.6240	0.0004	0.0419	0.0422	0.0003
	PUD(HP-) PRS	1.06	1.79E-12	(1.05 , 1.08)	0.6235	0.6248	0.0012	0.0419	0.0428	0.0009
	PU PRS	1.08	1.03E-19	(1.06 , 1.10)	0.6235	0.6255	0.0020	0.0419	0.0434	0.0015
	HP PRS	1.01	4.04E-01	(0.99 , 1.02)	0.6235	0.6236	0.0000	0.0419	0.0419	0.0000
	GU PRS	1.05	7.98E-08	(1.03 , 1.07)	0.6235	0.6242	0.0007	0.0419	0.0424	0.0005
	GU(HP+) PRS	1.02	3.00E-02	(1.00 , 1.04)	0.6235	0.6236	0.0001	0.0419	0.0420	0.0001
	GU(HP-) PRS	1.04	9.66E-07	(1.03 , 1.06)	0.6235	0.6241	0.0006	0.0419	0.0423	0.0004
	DU PRS	1.08	1.32E-17	(1.06 , 1.10)	0.6235	0.6253	0.0018	0.0419	0.0432	0.0013
	DU(HP+) PRS	1.05	3.91E-08	(1.03 , 1.07)	0.6235	0.6243	0.0008	0.0419	0.0424	0.0005
	DU(HP-) PRS	1.06	5.91E-12	(1.04 , 1.08)	0.6235	0.6247	0.0011	0.0419	0.0427	0.0008
	BU PRS	1.05	1.73E-09	(1.04 , 1.07)	0.6235	0.6244	0.0008	0.0419	0.0425	0.0006
	BU(HP+) PRS	1.02	2.79E-02	(1.00 , 1.04)	0.6235	0.6236	0.0001	0.0419	0.0420	0.0001
	BU(HP-) PRS	1.04	1.35E-06	(1.03 , 1.06)	0.6235	0.6241	0.0006	0.0419	0.0423	0.0004
DU	PUD(HP+) PRS	1.11	2.72E-17	(1.08 , 1.14)	0.6176	0.6297	0.0121	0.0466	0.0489	0.0023
	PUD(HP-) PRS	1.16	1.21E-34	(1.13 , 1.19)	0.6176	0.6359	0.0183	0.0466	0.0515	0.0049
	PU PRS	1.22	4.37E-58	(1.19 , 1.25)	0.6176	0.6412	0.0236	0.0466	0.0550	0.0084
	HP PRS	1.02	1.03E-01	(1.00 , 1.04)	0.6176	0.6189	0.0013	0.0466	0.0466	0.0001
	GU PRS	1.13	2.83E-22	(1.10 , 1.15)	0.6176	0.6318	0.0141	0.0466	0.0496	0.0031
	GU(HP+) PRS	1.06	4.66E-06	(1.03 , 1.08)	0.6176	0.6234	0.0057	0.0466	0.0472	0.0007
	GU(HP-) PRS	1.11	2.17E-17	(1.08 , 1.14)	0.6176	0.6295	0.0119	0.0466	0.0489	0.0023
	DU PRS	1.23	1.99E-64	(1.20 , 1.26)	0.6176	0.6428	0.0252	0.0466	0.0559	0.0094
	DU(HP+) PRS	1.12	1.51E-19	(1.09 , 1.14)	0.6176	0.6306	0.0130	0.0466	0.0492	0.0027
	DU(HP-) PRS	1.18	1.48E-43	(1.16 , 1.21)	0.6176	0.6385	0.0209	0.0466	0.0528	0.0062
	BU PRS	1.13	6.21E-25	(1.11 , 1.16)	0.6176	0.6332	0.0156	0.0466	0.0500	0.0035
	BU(HP+) PRS	1.04	2.46E-03	(1.01 , 1.06)	0.6176	0.6209	0.0033	0.0466	0.0469	0.0003
	BU(HP-) PRS	1.11	4.45E-18	(1.09 , 1.14)	0.6176	0.6297	0.0121	0.0466	0.0490	0.0024
BU	PUD(HP+) PRS	1.12	1.56E-06	(1.07 , 1.17)	0.6259	0.6378	0.0119	0.0365	0.0385	0.0020
	PUD(HP-) PRS	1.15	4.41E-09	(1.10 , 1.20)	0.6259	0.6404	0.0145	0.0365	0.0395	0.0030
	PU PRS	1.21	1.50E-16	(1.16 , 1.27)	0.6259	0.6470	0.0212	0.0365	0.0424	0.0059
	HP PRS	1.04	1.41E-01	(0.99 , 1.08)	0.6259	0.6279	0.0020	0.0365	0.0367	0.0002
	GU PRS	1.12	1.29E-06	(1.07 , 1.17)	0.6259	0.6373	0.0114	0.0365	0.0385	0.0020
	GU(HP+) PRS	1.07	7.07E-03	(1.02 , 1.12)	0.6259	0.6309	0.0050	0.0365	0.0371	0.0006
	GU(HP-) PRS	1.10	4.30E-05	(1.05 , 1.15)	0.6259	0.6348	0.0089	0.0365	0.0379	0.0015
	DU PRS	1.22	6.02E-18	(1.17 , 1.28)	0.6259	0.6492	0.0233	0.0365	0.0430	0.0065
	DU(HP+) PRS	1.12	6.86E-07	(1.07 , 1.18)	0.6259	0.6380	0.0121	0.0365	0.0386	0.0022
	DU(HP-) PRS	1.15	1.31E-09	(1.10 , 1.21)	0.6259	0.6418	0.0160	0.0365	0.0397	0.0032
	BU PRS	1.13	8.16E-08	(1.08 , 1.19)	0.6259	0.6397	0.0138	0.0365	0.0390	0.0025
	BU(HP+) PRS	1.05	5.85E-02	(1.00 , 1.09)	0.6259	0.6288	0.0029	0.0365	0.0368	0.0003
	BU(HP-) PRS	1.11	9.59E-06	(1.06 , 1.16)	0.6259	0.6366	0.0107	0.0365	0.0382	0.0017

3. 10 Pleiotropy of PUD risk variants on GC

Considering that DU appears to be a protective factor against GC¹⁰, two-sample Mendelian randomization (MR)⁴¹ was conducted to evaluate the causality of PUD or its subtypes on GC. Summary statistics for GC in EAS were obtained from a previous study in BBJ1-180K, which included approximately 6,500 cases⁴³ (Methods). Summary statistics for PUD and its subtypes were obtained by conducting a meta-analysis combining three replication datasets. Although PUD and its subtypes showed significant ($P < 0.05/15$) protective effects against GC using the Inverse variance weighted (IVW) method (Table 27), MR-Egger analysis suggested significant pleiotropy for the instruments (Table 28). MR-PRESSO⁴² was utilized to correct for the horizontal pleiotropic variants (ranging from 6 to 7 for each exposure). The outlier-corrected MR showed no significant effects of PUD or its subtypes on GC (Table 29). Of note, splitting samples and removing outliers may cause power loss (Figure 29).

To evaluate the pleiotropic effects of PUD risk variants on GC risk, the effect sizes of distinct signals were compared in EAS between PUD subtypes and GC. For 23 available variants existing in both datasets, the effect sizes for DU were found to be negatively correlated with that for GC (slope = -0.33, $SE_{\text{slope}} = 0.10$; Figure 15d; Figure 30; Table 30). It was noteworthy that lead variants linked to *EFNA1* (Ephrin A1, a member of the EFN family), *PTGER4* (receptor for prostaglandin E2, PGE2), and *PSCA* showed relatively strong but opposite effects on DU and GC (Table 30). This suggested that the alleles of these variants, which increased the risk for PUD, could decrease the risk for GC. When removing the three variants from the regression, the negative correlation was not observed for the 20 variants (slope = -0.06, $SE = 0.10$; Supplementary Figure 25), indicating that the negative correlation between DU and GC was largely affected by the three variants.

3. 11 Gene-based and gene-set analysis

To investigate the associations at gene level, a gene-level analysis using MAGMA^{50,51} was performed. In this analysis, 9 genes significantly associated with PUD ($P < 6.5 \times 10^{-7}$; Table 31), 45 genes associated with DU, and 15 genes associated with GU, were detected in the EAS population. In total, 47 distinct genes were associated with PUD or PUD subtypes. Of note, multiple genes identified in gene-level analysis were reportedly related to GC (*PTGER4*⁸⁴, *PRKAA1*⁸⁵, *GNAQ*⁸⁶, *GNAS*⁸⁷, *NHEJ1*⁸⁸, *IHH*⁸⁹, and *JUP*⁹⁰). Next, based on the gene-level statistics, pathway enrichment analysis was additionally conducted and identified one gene set after Bonferroni correction (nikolsky_breast_cancer_8q23_q24_amplicon⁹¹, including genes within amplicon 8q23-q24 identified in a study of breast tumors; $P < 8.0 \times 10^{-7}$; Table 32)

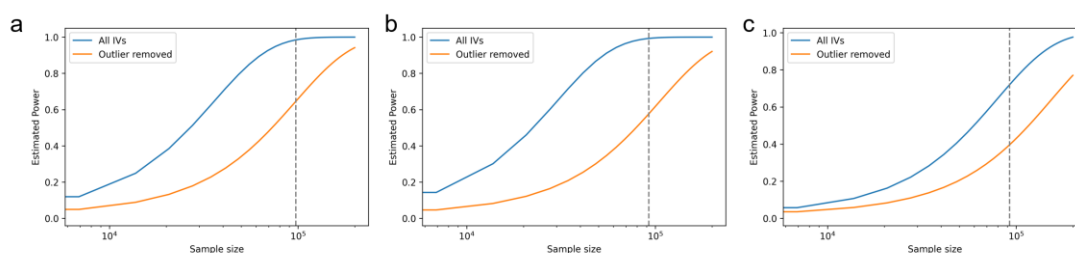


Figure 29. Statistical power estimation for Mendelian randomization analysis.

The proportion of variance (r^2) explained by genetic instruments was approximated by the sum of the explained

variance by the selected variants. Grey dashed lines represent the sample sizes. Significance, 0.0166; odds ratio, 1.2.

a, Power estimation for the analysis of PUD; the ratio of cases to controls, 1:8.7; sample size, 97,523. **b**, Power estimation for the analysis of DU; the ratio of cases to controls, 1:18; sample size, 92,294. **c**, Power estimation for the analysis of GU; the ratio of cases to controls, 1:13; sample size, 94,056.

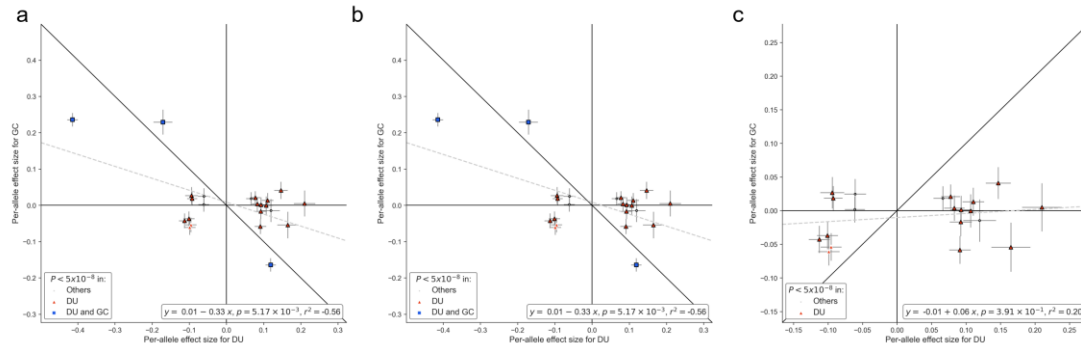


Figure 30. Effect size comparison of distinct signals between PUD and GC in East Asian ancestry individuals.

Per-allele effect size (β) comparison between DU and GC. Summary statistics for DU were obtained from EAS-specific meta-analysis. Summary statistics for GC were obtained from previous GWAS conducted in BBJ1 (Methods). Lead variants and independent secondary signals associated with PUD or any subtypes in the EAS population were selected for comparison. The most significant associations are shown if overlapping variants exist (interval < 500 kb). **a**, 23 available variants (existing in both datasets) are shown. **b**, 23 available variants with MAF > 0.01 are shown. **c**, 20 variants (excluding lead variants at *EFNA1*, *PTGER4*, and *PSCA*) are shown.

Table 27. Two-sample Mendelian randomization study of the causality of PUD or its subtypes on GC.

Exposure	Outcome	Method	Number of variants	BETA	SE	P
PUD	GC	Inverse variance weighted (fixed effects)	22	-0.47099	0.048993	7.03E-22
PUD	GC	Inverse variance weighted (multiplicative random effects)	22	-0.47099	0.14787	1.45E-03
PUD	GC	MR Egger	22	-1.09197	0.229971	1.23E-04
PUD	GC	Weighted median	22	-0.35791	0.160108	2.54E-02
PUD	GC	Weighted mode	22	-0.87744	0.105775	4.58E-08
GU	GC	Inverse variance weighted (fixed effects)	22	-0.45806	0.062082	1.60E-13
GU	GC	Inverse variance weighted (multiplicative random effects)	22	-0.45806	0.223923	4.08E-02
GU	GC	MR Egger	22	-1.32491	0.4235	5.29E-03
GU	GC	Weighted median	22	0.047611	0.132556	7.19E-01
GU	GC	Weighted mode	22	0.114757	0.131681	3.93E-01
DU	GC	Inverse variance weighted (fixed effects)	23	-0.35886	0.032549	2.89E-28
DU	GC	Inverse variance weighted (multiplicative random effects)	23	-0.35886	0.099775	3.22E-04
DU	GC	MR Egger	23	-0.63914	0.147855	3.00E-04
DU	GC	Weighted median	23	-0.53955	0.068781	4.35E-15
DU	GC	Weighted mode	23	-0.50813	0.054669	4.49E-09

Table 28. Tests for directional pleiotropy and heterogeneity in MR.

MR-egger	Exposure	Outcome	Egger intercept	SE	P
	DU	GC	0.049	0.020	2.59E-02
	GU	GC	0.073	0.031	3.02E-02
	PUD	GC	0.068	0.021	4.51E-03
Heterogeneity (Inverse variance weighted)	Exposure	Outcome	Q	Degree of freedom	P
	DU	GC	206.725	22	5.47E-32
	GU	GC	273.206	21	8.66E-46
	PUD	GC	191.296	21	1.85E-29

Table 29. Outlier-corrected MR using the MR-PRESSO method.

Outcome	Exposure	MR Analysis	Causal Estimate (BETA)	SD	T statistics	P	Horizontal pleiotropic outliers (SNPID)
GC	PUD	Raw	-0.471	0.148	-3.185	4.45E-03	13:28499583:cccctcctct:c,13:28567808:aaac:a,20:57471283:a:at,22:25000229:a:t,5:40729974:t:c,8:143761931:c:t
GC	PUD	Outlier-corrected	-0.067	0.097	-0.694	4.98E-01	13:28499583:cccctcctct:c,13:28567808:aaac:a,20:57471283:a:at,22:25000229:a:t,5:40729974:t:c,8:143761931:c:t
GC	GU	Raw	-0.458	0.224	-2.046	5.35E-02	1:155133066:g:ggg,13:28499583:cccctcctct:c,17:39913256:a:g,22:25000229:a:t,5:40729974:t:c,8:143761931:c:t
GC	GU	Outlier-corrected	-0.009	0.126	-0.068	9.47E-01	1:155133066:g:ggg,13:28499583:cccctcctct:c,17:39913256:a:g,22:25000229:a:t,5:40729974:t:c,8:143761931:c:t
GC	DU	Raw	-0.359	0.100	-3.597	1.60E-03	1:155133066:g:ggg,13:28499583:cccctcctct:c,20:57471283:a:at,22:25000229:a:t,3:169136565:c:g,5:40729974:t:c,8:143761931:c:t
GC	DU	Outlier-corrected	-0.056	0.071	-0.796	4.39E-01	1:155133066:g:ggg,13:28499583:cccctcctct:c,20:57471283:a:at,22:25000229:a:t,3:169136565:c:g,5:40729974:t:c,8:143761931:c:t

Table 30. Effect size comparison of PUD signals between DU and GC in East Asians.

8 variants were not available for effect comparison.

SNPID	EA	NEA	DU			GC			Q	HetP	HetP<0.05
			BETA	SE	P	BETA	SE	P			
1:155133066:G:GGT	G	GGT	-0.1709	0.0254	1.695E-11	0.2290	0.0343	2.456E-11	87.782	7.307E-21	True
2:219963550:C:T	T	C	0.0912	0.0155	3.704E-09	-0.0585	0.0207	4.629E-03	33.594	6.789E-09	True
3:16981683:G:T	T	G	0.0778	0.0136	9.773E-09	0.0208	0.0180	2.477E-01	6.381	1.153E-02	True
3:169136565:C:G	G	C	-0.1008	0.0151	2.539E-11	-0.0371	0.0202	6.605E-02	6.386	1.150E-02	True
3:171491650:T:G	G	T	-0.0939	0.0172	4.83E-08	0.0267	0.0233	2.529E-01	17.301	3.190E-05	True
4:47634802:A:AT	A	AT	0.1069	0.0184	5.875E-09	-0.0004	0.0247	9.861E-01	12.155	4.895E-04	True
4:124538120:T:C	C	T	-0.0606	0.0158	0.0001193	0.0246	0.0223	2.689E-01	9.742	1.801E-03	True
5:40729974:T:C	C	T	0.1189	0.0138	5.552E-18	-0.1643	0.0182	2.103E-19	153.310	3.277E-35	True
6:26162679:G:A	A	G	0.1651	0.0283	5.155E-09	-0.0544	0.0364	1.351E-01	22.651	1.943E-06	True
7:127560541:G:A	A	G	0.2101	0.0287	2.652E-13	0.0049	0.0357	8.909E-01	20.104	7.335E-06	True
8:143761931:C:T	T	C	-0.4143	0.0141	7.69E-189	0.2357	0.0187	2.570E-36	768.824	3.240E-169	True
9:80649703:A:C	C	A	0.0831	0.0148	2.105E-08	0.0037	0.0188	8.453E-01	11.025	8.989E-04	True
9:97031548:G:C	C	G	-0.0611	0.0146	2.909E-05	0.0018	0.0194	9.273E-01	6.725	9.506E-03	True
9:136132908:T:TC	TC	T	-0.0924	0.0135	8.956E-12	0.0185	0.0181	3.066E-01	24.183	8.760E-07	True
11:6268546:G:A	A	G	0.1198	0.0238	5.011E-07	-0.0147	0.0315	6.396E-01	11.628	6.495E-04	True
11:68826155:A:G	G	A	0.093	0.0159	4.984E-09	0.0012	0.0208	9.540E-01	12.331	4.454E-04	True
13:28499583:CCCCTCCTCT:C	CCCCTCCTCT	C	-0.0955	0.0151	2.303E-10	-0.0540	0.0204	7.929E-03	2.677	1.018E-01	False
13:28567808:AAAC:A	A	AAAC	0.1468	0.0181	4.897E-16	0.0409	0.0236	8.320E-02	12.650	3.757E-04	True
17:39867248:C:T	T	C	0.1107	0.0157	1.58E-12	0.0131	0.0208	5.264E-01	14.052	1.778E-04	True
17:39913256:A:G	G	A	0.0664	0.0136	1.068E-06	0.0182	0.0179	3.092E-01	4.580	3.235E-02	True
20:6673542:T:C	C	T	0.0925	0.0159	5.697E-09	-0.0168	0.0211	4.263E-01	17.079	3.586E-05	True
20:57471283:A:AT	A	AT	-0.1126	0.0152	1.318E-13	-0.0428	0.0202	3.453E-02	7.622	5.766E-03	True
22:25000229:A:T	T	A	-0.0988	0.0155	2.089E-10	-0.0609	0.0206	3.039E-03	2.161	1.416E-01	False

Table 31. Gene-based analysis for PUD and PUD subtypes in East Asian individuals.

Phenotype	Gene ID	Gene name	CHR	START(GRCh37)	STOP(GRCh37)	NSNPS	NPARAM	N	ZSTAT	P
PUD	ENSG00000167656	LY6D	8	143866296	143868008	11	5	230705	7.9347	1.0547E-15
PUD	ENSG00000180155	LYNX1	8	143845752	143859640	77	26	236832	7.9223	1.1657E-15
PUD	ENSG00000132356	PRKAA1	5	40759481	40798476	128	26	260123	7.8357	2.3315E-15
PUD	ENSG00000198576	ARC	8	143692405	143696833	20	6	229612	7.8327	2.387E-15
PUD	ENSG00000113638	TTC33	5	40714577	40756077	141	26	250574	7.7394	4.996E-15
PUD	ENSG000000087460	GNAS	20	57414773	57486247	197	45	237439	7.2172	2.6534E-13
PUD	ENSG00000183463	URAD	13	28551851	28562791	39	11	238823	7.0976	6.3488E-13
PUD	ENSG00000139515	PDX1	13	28494157	28500368	20	6	249201	6.8055	5.0338E-12
PUD	ENSG00000173801	JUP	17	39775692	39943183	653	49	239954	6.5685	2.5413E-11
PUD	ENSG00000154822	PLCL2	3	16844159	17132086	1061	53	244162	6.3643	9.8098E-11
PUD	ENSG00000100031	GGT1	22	24979718	25024972	153	22	233750	6.3097	1.3975E-10
PUD	ENSG00000100028	SNRPD3	22	24951471	25005947	192	33	239249	6.2717	1.7861E-10
PUD	ENSG00000213901	SLC23A3	2	219940051	220035549	294	28	249619	6.1396	4.1363E-10
PUD	ENSG00000171522	PTGER4	5	40679600	40693837	33	11	257521	6.1094	5E-10
PUD	ENSG00000167653	PSCA	8	143751726	143764142	80	8	265918	6.1094	5E-10
PUD	ENSG00000160886	LY6K	8	143781529	143786545	28	7	246929	6.1094	5E-10
PUD	ENSG00000130193	THEM6	8	143808621	143818345	32	11	256614	6.1094	5E-10
PUD	ENSG00000187736	NHEJ1	2	219940039	220025587	269	25	248942	6.0393	7.7401E-10
PUD	ENSG00000163501	IIHH	2	219919142	219925189	16	4	269813	5.9676	1.2037E-09
PUD	ENSG00000181790	BAI1	8	143530791	143626370	408	87	237109	5.7669	4.0363E-09
PUD	ENSG00000184502	GAST	17	39868578	39872221	13	6	234834	5.6329	8.8599E-09
PUD	ENSG00000130950	NUTM2F	9	97077605	97090926	55	18	242256	5.3728	3.8757E-08
PUD	ENSG00000173805	HAP1	17	39873994	39890896	88	9	239755	5.2981	5.8492E-08
PUD	ENSG00000110148	CCKBR	11	6280966	6293357	71	12	253670	5.256	7.3594E-08
PUD	ENSG00000175787	ZNF169	9	97021593	97063736	233	33	219153	5.1462	1.3288E-07
PUD	ENSG00000106328	FSCN3	7	127231463	127242198	49	18	246538	5.1162	1.5591E-07
PUD	ENSG000000085276	MECOM	3	168801287	169381406	2571	180	250154	5.1112	1.6006E-07
PUD	ENSG00000179562	GCC1	7	127220672	127233665	52	14	243854	5.0927	1.7647E-07
PUD	ENSG00000004059	ARF5	7	127228399	127231759	13	5	248133	4.9602	3.5218E-07
GU	ENSG00000198576	ARC	8	143692405	143696833	20	6	222085	7.7394	4.996E-15
GU	ENSG00000130193	THEM6	8	143808621	143818345	32	11	248265	7.6627	9.1038E-15
GU	ENSG00000160886	LY6K	8	143781529	143786545	28	7	238932	7.5607	2.004E-14
GU	ENSG00000167653	PSCA	8	143751726	143764142	80	8	257305	7.4116	6.2395E-14
GU	ENSG00000100031	GGT1	22	24979718	25024972	153	22	226144	6.1094	5E-10
GU	ENSG00000180155	LYNX1	8	143845752	143859640	77	26	229142	5.9796	1.1186E-09
GU	ENSG00000171522	PTGER4	5	40679600	40693837	33	11	249175	5.979	1.1227E-09
GU	ENSG00000113638	TTC33	5	40714577	40756077	141	26	242463	5.9575	1.2805E-09
GU	ENSG000000087460	GNAS	20	57414773	57486247	197	45	229496	5.735	4.8747E-09
GU	ENSG00000167656	LY6D	8	143866296	143868008	11	5	223291	5.6716	7.0736E-09
GU	ENSG00000132356	PRKAA1	5	40759481	40798476	128	26	251696	5.55	1.4286E-08
GU	ENSG00000173801	JUP	17	39775692	39943183	653	49	232136	5.4812	2.1121E-08
GU	ENSG00000139515	PDX1	13	28494157	28500368	20	6	241111	5.0365	2.371E-07
GU	ENSG00000213901	SLC23A3	2	219940051	220035549	294	28	241526	4.9044	4.6862E-07
GU	ENSG00000154822	PLCL2	3	16844159	17132086	1061	53	236122	4.8447	6.3388E-07
DU	ENSG00000183463	URAD	13	28551851	28562791	39	11	223185	8.0766	3.3307E-16
DU	ENSG00000167656	LY6D	8	143866296	143868008	11	5	215358	7.9817	7.2164E-16
DU	ENSG00000130193	THEM6	8	143808621	143818345	32	11	239797	7.7731	3.8303E-15
DU	ENSG00000171522	PTGER4	5	40679600	40693837	33	11	240546	7.5099	2.9587E-14
DU	ENSG00000167653	PSCA	8	143751726	143764142	80	8	248164	7.4853	3.5694E-14
DU	ENSG00000181790	BAI1	8	143530791	143626370	408	88	221692	7.2831	1.6306E-13
DU	ENSG000000087460	GNAS	20	57414773	57486247	197	45	221588	6.8862	2.8659E-12
DU	ENSG00000169242	EFNA1	1	155099936	155107333	22	9	182827	6.6502	1.4636E-11
DU	ENSG00000100028	SNRPD3	22	24951471	25005947	192	33	223393	6.2781	1.7137E-10
DU	ENSG00000158373	HIST1H2BD	6	26158349	26171577	64	12	236677	6.1792	3.2211E-10
DU	ENSG00000113638	TTC33	5	40714577	40756077	141	26	234014	6.1094	5E-10
DU	ENSG00000132356	PRKAA1	5	40759481	40798476	128	26	242939	6.1094	5E-10
DU	ENSG00000198576	ARC	8	143692405	143696833	20	6	214700	6.1094	5E-10
DU	ENSG00000160886	LY6K	8	143781529	143786545	28	7	230654	6.1094	5E-10
DU	ENSG00000180155	LYNX1	8	143845752	143859640	77	26	221289	6.1094	5E-10
DU	ENSG00000100031	GGT1	22	24979718	25024972	153	22	218411	6.1094	5E-10
DU	ENSG00000213901	SLC23A3	2	219940051	220035549	294	28	233199	5.7578	4.2614E-09
DU	ENSG00000158406	HIST1H4H	6	26281283	26285762	16	6	246308	5.7546	4.3435E-09
DU	ENSG00000173801	JUP	17	39775692	39943183	653	49	224217	5.7438	4.6296E-09
DU	ENSG00000139515	PDX1	13	28494157	28500368	20	6	232897	5.7099	5.6524E-09
DU	ENSG00000163501	IIHH	2	219919142	219925189	16	4	252077	5.6807	6.707E-09
DU	ENSG00000187736	NHEJ1	2	219940039	220025587	269	25	232566	5.6598	7.5757E-09
DU	ENSG00000206503	HLA-A	6	29909037	29913661	123	16	248408	5.4728	2.2151E-08
DU	ENSG00000126233	SLURP1	8	143822362	143823829	3	2	252639	5.4544	2.4565E-08
DU	ENSG00000145244	CORIN	4	47596015	47840123	854	66	234165	5.4499	2.5201E-08
DU	ENSG00000154822	PLCL2	3	16844159	17132086	1060	52	228239	5.4358	2.728E-08
DU	ENSG00000184502	GAST	17	39868578	39872221	13	6	219486	5.4245	2.9058E-08
DU	ENSG00000004059	ARF5	7	127228399	127231759	13	5	231821	5.3485	4.4338E-08
DU	ENSG00000185499	MUC1	1	155158300	155162707	5	3	190879	5.3236	5.0869E-08
DU	ENSG00000179562	GCC1	7	127220672	127233665	52	14	227830	5.274	6.6728E-08
DU	ENSG00000106328	FSCN3	7	127231463	127242198	49	18	230321	5.2398	8.0372E-08
DU	ENSG000000085276	MECOM	3	168801287	169381406	2569	170	233761	5.1656	1.1986E-07
DU	ENSG00000163463	KRTCAP2	1	155141884	155145951	10	3	252639	5.1579	1.2487E-07
DU	ENSG00000173805	HAP1	17	39873994	39890896	88	9	224049	5.087	1.8184E-07
DU	ENSG00000198366	HIST1H3A	6	26020718	26021186	3	2	252639	5.0518	2.1881E-07
DU	ENSG00000143590	EFNA3	1	155036224	155060014	34	9	217223	5.0329	2.4152E-07
DU	ENSG00000100101	NOL12	22	38077680	38170137	282	35	219278	5.0243	2.5267E-07
DU	ENSG00000251246	EFNA3	1	155036224	155059283	33	10	216150	5.0148	2.6538E-07
DU	ENSG00000124508	BTN2A2	6	26383324	26395102	55	13	239679	4.9946	2.9475E-07
DU	ENSG00000104499	GML	8	143915663	143997922	434	23	237375	4.9709	3.3317E-07
DU	ENSG00000187837	HIST1H1C	6	26055968	26056699	8	5	245619	4.9685	3.3729E-07
DU	ENSG00000273088	RP11-201K1	1	155141885	155159748	37	10	230880	4.9191	4.3475E-07
DU	ENSG00000112763	BTN2A1	6	26458150	26476849	99	21	243413	4.9045	4.6828E-07
DU	ENSG00000169231	THBS3	1	155165379	155178842	25	9	197143	4.8981	4.8374E-07
DU	ENSG00000156052	GAOQ	9	80331003	80646374	833	57	221068	4.8501	6.1703E-07
BU	ENSG00000160886	LY6K	8	143781529	143786545	28	7	222657	7.9223	1.1657E-15
BU	ENSG00000198576	ARC	8	143692405	143696833	20	6	207172	7.7887	3.3862E-15
BU	ENSG00000167653	PSCA	8	143751726	143764142	80	8	239558	7.5222	2.6923E-14
BU	ENSG00000180155	LYNX1	8	143845752	143859640	77	26	213600	6.5107	3.7405E-11
BU	ENSG00000167656	LY6D	8	143866296	143868008	11	5	207944	6.3994	7.7982E-11
BU	ENSG00000130193	THEM6	8	143808621	143818345	32	11	231447	6.1094	5E-10
BU	ENSG000000087460	GNAS	20	57414773	57486247	197	45	213843	5.4236	2.9211E-08
BU	ENSG00000181790	BAI1	8	143530791	143626370	408	88	213644	5.2018	9.8674E-08

Table 32. Gene-set enrichment analysis for PUD and PUD subtypes in East Asian individuals.

Top 10 gene sets for each phenotype were shown.

Phenotype	Gene set	NGENES	BETA	BETA STD	SE	P
PUD	Curated gene sets:nikolsky breast cancer 8q23 q24 amplicon	154	0.56929	0.051002	0.11365	2.76E-07
PUD	Curated gene sets:doane breast cancer esr1 dn	45	0.55052	0.026738	0.14222	5.44E-05
PUD	GO bp:go regulation of defense response to virus by host	35	0.54838	0.023495	0.1465	9.11E-05
PUD	GO bp:go negative regulation of type b pancreatic cell apoptotic process	6	1.609	0.028565	0.43682	0.000115
PUD	GO bp:go peripheral nervous system myelin maintenance	7	0.99849	0.019146	0.27751	0.000161
PUD	GO bp:go positive regulation of molecular function	1655	0.082052	0.02312	0.023119	0.000194
PUD	Curated gene sets:reactome organic cation transport	10	1.1135	0.025516	0.3198	0.00025
PUD	GO mf:go fucosyltransferase activity	13	0.85026	0.022214	0.24463	0.000255
PUD	GO bp:go response to oxygen radical	26	0.52166	0.019268	0.1521	0.000303
PUD	GO bp:go type b pancreatic cell apoptotic process	10	1.1321	0.025943	0.33028	0.000305
GU	GO bp:go regulation of arp2 3 complex mediated actin nucleation	15	0.91969	0.025809	0.21525	9.71E-06
GU	GO bp:go histone h3 k36 methylation	11	1.1806	0.028374	0.30018	4.21E-05
GU	GO bp:go arp2 3 complex mediated actin nucleation	31	0.54543	0.021995	0.14675	0.000101
GU	GO bp:go negative regulation of type b pancreatic cell apoptotic process	6	1.5776	0.028007	0.42753	0.000112
GU	GO bp:go negative regulation of b cell receptor signaling pathway	8	1.024	0.02099	0.28116	0.000136
GU	GO bp:go peptide modification	4	1.57	0.022758	0.43789	0.000169
GU	GO bp:go regulation of defense response to virus by host	35	0.51076	0.021883	0.14334	0.000184
GU	GO bp:go regulation of actin nucleation	27	0.57759	0.021739	0.1645	0.000224
GU	GO bp:go endosome to lysosome transport via multivesicular body sorting pathway	6	1.4072	0.024982	0.41954	0.000399
GU	GO bp:go actin nucleation	47	0.40698	0.0202	0.12273	0.000457
DU	Curated gene sets:nikolsky breast cancer 8q23 q24 amplicon	154	0.64576	0.057853	0.11302	5.62E-09
DU	Curated gene sets:reactome glucocorticoid biosynthesis	9	1.2342	0.026832	0.28993	1.04E-05
DU	Curated gene sets:nikolsky breast cancer 1q21 amplicon	38	1.0192	0.045496	0.24511	1.61E-05
DU	GO mf:go dopamine neurotransmitter receptor activity coupled via gi go	7	1.2213	0.023418	0.31972	6.70E-05
DU	GO bp:go endochondral bone morphogenesis	70	0.42987	0.026022	0.11277	6.92E-05
DU	GO bp:go adenylate cyclase inhibiting dopamine receptor signaling pathway	9	1.0685	0.02323	0.28232	7.72E-05
DU	Curated gene sets:udayakumar med1 targets dn	223	0.22964	0.024712	0.060961	8.29E-05
DU	GO bp:go bone morphogenesis	110	0.33456	0.025361	0.09221	0.000143
DU	Curated gene sets:wang barretts esophagus dn	24	0.70151	0.024896	0.19486	0.00016
DU	GO bp:go regulation of microtubule polymerization	43	0.50455	0.023956	0.14107	0.000174
BU	GO bp:go regulation of neurotransmitter receptor activity	65	0.49628	0.028954	0.11817	1.34E-05
BU	GO bp:go regulation of t helper 1 type immune response	27	0.69718	0.026241	0.16761	1.60E-05
BU	GO bp:go regulation of mesoderm development	16	0.83849	0.024301	0.20599	2.36E-05
BU	GO bp:go positive regulation of t helper 1 type immune response	20	0.80574	0.026106	0.19985	2.78E-05
BU	GO cc:go ficolin 1 rich granule	121	0.30589	0.024313	0.07638	3.12E-05
BU	Curated gene sets:bhat esr1 targets via akt1 dn	83	0.38364	0.025279	0.097131	3.93E-05
BU	Curated gene sets:bystrykh hematopoiesis stem cell qtl trans	723	0.12196	0.023315	0.031604	5.71E-05
BU	GO mf:go acetylcholine receptor binding	11	1.0521	0.025287	0.27296	5.82E-05
BU	GO cc:go spermatoproteasome complex	4	1.2906	0.018708	0.33576	6.08E-05
BU	Curated gene sets:lin npas4 targets dn	68	0.398	0.023747	0.1048	7.33E-05

3. 12 Tissue- and cell-type specificity analysis

To investigate the tissue types related to PUD and its subtypes, the tissue-level specificity was tested employing MAGMA⁵⁰ with GTEx v8 datasets⁷⁵ in EAS individuals. Significant genetic enrichments (FDR < 5%) were observed in the stomach, pancreas, small intestine, and kidney for PUD, in the stomach, pancreas, and prostate for DU, and in the stomach for GU (Figure 31a).

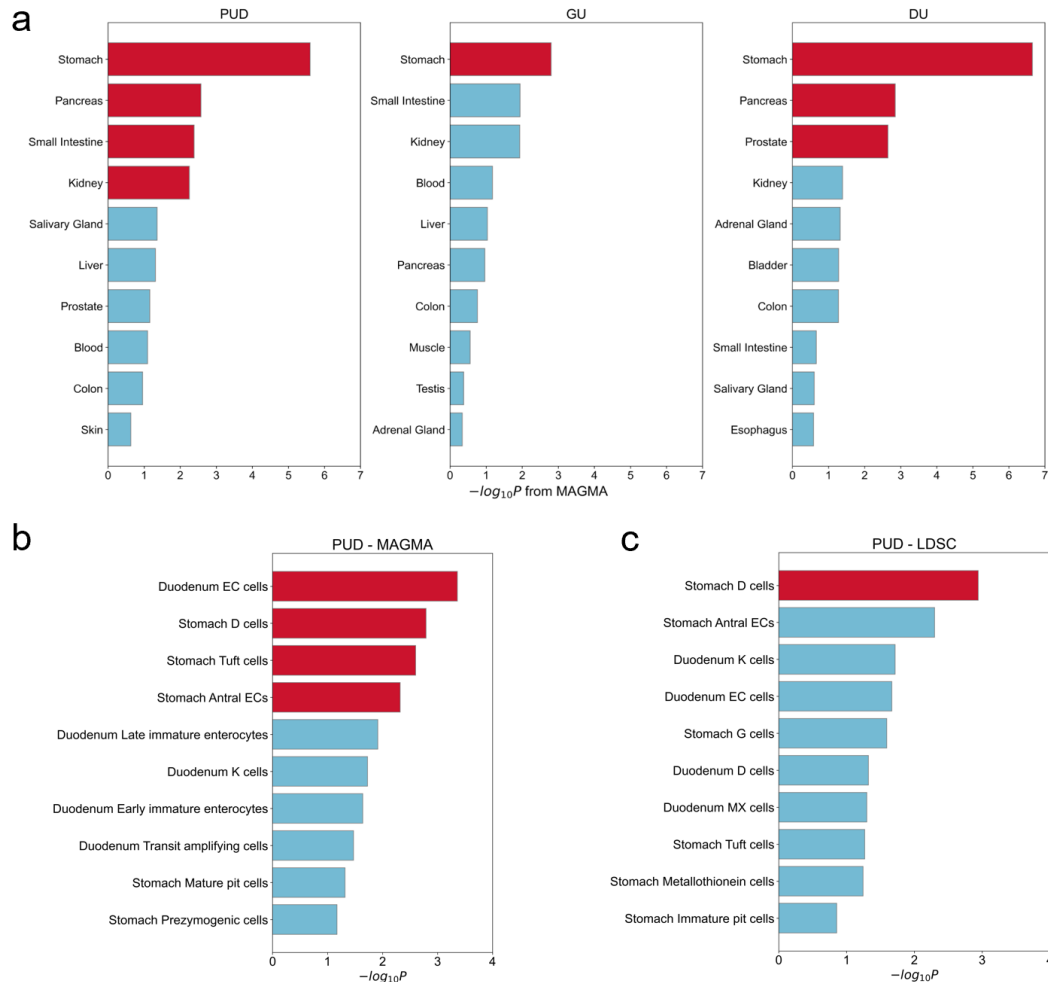


Figure 31. Tissue- and cell-type specificity analysis.

a. Associations across peptic ulcer phenotypes and 30 general tissue types were analyzed using MAGMA with East Asian-specific summary statistics and the GTEx version 8 dataset. b, c. Associations between PUD and cell types in the stomach and duodenum were analyzed using MAGMA and LDSC (testing for enrichment of the 10% most specific genes in each cell type). Inverse variance weighted meta-analysis combined statistics from East Asian and European ancestries for each method. x-axis, $-\log_{10}(P)$ derived from meta-analyzed estimates. a, b, c. Red bars imply significant associations after corrections (FDR < 5%). Only the ten most significant associations for each phenotype are shown.

To further characterize specific cell types associated with PUD in the gastric and duodenal tissues, publicly available single-cell RNA sequencing (scRNA-seq) datasets of the human stomach and duodenum⁵³ were obtained. The top 10% of the most specific genes for each cell type were selected as the cell-specific gene sets (Methods). Cell specificity analysis was performed using LDSC⁹² and

MAGMA⁵⁰ in EAS and EUR individuals, respectively. To increase statistical power, a fixed-effect meta-analysis was conducted combining EAS and EUR results for each method (Methods; Figure 32. Cell-type specificity analysis in East Asian ancestry individuals using LDSC.). For PUD, stomach D cells reached the significance threshold ($FDR < 5\%$) in both the analyses of MAGMA and LDSC (Figure 31b-c; Figure 32 - Figure 37; Table 33; Table 34). Additionally, duodenal enterochromaffin cells (EC cells), stomach antral ECs, and stomach tuft cells were significantly ($FDR < 5\%$) associated with PUD, as per MAGMA (Fig. 6b). Somatostatin produced by stomach D cells inhibits the secretion of a variety of gastrointestinal hormones, including the gastrin secreted by stomach G cells that stimulate gastric acid secretion. EC cells secrete serotonin (5-HT, a neurotransmitter) with diverse gastrointestinal functions, and tuft cells (chemosensory epithelial cells) secrete interleukin-25 (IL-25) driving type 2 immune response to parasitic infection. Together, the findings suggested the important role of gastrointestinal hormone regulation and immune response in PUD etiology.

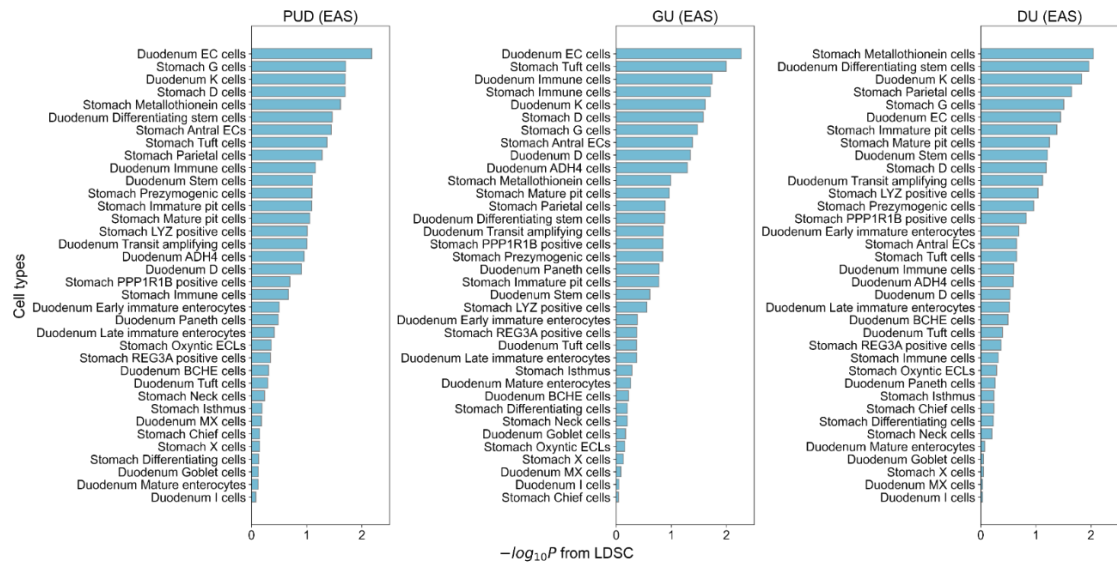


Figure 32. Cell-type specificity analysis in East Asian ancestry individuals using LDSC.

Associations between PUD and cell types in the stomach and duodenum were analyzed using LDSC (testing for the enrichment of the 10% most specific genes in each cell type; **Methods**). East Asian-specific summary statistics were used in the analysis. X axis, $-\log_{10}(P)$ derived from LDSC estimates. Color bars indicate whether the association is significant (Red, FDR < 5%; Light blue, FDR ≥ 5%).

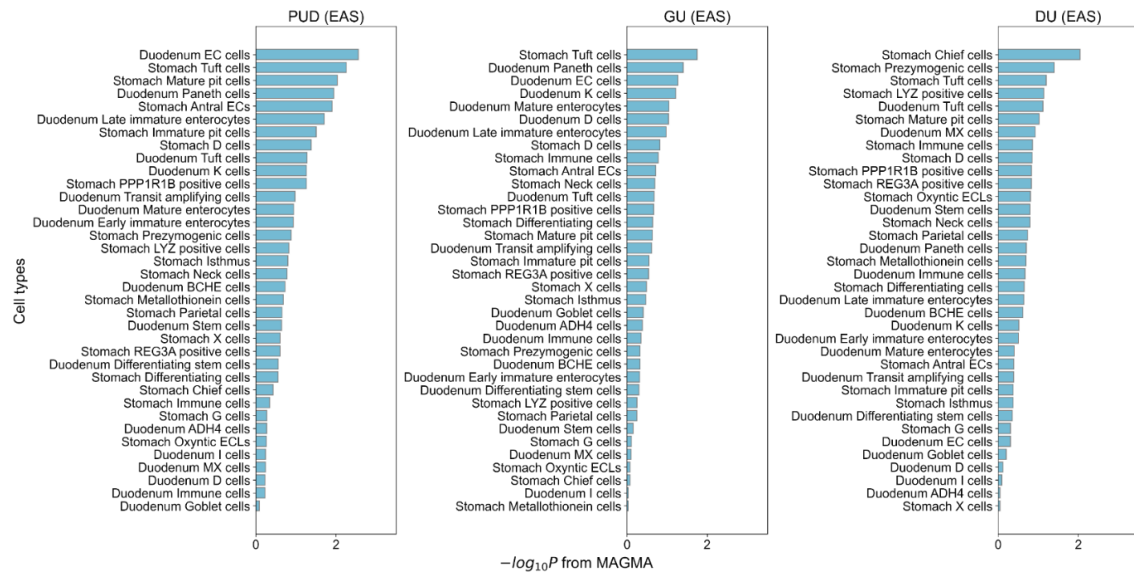


Figure 33. Cell-type specificity analysis in East Asian ancestry individuals using MAGMA.

Associations between PUD and cell types in the stomach and duodenum were analyzed using MAGMA (testing for the enrichment of the 10% most specific genes in each cell type; **Methods**). East Asian-specific summary statistics were used in the analysis. X axis, $-\log_{10}(P)$ derived from MAGMA estimates. Color bars indicate whether the association is significant (Red, FDR < 5%; Light blue, FDR ≥ 5%).

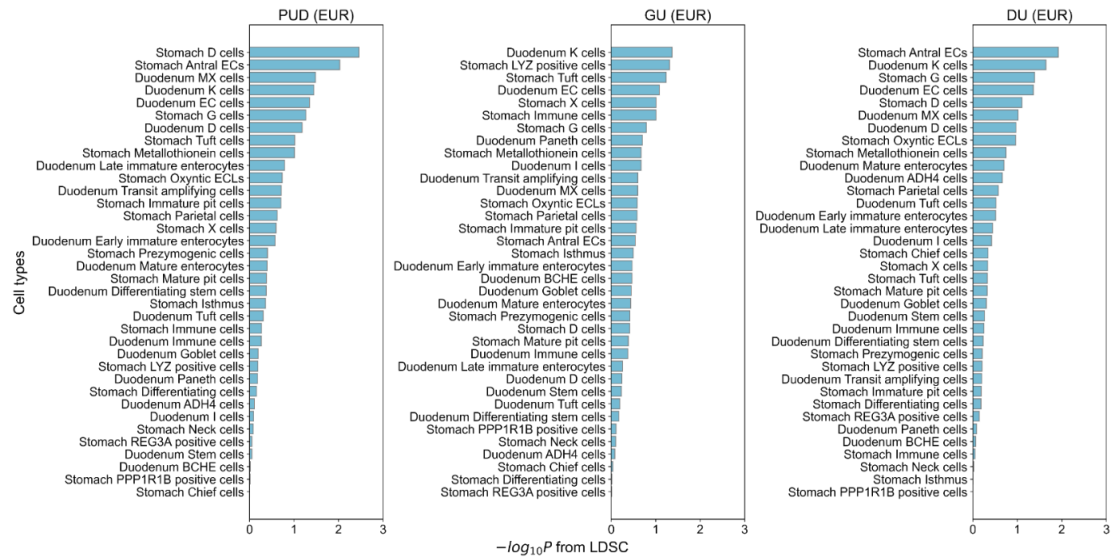


Figure 34. Cell-type specificity analysis in European ancestry individuals using LDSC.

Associations between PUD and cell types in the stomach and duodenum were analyzed using LDSC (testing for the enrichment of the 10% most specific genes in each cell type; **Methods**). European-specific summary statistics were used in the analysis. X axis, $-\log_{10}(P)$ derived from LDSC estimates. Color bars indicate whether the association is significant (Red, FDR < 5%; Light blue, FDR ≥ 5%).

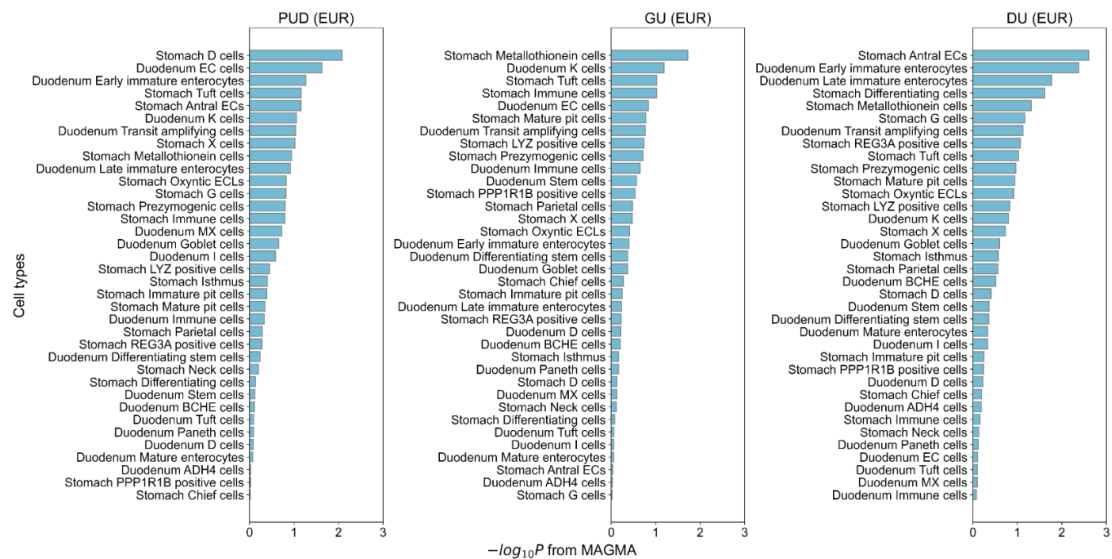


Figure 35. Cell-type specificity analysis in European ancestry individuals using MAGMA.

Associations between PUD and cell types in the stomach and duodenum were analyzed using MAGMA (testing for the enrichment of the 10% most specific genes in each cell type; **Methods**). European-specific summary statistics were used in the analysis. X axis, $-\log_{10}(P)$ derived from MAGMA estimates. Color bars indicate whether the association is significant (Red, FDR < 5%; Light blue, FDR ≥ 5%).

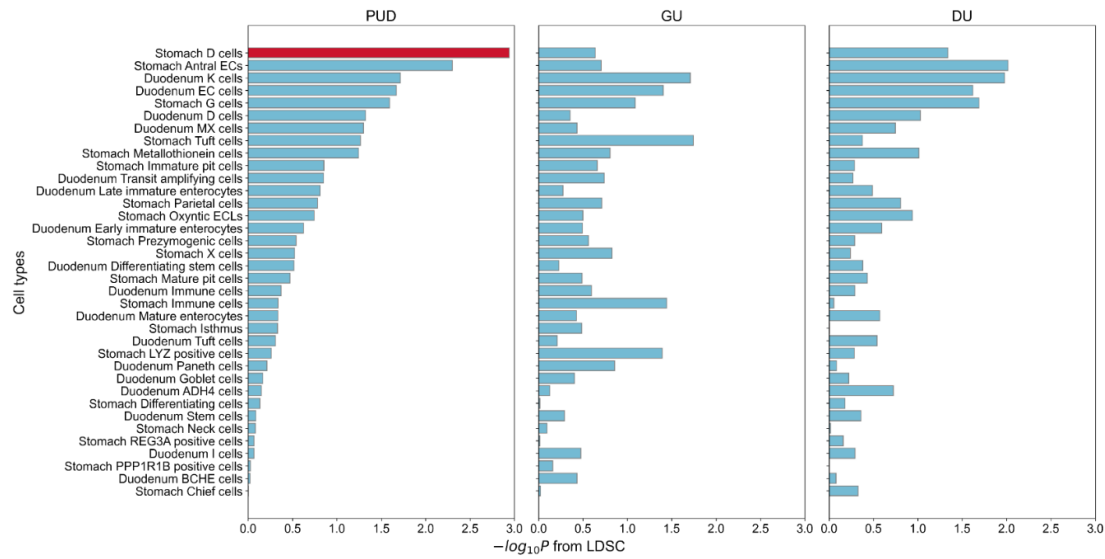


Figure 36. Cross-ancestry meta-analysis of cell-type specificity using LDSC.

Associations between PUD and cell types in the stomach and duodenum were analyzed using LDSC (testing for the enrichment of the 10% most specific genes in each cell type). An inverse variance-weighted meta-analysis was performed by combining statistics from EAS and EUR ancestries. X axis, $-\log_{10}(P)$ derived from meta-analyzed estimates. Color bars indicate whether the enrichment is significant (Red, FDR < 5%; Light blue, FDR ≥ 5%) in meta-analyzed results.

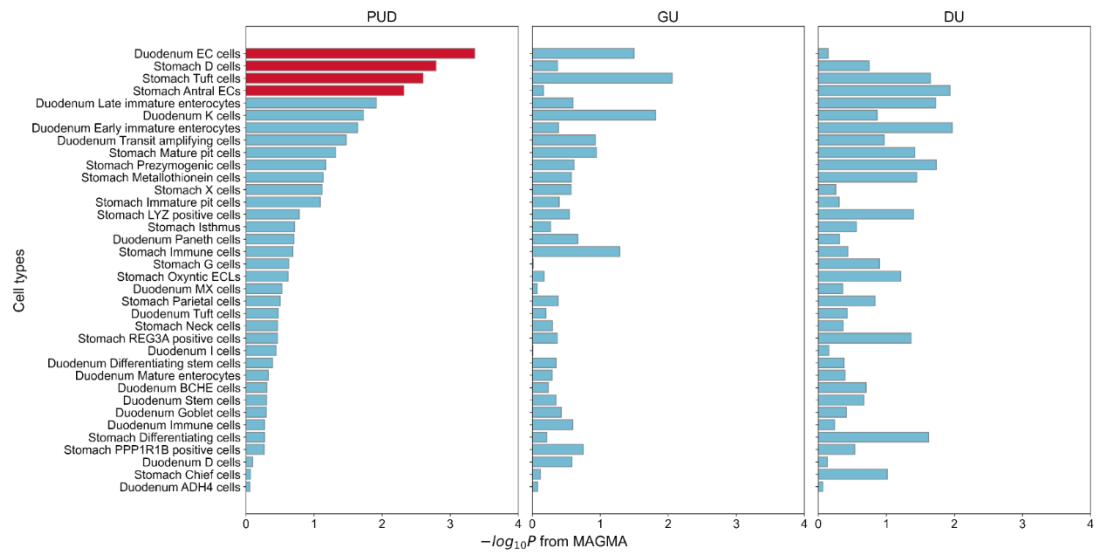


Figure 37. Cross-ancestry meta-analysis of cell-type specificity using MAGMA.

Associations between PUD and cell types in the stomach and duodenum were analyzed using MAGMA (testing for the enrichment of the 10% most specific genes in each cell type). An inverse variance-weighted meta-analysis was performed by combining statistics from EAS and EUR ancestries. X axis, $-\log_{10}(P)$ derived from meta-analyzed estimates. Color bars indicate whether the enrichment is significant (Red, FDR < 5%; Light blue, FDR ≥ 5%) in meta-analyzed results.

Table 33. Cell type specificity analysis using LDSC.

Meta-analysis was performed using inverse variance weighted (IVW) methods. P-value was derived from upper-tail tests.

Phenotype	Cell type	EAS			EUR			Meta-analysis			
		Coefficient_EAS	Coefficient_std_error_EAS	Coefficient_P_value_EAS	Coefficient_t_EUR	Coefficient_std_error_EUR	Coefficient_P_value_EUR	Coefficient_FE	SE_META	Z_META	P_META (one-sided)
PUD	Stomach D cells	4.89E-09	2.38E-09	2.01E-02	1.29E-09	4.77E-10	3.46E-03	1.43E-09	4.68E-10	3.05	1.14E-03
PUD	Stomach Antral ECs	6.64E-09	3.68E-09	3.57E-02	1.20E-09	5.09E-10	9.38E-03	1.30E-09	5.04E-10	2.57	5.01E-03
PUD	Duodenum K cells	8.46E-09	4.12E-09	2.00E-02	1.04E-09	5.80E-10	3.59E-02	1.19E-09	5.74E-10	2.07	1.93E-02
PUD	Duodenum EC cells	9.22E-09	3.72E-09	6.60E-03	8.81E-10	5.19E-10	4.47E-02	1.04E-09	5.14E-10	2.02	2.14E-02
PUD	Stomach G cells	4.96E-09	2.41E-09	1.96E-02	7.11E-10	4.43E-10	5.43E-02	8.50E-10	4.36E-10	1.95	2.55E-02
PUD	Duodenum D cells	3.45E-09	2.98E-09	1.24E-01	7.07E-10	4.69E-10	6.56E-02	7.73E-10	4.63E-10	1.67	4.74E-02
PUD	Duodenum MX cells	-6.94E-10	1.69E-09	6.59E-01	9.38E-10	5.10E-10	3.29E-02	8.02E-10	4.88E-10	1.64	5.02E-02
PUD	Stomach Tuft cells	3.73E-09	2.17E-09	4.28E-02	5.46E-10	4.19E-10	9.61E-02	6.61E-10	4.11E-10	1.61	5.41E-02
PUD	Stomach Metallothionein cells	6.17E-09	3.13E-09	2.43E-02	6.27E-10	4.85E-10	9.79E-02	7.58E-10	4.79E-10	1.58	5.70E-02
PUD	Stomach Immature pit cells	3.84E-09	2.75E-09	8.12E-02	4.13E-10	4.85E-10	1.98E-01	5.16E-10	4.78E-10	1.08	1.40E-01
PUD	Duodenum Transit amplifying cells	4.15E-09	3.22E-09	9.86E-02	4.97E-10	5.78E-10	1.95E-01	6.11E-10	5.68E-10	1.08	1.41E-01
PUD	Duodenum Late immature enterocytes	6.57E-10	2.32E-09	3.89E-01	5.04E-10	5.13E-10	1.63E-01	5.11E-10	5.01E-10	1.02	1.54E-01
PUD	Stomach Parietal cells	4.01E-09	2.47E-09	5.26E-02	3.00E-10	4.23E-10	2.39E-01	4.05E-10	4.17E-10	0.97	1.66E-01
PUD	Stomach Oxyntic ECLs	3.08E-10	1.96E-09	4.38E-01	4.29E-10	4.76E-10	1.84E-01	4.23E-10	4.63E-10	0.91	1.81E-01
PUD	Duodenum Early immature enterocytes	1.16E-09	2.39E-09	3.13E-01	3.08E-10	4.90E-10	2.65E-01	3.42E-10	4.80E-10	0.71	2.38E-01
PUD	Stomach Prezygogenic cells	3.06E-09	2.18E-09	8.04E-02	1.27E-10	4.81E-10	3.96E-01	2.63E-10	4.70E-10	0.56	2.88E-01
PUD	Stomach X cells	-9.28E-10	1.71E-09	7.06E-01	2.65E-10	3.99E-10	2.54E-01	2.03E-10	3.89E-10	0.52	3.01E-01
PUD	Duodenum Differentiating stem cells	6.54E-09	3.60E-09	3.44E-02	1.27E-10	6.14E-10	4.18E-01	3.09E-10	6.05E-10	0.51	3.05E-01
PUD	Stomach Mature pit cells	3.89E-09	2.86E-09	8.71E-02	9.30E-11	4.37E-10	4.16E-01	1.79E-10	4.32E-10	0.42	3.39E-01
PUD	Duodenum Immune cells	3.63E-09	2.45E-09	6.93E-02	-4.53E-11	4.78E-10	5.38E-01	8.91E-11	4.69E-10	0.19	4.25E-01
PUD	Stomach Immune cells	1.33E-09	1.68E-09	2.15E-01	-3.66E-11	4.08E-10	5.36E-01	3.93E-11	3.96E-10	0.10	4.61E-01
PUD	Duodenum Mature enterocytes	-1.84E-09	2.65E-09	7.57E-01	1.42E-10	5.73E-10	4.02E-01	5.34E-11	5.60E-10	0.10	4.62E-01
PUD	Stomach Isthmus	-9.05E-10	2.45E-09	6.44E-01	8.54E-11	5.08E-10	4.33E-01	4.45E-11	4.97E-10	0.09	4.64E-01
PUD	Duodenum Tuft cells	-4.26E-11	2.56E-09	5.07E-01	1.05E-11	5.03E-10	4.92E-01	8.56E-12	4.93E-10	0.02	4.93E-01
PUD	Stomach LYZ positive cells	3.51E-09	2.71E-09	9.76E-02	-1.89E-10	5.06E-10	6.46E-01	-6.49E-11	4.97E-10	-0.13	5.52E-01
PUD	Duodenum Paneth cells	9.79E-10	2.18E-09	3.27E-01	-2.12E-10	5.21E-10	6.58E-01	-1.48E-11	5.07E-10	-0.29	6.15E-01
PUD	Duodenum Goblet cells	-1.40E-09	2.06E-09	7.51E-01	-1.55E-10	4.59E-10	6.32E-01	-2.13E-10	4.48E-10	-0.48	6.83E-01
PUD	Duodenum ADH4 cells	3.82E-09	3.13E-09	1.11E-01	-3.37E-10	4.54E-10	7.71E-01	-2.51E-10	4.49E-10	-0.56	7.12E-01
PUD	Stomach Differentiating cells	-1.20E-09	1.87E-09	7.40E-01	-2.03E-10	4.06E-10	6.91E-01	-2.48E-10	3.97E-10	-0.63	7.34E-01
PUD	Duodenum Stem cells	3.48E-09	2.46E-09	7.89E-02	-5.34E-10	4.49E-10	8.83E-01	-4.03E-10	4.42E-10	-0.92	8.20E-01
PUD	Stomach Neck cells	-4.66E-10	2.31E-09	5.80E-01	-4.84E-10	5.23E-10	8.23E-01	-4.83E-10	5.10E-10	-0.95	8.28E-01
PUD	Stomach REG3A positive cells	1.96E-10	1.60E-09	4.51E-01	-6.48E-10	5.50E-10	8.81E-01	-5.59E-10	5.20E-10	-1.07	8.59E-01
PUD	Duodenum I cells	-1.72E-09	1.80E-09	8.31E-01	-3.61E-10	4.05E-10	8.14E-01	-4.27E-10	3.95E-10	-1.08	8.60E-01
PUD	Stomach PPP1R1B positive cells	2.23E-09	2.67E-09	2.02E-01	-8.84E-10	4.96E-10	9.63E-01	-7.80E-10	4.87E-10	-1.60	9.45E-01
PUD	Duodenum BCHE cells	5.51E-11	2.19E-09	4.90E-01	-7.34E-10	4.40E-10	9.52E-01	-7.03E-10	4.32E-10	-1.63	9.48E-01
PUD	Stomach Chief cells	-1.17E-09	2.16E-09	7.05E-01	-9.99E-10	4.51E-10	9.87E-01	-1.01E-09	4.42E-10	-2.28	9.89E-01
GU	Stomach Tuft cells	3.52E-09	1.51E-09	1.01E-02	6.10E-10	3.89E-10	5.84E-02	7.89E-10	3.76E-10	2.10	1.80E-02
GU	Duodenum K cells	5.36E-09	2.71E-09	2.41E-02	8.97E-10	5.21E-10	4.25E-02	1.06E-09	5.12E-10	2.06	1.95E-02
GU	Stomach Immune cells	2.88E-09	1.40E-09	1.94E-02	5.00E-10	3.87E-10	9.79E-02	6.70E-10	3.73E-10	1.80	3.60E-02
GU	Duodenum EC cells	6.43E-09	2.52E-09	5.39E-03	5.33E-10	3.84E-10	8.23E-02	6.67E-10	3.79E-10	1.76	3.94E-02
GU	Stomach LYZ positive cells	1.03E-09	1.75E-09	2.78E-01	6.30E-10	3.80E-10	4.88E-02	6.48E-10	3.72E-10	1.74	4.06E-02
GU	Stomach G cells	2.87E-09	1.56E-09	3.34E-02	3.75E-10	3.79E-10	1.61E-01	5.14E-10	3.69E-10	1.39	8.17E-02
GU	Duodenum Paneth cells	1.45E-09	1.49E-09	1.65E-01	3.66E-10	4.31E-10	1.98E-01	4.50E-10	4.14E-10	1.09	1.39E-01
GU	Stomach X cells	-8.12E-10	1.25E-09	7.42E-01	6.12E-10	3.95E-10	9.74E-02	3.92E-10	3.77E-10	1.04	1.49E-01
GU	Stomach Metallothionein cells	2.71E-09	2.13E-09	1.03E-01	3.07E-10	3.84E-10	2.12E-01	3.82E-10	3.78E-10	1.01	1.56E-01
GU	Duodenum Transit amplifying cells	2.05E-09	1.89E-09	1.40E-01	3.11E-10	4.63E-10	2.51E-01	4.08E-10	4.50E-10	0.91	1.82E-01
GU	Stomach Parietal cells	1.95E-09	1.72E-09	1.29E-01	2.37E-10	3.72E-10	2.62E-01	3.14E-10	3.64E-10	0.86	1.94E-01
GU	Stomach Antral ECs	3.96E-09	2.28E-09	4.13E-02	2.23E-10	3.93E-10	2.86E-01	3.31E-10	3.88E-10	0.85	1.97E-01
GU	Stomach Immature pit cells	1.85E-09	1.92E-09	1.68E-01	2.37E-10	3.98E-10	2.75E-01	3.04E-10	3.89E-10	0.78	2.18E-01
GU	Stomach D cells	3.11E-09	1.60E-09	2.63E-02	1.15E-10	3.74E-10	3.79E-01	2.69E-10	3.64E-10	0.74	2.30E-01
GU	Duodenum Immune cells	4.02E-09	1.92E-09	1.82E-02	8.94E-11	4.39E-10	4.19E-01	2.83E-10	4.28E-10	0.66	2.53E-01
GU	Stomach Prezygogenic cells	1.59E-09	1.49E-09	1.43E-01	1.36E-10	4.37E-10	3.78E-01	2.51E-10	4.19E-10	0.60	2.75E-01
GU	Stomach Oxyntic ECLs	-6.72E-10	1.36E-09	6.90E-01	2.60E-10	4.02E-10	2.59E-01	1.85E-10	3.85E-10	0.48	3.16E-01
GU	Duodenum Early immature enterocytes	3.55E-10	1.55E-09	4.09E-01	1.76E-10	4.20E-10	3.38E-01	1.88E-10	4.05E-10	0.46	3.21E-01
GU	Stomach Mature pit cells	2.44E-09	1.97E-09	1.08E-01	8.62E-11	3.91E-10	4.13E-01	1.07E-10	3.83E-10	0.46	3.24E-01
GU	Stomach Isthmus	-5.52E-11	1.66E-09	5.13E-01	1.74E-10	3.72E-10	3.20E-01	1.63E-10	3.63E-10	0.45	3.26E-01
GU	Duodenum I cells	-1.57E-09	1.31E-09	8.85E-01	3.06E-10	3.83E-10	2.12E-01	1.58E-10	3.68E-10	0.43	3.34E-01
GU	Duodenum MX cells	-1.09E-09	1.22E-09	8.14E-01	2.83E-10	4.25E-10	2.53E-01	1.34E-10	4.02E-10	0.33	3.69E-01
GU	Duodenum BCHE cells	-3.57E-10	1.54E-09	5.92E-01	1.87E-10	4.52E-10	3.40E-01	1.44E-10	4.33E-10	0.33	3.70E-01
GU	Duodenum Mature enterocytes	-2.00E-10	1.88E-09	5.42E-01	1.63E-10	4.63E-10	3.62E-01	1.42E-10	4.50E-10	0.32	3.76E-01
GU	Duodenum Goblet cells	-6.65E-10	1.52E-09	6.69E-01	1.39E-10	3.66E-10	3.52E-01	9.44E-11	3.56E-10	0.27	3.95E-01
GU	Duodenum D cells	3.63E-09	2.14E-09	4.51E-02	-7.61E-11	4.15E-10	5.73E-01	5.80E-11	4.08E-10	0.14	4.43E-01
GU	Duodenum Stem cells	1.06E-09	1.52E-09	2.43E-01	-8.86E-11	4.09E-10	5.86E-01	-1.11E-11	3.95E-10	-0.03	5.11E-01
GU	Duodenum Late immature enterocytes	3.19E-10	1.64E-09	4.23E-01	-5.22E-11	3.96E-10	5.53E-01	-3.17E-11	3.85E-10	-0.08	5.33E-01
GU	Duodenum Differentiating stem cells	2.62E-09	2.35E-09	1.32E-01	-1.97E-10	4.47E-10	6.71E-01	-9.93E-11	4.39E-10	-0.23	5.90E-01
GU	Duodenum Tuft cells	3.43E-10	1.74E-09	4.22E-01	-1.48E-10	4.22E-10	6.37E-01	-1.21E-10	4.10E-10	-0.29	6.16E-01
GU	Stomach PPP1R1B positive cells	2.00E-09	1.86E-09	1.41E-01	-2.89E-10	3.91E-10	7.70E-01	-1.93E-10	3.83E-10	-0.50	6.93E-01
GU	Duodenum ADH4 cells	3.47E-09	2.12E-09	5.10E-02	-3.01E-10	3.25E-10	8.22E-01	-2.14E-10	3.21E-10	-0.67	7.48E-01
GU	Stomach Neck cells	-4.94E-10	1.53E-09	6.27E-01	-3.22E-10	4.04E-10	7.87E-01	-3.33E-10	3.91E-10	-0.85	8.03E-01
GU	Stomach Chief cells	-1.87E-09	1.46E-09	8.99E-01	-4.82E-10	3.54E-10	9.13E-01	-5.58E-10	3.44E-10	-1.62	9.48E-01
GU	Stomach Differentiating cells	-4.17E-10	1.31E-09	6.25E-01	-5.98E-10	3.38E-10	9.61E-01	-5.86E-10	3.27E-10	-1.79	9.63E-01
GU	Stomach REG3A positive cells	2.51E-10	1.24E-09	4.20E-01	-7.00E-10	3.60E-10	9.74E-01	-6.23E-10	3.46E-10	-1.81	9.65E-01
DU	Stomach Antral ECs	2.46E-09	3.24E-09	2.24E-01	1.08E-09	4.77E-10	1.21E-02	1.11E-09	4.72E-10	2.34	9.62E-03
DU	Duodenum K cells	7.33E-09	3.37E-09	1.47E-02	1.05E-09	5.25E-10	2.29E-02	1.20E-09	5.19E-10	2.31	1.05E-02
DU	Stomach G cells	4.28E-09	2.30E-09	3.12E-02	7.23E-10	4.16E-10	4.08E-02	8.36E-10	4.09E-10	2.04	2.04E-02
DU	Duodenum EC cells	5.49E-09	3.05E-09	3.57E-02	8.50E-10	4.97E-10	4.36E-02	9.71E-10	4.91E-10	1.98	2.39E-02
DU	Stomach D cells	3.35E-09	2.21E-09	6.46E-02	6.34E-10	4.48E-10	7.86E-02	7.42E-10	4.39E-10	1.69	4.57E-02
DU	Duodenum D cells	1.30E-09	2.42E-09	2.96E-01	5.70E-10	4.58E-10	1.07E-01	5.96E-10	4.50E-10	1.32	9.30E-02
DU	Stomach Metallothionein cells	6.09E-09	2.58E-09	9.13E-03	3.99E-10	4.35E-10	1.80E-01	5.56E-10	4.29E-10	1.30	9.74E-02
DU	Stomach Oxyntic ECLs	-7.35E-11	1.99E-09	5.15E-01	4.95E-10	4.02E-10	1.09E-01	4.73E-10	3.94E-10	1.20	1.15E-01
DU	Stomach Parietal cells	4.07E-09	2.03E-09	2.24E-02	2.57E-10	4.14E-10	2.67E-01	4.10E-10	4.06		

Table 34. Cell type specificity analysis using MAGMA.

Meta-analysis was performed using inverse variance weighted (IVW) methods. P-value was derived from upper-tail tests.

Phenotype	Cell Type	EAS			EUR			Meta-analysis			
		BETA_EAS	SE_EAS	P_EAS	BETA_EUR	SE_EUR	P_EUR	BETA_FE	SE_META	Z_META	P_META (one-sided)
PUD	Duodenum EC cells	0.085	0.031	2.78E-03	0.054	0.027	2.33E-02	0.068	0.020	3.329	4.36E-04
PUD	Stomach D cells	0.050	0.029	4.13E-02	0.062	0.026	8.22E-03	0.057	0.019	2.945	1.61E-03
PUD	Stomach Tuft cells	0.072	0.028	5.54E-03	0.038	0.026	6.87E-02	0.053	0.019	2.807	2.50E-03
PUD	Stomach Antral ECs	0.066	0.029	1.24E-02	0.038	0.026	6.98E-02	0.050	0.019	2.591	4.79E-03
PUD	Duodenum Late immature enterocytes	0.062	0.030	1.96E-02	0.032	0.027	1.20E-01	0.045	0.020	2.251	1.22E-02
PUD	Duodenum K cells	0.050	0.031	5.47E-02	0.037	0.027	8.66E-02	0.043	0.020	2.080	1.88E-02
PUD	Duodenum Early immature enterocytes	0.036	0.030	1.15E-01	0.044	0.027	5.40E-02	0.041	0.020	1.999	2.28E-02
PUD	Duodenum Transit amplifying cells	0.038	0.030	1.05E-01	0.036	0.027	9.12E-02	0.037	0.020	1.830	3.36E-02
PUD	Stomach Mature pit cells	0.068	0.029	9.36E-03	0.003	0.026	4.47E-01	0.032	0.019	1.663	4.81E-02
PUD	Stomach Prezygogenic cells	0.032	0.029	1.32E-01	0.026	0.026	1.57E-01	0.029	0.019	1.496	6.73E-02
PUD	Stomach Metallothionein cells	0.024	0.029	2.07E-01	0.031	0.025	1.13E-01	0.028	0.019	1.452	7.33E-02
PUD	Stomach X cells	0.020	0.030	2.48E-01	0.034	0.026	9.51E-02	0.028	0.019	1.435	7.56E-02
PUD	Stomach Immature pit cells	0.055	0.029	3.12E-02	0.006	0.025	4.05E-01	0.027	0.019	1.400	8.07E-02
PUD	Stomach LYZ positive cells	0.030	0.029	1.49E-01	0.010	0.026	3.53E-01	0.019	0.019	0.972	1.65E-01
PUD	Stomach Isthmus	0.029	0.029	1.57E-01	0.007	0.025	3.94E-01	0.017	0.019	0.868	1.93E-01
PUD	Duodenum Paneth cells	0.068	0.030	1.15E-02	-0.022	0.026	8.04E-01	0.017	0.020	0.849	1.98E-01
PUD	Stomach Immune cells	0.004	0.029	4.49E-01	0.026	0.026	1.60E-01	0.016	0.020	0.827	2.04E-01
PUD	Stomach G cells	-0.002	0.029	5.30E-01	0.027	0.026	1.51E-01	0.014	0.019	0.722	2.35E-01
PUD	Stomach Oxyntic ECLs	-0.003	0.029	5.46E-01	0.027	0.026	1.50E-01	0.014	0.019	0.707	2.40E-01
PUD	Duodenum MX cells	-0.006	0.030	5.79E-01	0.024	0.027	1.87E-01	0.011	0.020	0.535	2.96E-01
PUD	Stomach Parietal cells	0.022	0.029	2.22E-01	-0.001	0.026	5.15E-01	0.009	0.019	0.481	3.15E-01
PUD	Duodenum Tuft cells	0.050	0.031	5.33E-02	-0.022	0.027	7.97E-01	0.009	0.020	0.428	3.34E-01
PUD	Stomach Neck cells	0.028	0.029	1.67E-01	-0.008	0.025	6.22E-01	0.008	0.019	0.402	3.44E-01
PUD	Stomach REG3A positive cells	0.020	0.029	2.49E-01	-0.002	0.025	5.24E-01	0.008	0.019	0.400	3.45E-01
PUD	Duodenum I cells	-0.005	0.030	5.72E-01	0.017	0.027	2.60E-01	0.007	0.020	0.358	3.60E-01
PUD	Duodenum Differentiating stem cells	0.018	0.031	2.78E-01	-0.005	0.027	5.75E-01	0.005	0.020	0.242	4.04E-01
PUD	Duodenum Mature enterocytes	0.037	0.031	1.14E-01	-0.027	0.027	8.35E-01	0.002	0.020	0.076	4.70E-01
PUD	Duodenum BChE cells	0.027	0.031	1.87E-01	-0.020	0.027	7.74E-01	0.000	0.020	0.016	4.94E-01
PUD	Duodenum Stem cells	0.023	0.031	2.27E-01	-0.018	0.028	7.47E-01	0.000	0.021	0.003	4.99E-01
PUD	Duodenum Goblet cells	-0.029	0.031	8.18E-01	0.021	0.027	2.22E-01	0.000	0.021	-0.012	5.05E-01
PUD	Duodenum Immune cells	-0.007	0.031	5.92E-01	0.003	0.027	4.62E-01	-0.002	0.020	-0.081	5.32E-01
PUD	Stomach Differentiating cells	0.017	0.029	2.80E-01	-0.015	0.025	7.30E-01	-0.002	0.019	-0.083	5.33E-01
PUD	Stomach PPP1R1B positive cells	0.046	0.029	5.49E-02	-0.040	0.026	9.41E-01	-0.002	0.019	-0.099	5.39E-01
PUD	Duodenum D cells	-0.007	0.030	5.91E-01	-0.024	0.027	8.16E-01	-0.017	0.020	-0.828	7.96E-01
PUD	Stomach Chief cells	0.010	0.030	3.70E-01	-0.043	0.026	9.53E-01	-0.020	0.019	-1.055	8.54E-01
PUD	Duodenum ADH4 cells	-0.002	0.030	5.32E-01	-0.041	0.027	9.34E-01	-0.024	0.020	-1.175	8.80E-01
GU	Stomach Tuft cells	0.058	0.028	1.76E-02	0.032	0.024	9.44E-02	0.043	0.018	2.377	8.73E-03
GU	Duodenum K cells	0.047	0.030	6.04E-02	0.039	0.026	6.48E-02	0.042	0.020	2.159	1.54E-02
GU	Duodenum EC cells	0.048	0.030	5.36E-02	0.027	0.026	1.45E-01	0.036	0.019	1.854	3.19E-02
GU	Stomach Immune cells	0.028	0.029	1.66E-01	0.032	0.025	9.44E-02	0.031	0.019	1.628	5.17E-02
GU	Stomach Mature pit cells	0.021	0.028	2.32E-01	0.024	0.024	1.66E-01	0.022	0.018	1.215	1.12E-01
GU	Duodenum Transit amplifying cells	0.021	0.030	2.40E-01	0.024	0.026	1.70E-01	0.023	0.019	1.184	1.18E-01
GU	Stomach PPP1R1B positive cells	0.022	0.028	2.17E-01	0.013	0.024	2.90E-01	0.017	0.018	0.929	1.76E-01
GU	Duodenum Paneth cells	0.051	0.029	3.98E-02	-0.012	0.026	6.83E-01	0.015	0.019	0.795	2.13E-01
GU	Stomach Prezygogenic cells	0.002	0.028	4.73E-01	0.021	0.024	1.90E-01	0.013	0.018	0.705	2.40E-01
GU	Duodenum Late immature enterocytes	0.037	0.029	1.06E-01	-0.005	0.026	5.79E-01	0.013	0.019	0.676	2.50E-01
GU	Duodenum Immune cells	0.004	0.030	4.45E-01	0.020	0.026	2.25E-01	0.013	0.020	0.666	2.53E-01
GU	Duodenum D cells	0.040	0.030	9.09E-02	-0.007	0.025	6.09E-01	0.012	0.019	0.651	2.58E-01
GU	Stomach Metallothionein cells	-0.044	0.029	9.34E-01	0.050	0.024	1.89E-02	0.012	0.018	0.632	2.64E-01
GU	Stomach X cells	0.013	0.029	3.26E-01	0.011	0.025	3.33E-01	0.012	0.019	0.620	2.68E-01
GU	Stomach LYZ positive cells	-0.005	0.028	5.63E-01	0.023	0.025	1.82E-01	0.011	0.019	0.576	2.82E-01
GU	Duodenum Goblet cells	0.008	0.031	3.92E-01	0.005	0.026	4.21E-01	0.006	0.020	0.328	3.71E-01
GU	Stomach Immature pit cells	0.017	0.029	2.81E-01	-0.004	0.024	5.65E-01	0.005	0.018	0.249	4.02E-01
GU	Duodenum Early immature enterocytes	0.001	0.030	4.81E-01	0.007	0.026	3.94E-01	0.005	0.019	0.233	4.08E-01
GU	Stomach Parietal cells	-0.005	0.028	5.68E-01	0.011	0.025	3.30E-01	0.004	0.018	0.217	4.14E-01
GU	Stomach D cells	0.029	0.028	1.52E-01	-0.016	0.025	7.38E-01	0.004	0.019	0.197	4.22E-01
GU	Stomach REG3A positive cells	0.016	0.028	2.88E-01	-0.006	0.024	5.91E-01	0.003	0.018	0.189	4.25E-01
GU	Duodenum Differentiating stem cells	0.000	0.030	5.04E-01	0.005	0.026	4.20E-01	0.003	0.019	0.148	4.41E-01
GU	Duodenum Stem cells	-0.016	0.031	6.94E-01	0.017	0.026	2.66E-01	0.003	0.020	0.142	4.44E-01
GU	Stomach Neck cells	0.023	0.028	2.03E-01	-0.017	0.024	7.63E-01	0.000	0.018	0.005	4.98E-01
GU	Duodenum Mature enterocytes	0.040	0.030	9.04E-02	-0.030	0.026	8.78E-01	0.000	0.019	-0.011	5.05E-01
GU	Stomach Isthmus	0.012	0.028	3.38E-01	-0.011	0.024	6.82E-01	-0.002	0.018	-0.087	5.35E-01
GU	Duodenum BChE cells	0.002	0.030	4.73E-01	-0.008	0.026	6.20E-01	-0.004	0.019	-0.190	5.76E-01
GU	Stomach Differentiating cells	0.021	0.028	2.30E-01	-0.024	0.024	8.40E-01	-0.005	0.018	-0.273	6.07E-01
GU	Duodenum Tuft cells	0.025	0.030	2.09E-01	-0.028	0.025	8.66E-01	-0.007	0.019	-0.337	6.32E-01
GU	Stomach Oxyntic ECLs	-0.028	0.029	8.37E-01	0.007	0.025	3.84E-01	-0.008	0.019	-0.418	6.62E-01
GU	Stomach Antral ECs	0.025	0.029	1.91E-01	-0.033	0.024	9.13E-01	-0.009	0.019	-0.469	6.80E-01
GU	Stomach Chief cells	-0.028	0.029	8.38E-01	-0.002	0.024	5.29E-01	-0.013	0.019	-0.694	7.56E-01
GU	Duodenum ADH4 cells	0.007	0.029	4.09E-01	-0.037	0.026	9.27E-01	-0.018	0.019	-0.947	8.28E-01
GU	Duodenum MX cells	-0.023	0.030	7.84E-01	-0.016	0.025	7.40E-01	-0.019	0.019	-1.000	8.41E-01
GU	Stomach G cells	-0.021	0.028	7.76E-01	-0.040	0.024	9.48E-01	-0.032	0.018	-1.726	9.58E-01
GU	Duodenum I cells	-0.043	0.030	9.26E-01	-0.029	0.026	8.72E-01	-0.035	0.019	-1.804	9.64E-01
DU	Duodenum Early immature enterocytes	0.015	0.030	3.12E-01	0.070	0.027	4.15E-03	0.046	0.020	2.300	1.07E-02
DU	Stomach Antral ECs	0.007	0.029	4.03E-01	0.071	0.025	2.49E-03	0.043	0.019	2.273	1.15E-02
DU	Stomach Prezygogenic cells	0.049	0.028	3.99E-02	0.031	0.025	1.08E-01	0.039	0.019	2.089	1.83E-02
DU	Duodenum Late immature enterocytes	0.022	0.030	2.29E-01	0.057	0.027	1.69E-02	0.041	0.020	2.077	1.89E-02
DU	Stomach Tuft cells	0.043	0.028	6.28E-02	0.033	0.025	9.33E-02	0.037	0.019	2.007	2.24E-02
DU	Stomach Differentiating cells	0.021	0.028	2.23E-01	0.049	0.025	2.43E-02	0.037	0.019	1.980	2.39E-02
DU	Stomach Metallothionein cells	0.024	0.029	2.02E-01	0.041	0.025	4.89E-02	0.034	0.019	1.804	3.56E-02
DU	Stomach Mature pit cells	0.037	0.029	9.50E-02	0.030	0.025	1.13E-01	0.033	0.019	1.774	3.80E-02
DU	Stomach LYZ positive cells	0.042	0.029	7.28E-02	0.027	0.026	1.46E-01	0.034	0.019	1.754	3.97E-02
DU	Stomach REG3A positive cells	0.029	0.028	1.49E-01	0.034	0.025	8.54E-02	0.032	0.019	1.716	4.31E-02
DU	Stomach Oxyntic ECLs	0.029	0.029	1.56E-01	0.030	0.026	1.21E-01	0.030	0.019	1.547	6.09E-02
DU	Stomach Chief cells	0.069	0.029	9.10E-03	-0.008	0.025	6.26E-01	0.025	0.019	1.303	9.63E-02
DU	Duodenum Transit amplifying cells	0.007	0.030	4.06E-01	0.038	0.026	7.42E-02	0.025	0.020	1.246	1.06E-01
DU	Stomach G cells	0.001	0.029	4.87E-01	0.037	0.025	6.74E-02	0.022	0.019	1.147	1.26E-01
DU	Duodenum K cells	0.016	0.031	3.05E-01	0.027	0.027	1.56E-01	0.022	0.020	1.098	1.36E-01
DU	Stomach Parietal cells	0.026	0.029	1.83E-01	0.015	0.025	2.72E-01	0.020	0.019	1.053	1.46E-01
DU	Stomach D cells	0.030	0.029	1.42E-01	0.007	0.025	3.87E-01	0.017	0.019	0.922	1.78E-01
DU	Duodenum BChE cells	0.021	0.031	2.43E-01	0.014	0.026	3.00E-01	0.017	0.020	0.851	1.97E-01
DU	Duodenum Stem cells	0.030	0.031	1.60E-01	0.005	0.027	4.30E-01	0.016	0.020	0.791	2.15E-01
DU	Stomach Isthmus	0.006	0.029	4							

3. 13 Sensitivity analyses

In this section, association results of the previous chapter were further evaluated for their robustness by a range of sensitivity analyses. Potential factors that could affect the association results were evaluated, including sample overlap across BBJ cohorts, the heterogeneity across cohorts in East-Asian meta-analysis, the reference genome version used for imputation, statistical power for HP-stratification analysis, and the window size selected for gene- and gene-set-level analyses.

Sample overlap among Biobank Japan cohorts

Sample overlap among the three BBJ datasets was checked by merging the genotype datasets and estimating the identical-by-descent (IBD) sharing. The potential sample overlap was estimated to be minimal (<0.06%; Figure 38)

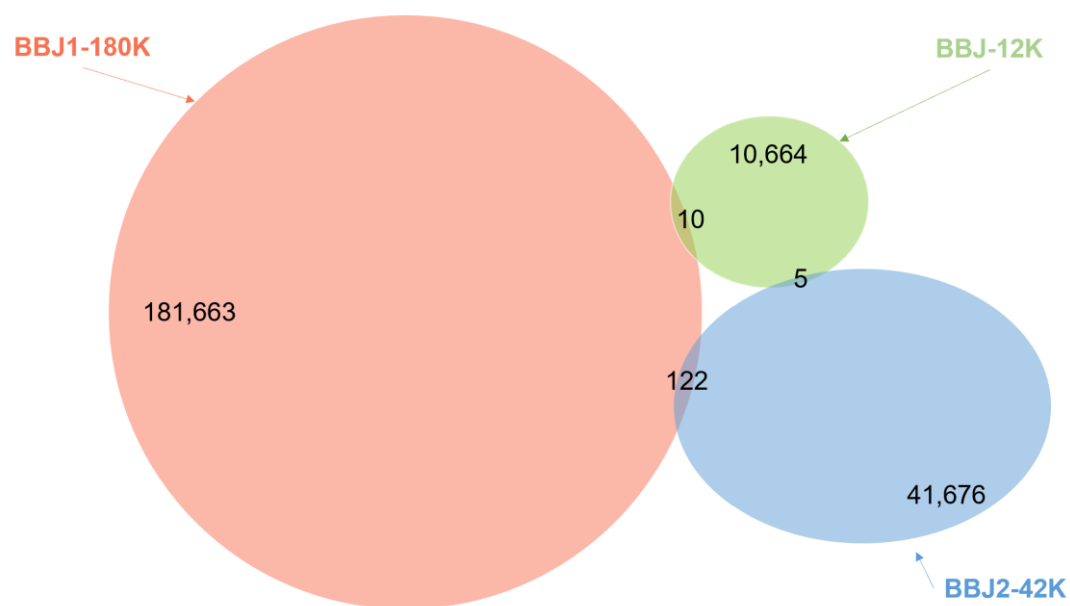


Figure 38. Venn Plot of the potential sample overlap within Biobank Japan cohorts.

Sample overlap was estimated by IBD sharing.

Heterogeneity analysis for East Asian-specific meta-analysis

The random-effects model implemented in GWAMA⁹³ was employed to further evaluate the heterogeneity (Figure 39).

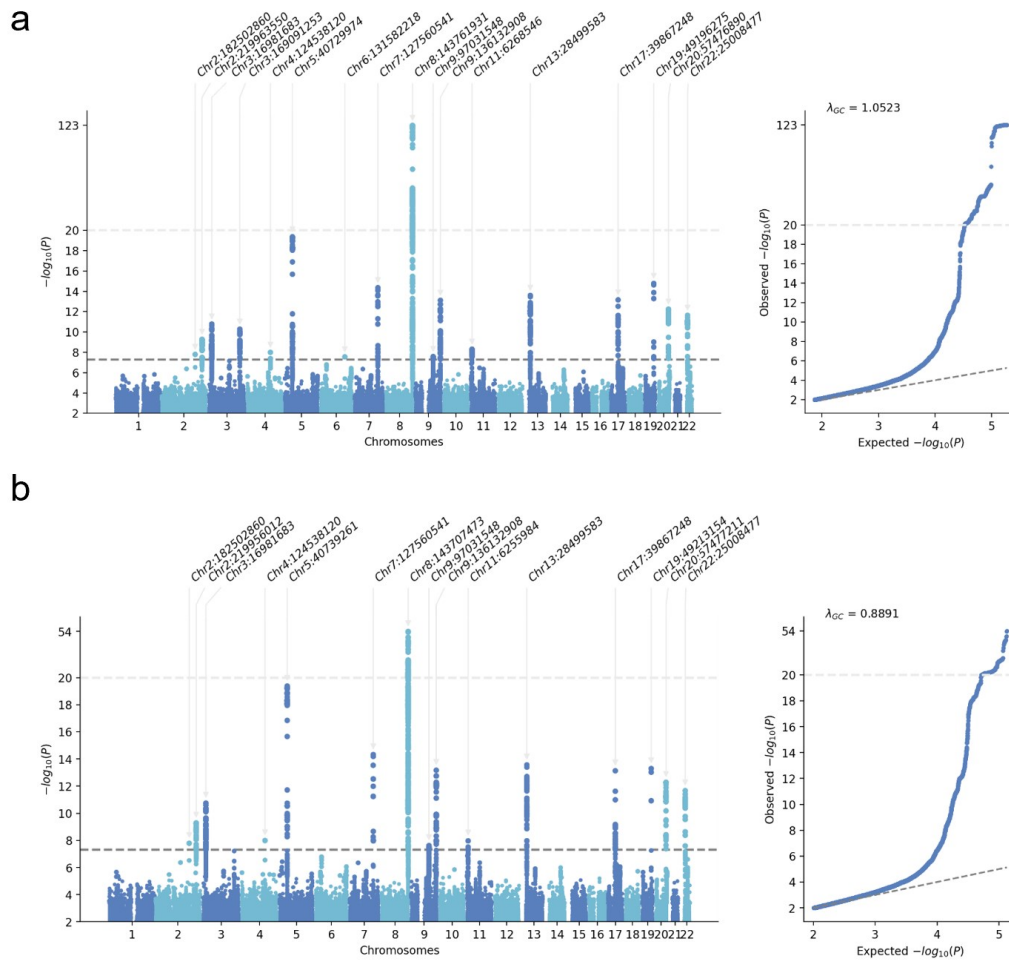


Figure 39. Manhattan plots and Q-Q plots for EAS-specific meta-analysis of PUD using different models.

For variants above the top light grey dashed line ($-\log_{10}(P) > 20$), values are rescaled. Lead variants are annotated with the nearest gene name. Variants are plotted against GRCh37 (hg19). The bottom dark grey dashed line indicates the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$). Variants with $-\log_{10}(P) < 2$ are omitted. **a**, Manhattan plot and Q-Q plot for the EAS-specific fixed-effect meta-analysis of PUD using METAL. **b**, Manhattan plot and Q-Q plot for the EAS-specific random-effects meta-analysis of PUD using GWAMA.

Investigation for signals in GRCh38-specific regions

Since the main study was based on GRCh37, to further investigate potential associations in GRCh38-specific regions, a cross-ancestry meta-analysis combining GRCh38-based datasets was conducted (Figure 40; Figure 41).

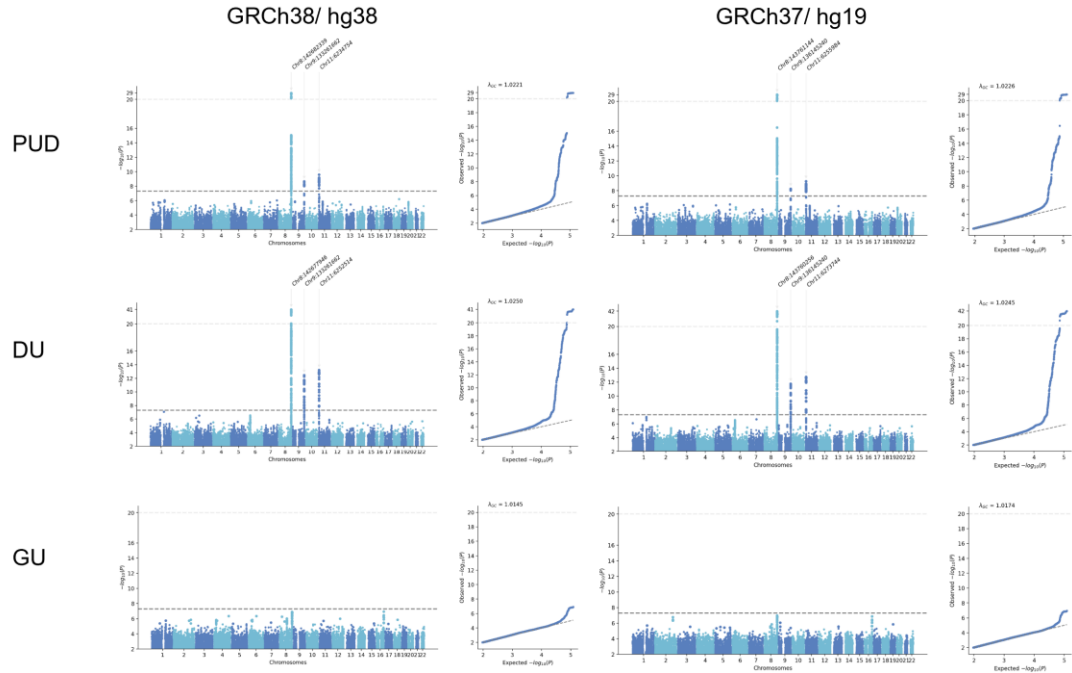


Figure 40. Manhattan plots and Q-Q plots for cross-ancestry meta-analyses based on different versions of the reference genome.

GWASs of PUD, DU, and GU conducted in BBJ1-12K, BBJ2-42K, and FinnGen were meta-analyzed (Methods). Variants existing in at least two cohorts are shown. For variants above the top light grey dashed line ($-\log_{10}(P) > 20$), values are rescaled. The bottom dark grey dashed line indicates the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$). Left panels, GRCh38-based datasets were used; variants are plotted against GRCh38 (hg38). Right panels, GRCh37-based datasets were used; variants are plotted against GRCh37 (hg19).

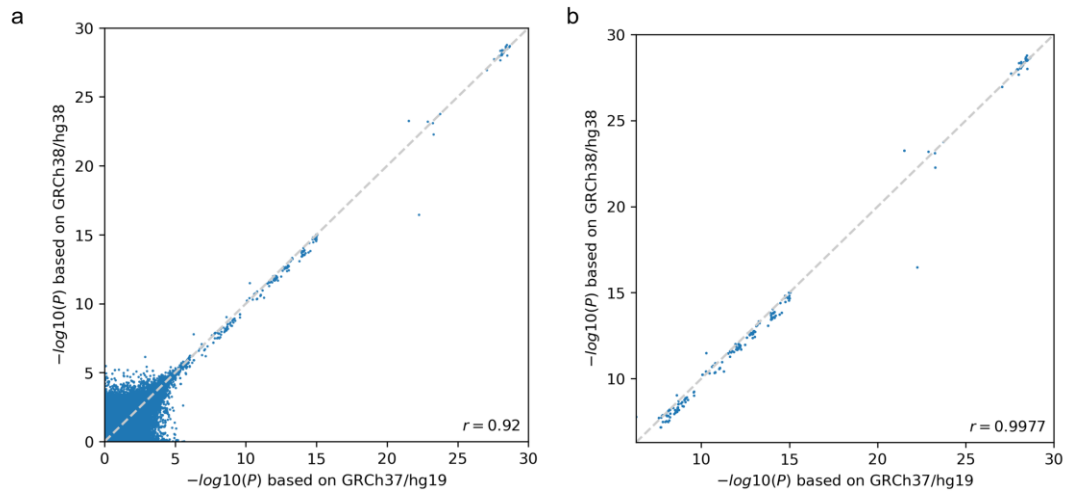


Figure 41. Comparison of $-\log_{10}(P)$ for SNPs in GWASs based on different versions of the reference genome.

a, comparison of all shared SNPs between GRCh37-based datasets and GRCh38-based datasets. The grey dashed line, a 45-degree line. Pearson correlation coefficient r is shown in the bottom right corner. **b**, comparison of SNPs reaching the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$).

Power analysis for HP-stratified analysis

Since stratified analysis could reduce power and lead to false negative results, the power of GWAS of HP-negative PUD for the identified HP-positive-specific variant was estimated. (Figure 42).

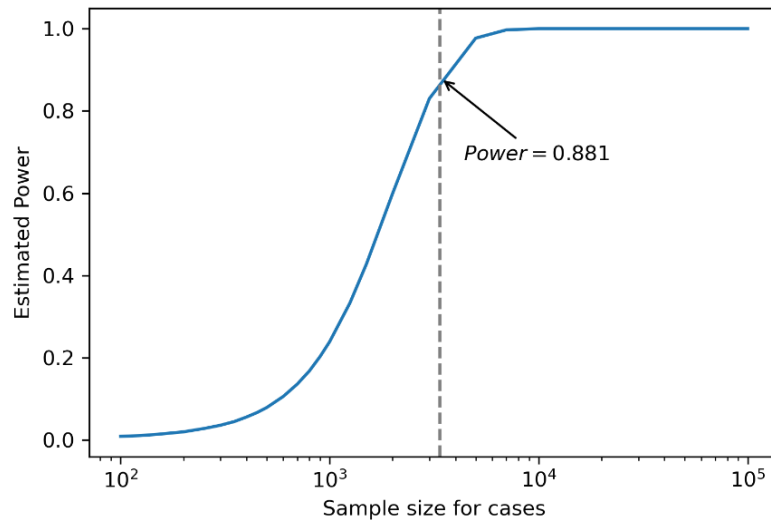


Figure 42. Power analysis for GWAS of HP-negative PUD in TMM-50K.

Statistical power was estimated using GAS Power Calculator (https://csg.sph.umich.edu/abecasis/gas_power_calculator/). The dashed line represents the number of cases ($N_{\text{case, HP-negative}} = 3,372$) used for GWAS of PUD in HP-negative individuals from TMM-50K. Other settings: the number of controls, 26,432; significance level, 0.0016; disease model, additive; prevalence, 0.109; disease allele frequency, 0.093; genotype relative risk, 1.189.

Analysis on the effect of window size selection for MAGMA

Pair-wise comparisons of the $-\log_{10}(p)$ values generated by MAGMA using a range of window sizes were conducted to evaluate the robustness of the association results in gene-based, tissue-type specificity, and cell-type specificity analyses discovered by MAGMA; the effect of window size selection was marginal (Figure 43 - Figure 45).

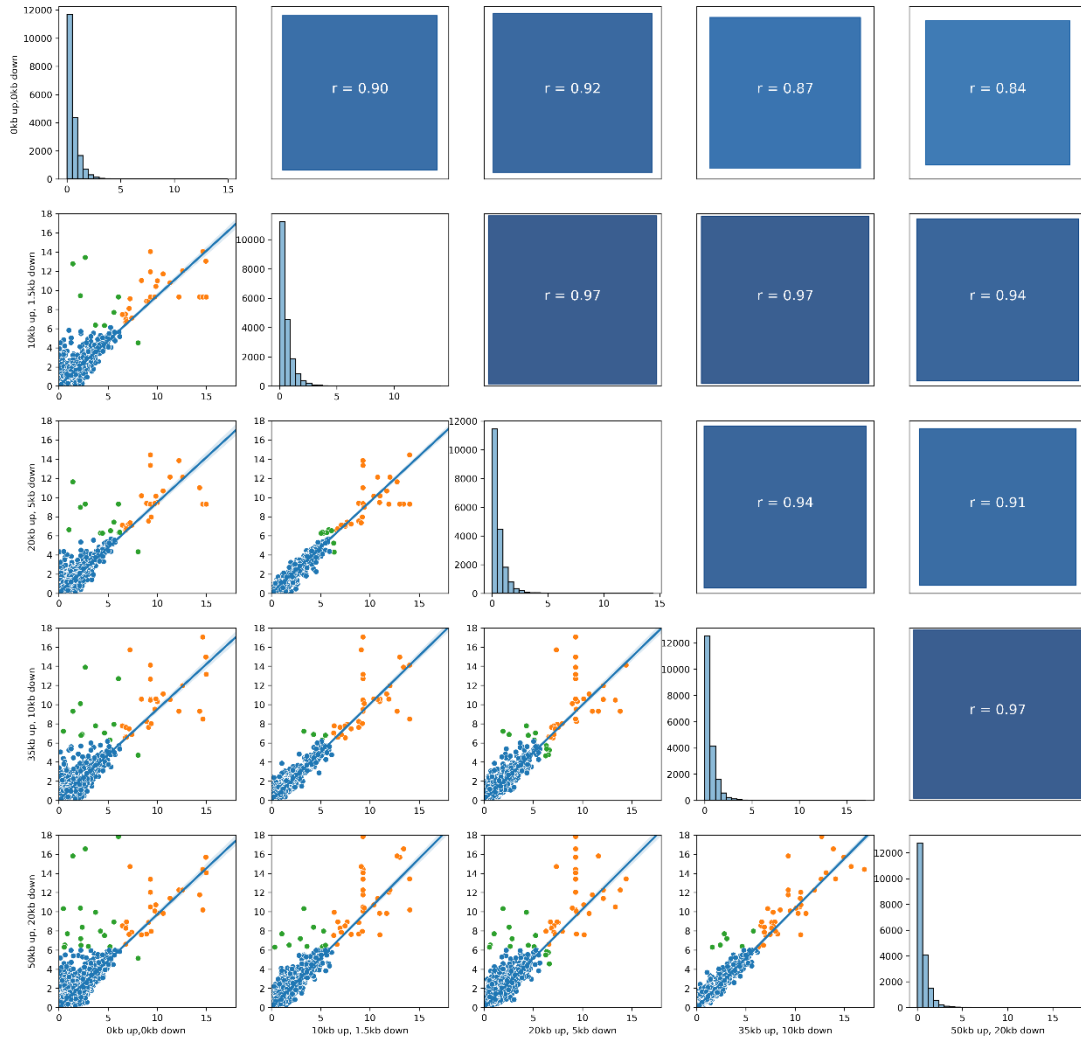


Figure 43. Comparison of MAGMA results using different window sizes around genes for gene-based analysis.

Comparison of $-\log_{10}(P)$ values obtained by gene-based tests for PUD in EAS using different window sizes for MAGMA (implemented in FUMA). Blue marker, P value after Bonferroni correction > 0.05 for both window sizes; green marker, P value after Bonferroni correction < 0.05 for only one of the two window sizes; yellow marker, P value after Bonferroni correction < 0.05 for both window sizes. Blue line, the linear regression line with confidence interval. The sizes of squares in the upper right are proportional to the Pearson correlation coefficient r .

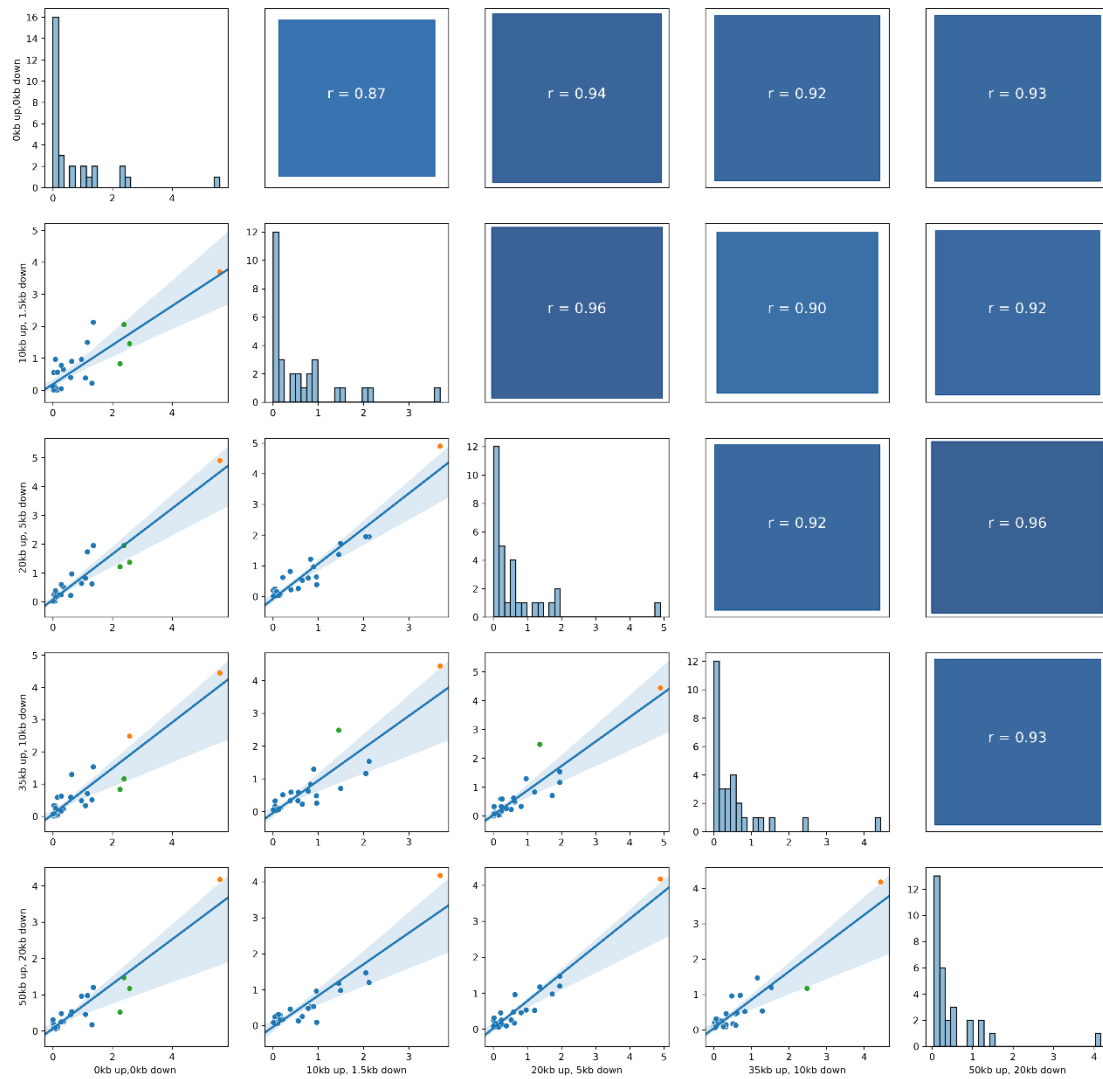


Figure 44. Comparison of MAGMA results using different window sizes around genes for tissue-specificity analysis.

Comparison of $-\log_{10}(P)$ values obtained by tissue-specificity analysis for PUD in EAS using different window sizes for MAGMA (implemented in FUMA). Blue marker, $\text{FDR} > 5\%$ for both window sizes; green marker, $\text{FDR} < 5\%$ for only one of the two window sizes; yellow marker, $\text{FDR} < 5\%$ for both window sizes. Blue line, the linear regression line with confidence interval. The sizes of squares in the upper right are proportional to the Pearson correlation coefficient r .

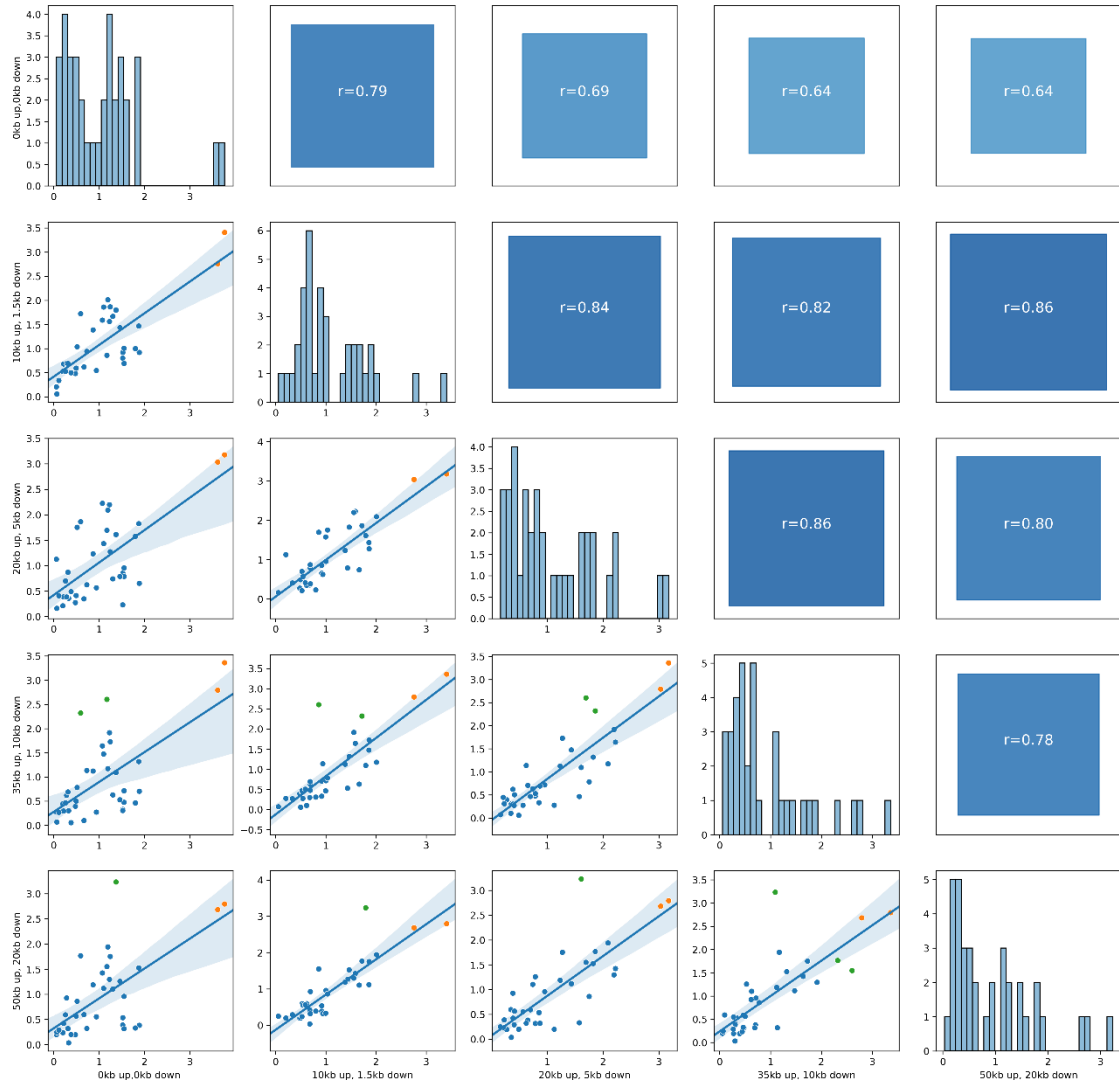


Figure 45. Comparison of MAGMA results using different window sizes around genes for cell-type-specificity analysis.

Comparison of meta-analyzed $-\log_{10}(P)$ values obtained by cell-type-specificity analysis for PUD using different window sizes. Blue marker, $\text{FDR} > 5\%$ for both window sizes; green marker, $\text{FDR} < 5\%$ for only one of the two window sizes; yellow marker, $\text{FDR} < 5\%$ for both window sizes. Blue line, the linear regression line with confidence interval. The sizes of squares in the upper right are proportional to the Pearson correlation coefficient r .

Chapter IV - Discussion

4. 1 Genetic architecture of PUD across ancestries

In this study, GWAS meta-analyses of PUD and its subtypes discovered 33 autosomal susceptibility loci, of which 25 had not been reported in previous GWAS (19 in East Asian-specific analysis and 6 in cross-ancestry analysis). The loci were mostly shared across ancestries with strong correlations of effect sizes. The cross-ancestry analysis emphasized the high genetic correlation of DU and suggested the heterogeneity of GU across ancestries. The larger effect sizes of *MUC1* and *MUC6* in populations of EUR ancestry are in reasonable agreement with their critical roles in the protection from NSAID-induced injury, given the much lower prevalence of HP infection in Western countries compared to that in East Asian populations²; despite the high genetic correlation of the available variants, we also identified variants that are only common in East-Asian population such as rs2233580 in *PAX4*; the cross-ancestry genetic correlation suggested that the genetic architecture of GU might be different between East Asians and Europeans based on ρ_{gi} which is estimated taking both effect sizes and allele frequencies into account. These population-specific risk variants and the genetic architecture of GU may potentially contribute to the difference in PUD prevalence between East Asians and Europeans in addition to environmental factors.

4. 2 Roles of cell differentiation and hormone regulation in PUD etiology

In the association analyses, multiple novel loci (*PAX4*⁹⁴, *PDX1*⁹⁵, *IHH*⁹⁶, and *SLC22A3*⁹⁷) and reported loci (*CCKBR*, *CDX2*, and *GAST*) were found to be related to cell differentiation and gastrin signaling. By integrating scRNA-seq datasets, the association of PUD with certain hormone-secreting cells was identified, including stomach D cells (somatostatin), and stomach antral and duodenal EC cells (serotonin). Gastrointestinal D cells are estimated to secrete ~65% circulating somatostatin, suppressing the release of gastric hormones and gastric acid^{98,99}. The PUD-associated rs2233580 in *PAX4* is a missense variant predicted to be highly deleterious (CADD > 20). It has been shown that *PAX4* is a transcriptional repressor for somatostatin¹⁰⁰ and regulates duodenal hormone-secreting cells and serotonin/somatostatin-producing cells of the distal stomach^{101,102}. Significant associations of *PDX1* were also detected in gene-based tests with independent signals (rs139276646) in its regulatory region. *PDX1* activates somatostatin transcription by interacting with its promoter. These suggested that D cell/somatostatin dysregulation may contribute to PUD development. On the other hand, EC cells are the predominant source of body serotonin and play key roles in the gut-brain axis as chemosensors¹⁰³, affecting a wide range of physiological processes, including gastrointestinal motility and secretion, nausea, and visceral hypersensitivity. Psychological conditions, including stress and depression, were associated with a higher risk of PUD^{9,104}. A previous large-scale study identified the causal effect of major depression (MD) on PUD¹². Given the important role of serotonin in psychological conditions, the association of EC cells/serotonin with PUD, and the bidirectional effects of the brain-gut axis, it is worth further investigating whether serotonin might be a key factor in the link between depression and PUD. Given these variant-level and cell-type-level findings, proteins such as *PAX4* and *PDX1*, and gastrointestinal hormones including somatostatin and serotonin, could potentially serve as the targets for drug development or drug repurposing for PUD prevention or treatment.

The HP-stratified analysis also showed the signal at *CCKBR* (receptor for gastrin) to be HP-positive-specific. The PUD risk allele of the lead SNP (rs12792379) is in LD with the eQTL allele associated with higher *CCKBR* expression in multiple tissues⁷⁵, including esophagus mucosa. It has been widely shown that HP-elicited cytokines stimulate gastrin release¹⁰⁵. It is likely that the increased gastrin level induced by HP will interact with altered expression in *CCKBR*, leading to dysregulated gastric acid secretion and altered susceptibility to apoptosis^{106,107}.

Taken together, our results provided genetic evidence of gastrointestinal cell differentiation and hormone regulation being critical in PUD etiology.

4.3 The similarities and differences between GU and DU

As expected, high genetic correlation between GU and DU and nominally significant genetic correlations between PUD and its risk factors were observed. It has long been known cigarette smoking increases the risk for PUD⁷. In the genetic correlation analysis, a nominally significant ($P < 0.05$) negative correlation was detected between PUD and age at smoking initiation, but such correlation was not detected for PUD and cigarettes per day, which suggest that PUD might share genetic components with long-term smoking behavior. Significant genetic correlations were not detected between PUD and drinking-related traits. The PUD risk allele for rs3859862 at *GGT1* (gamma-glutamyl transferase 1) identified in this study was linked with the alleles of a cis-eQTL and a cis-pQTL for *GGT1* that decrease the expression of *GGT1* in stomach and protein level of GGT in serum. It has been shown that alcohol consumption and smoking could lead to an increased level of GGT^{108,109}. It is worth investigating if GGT plays a role in the association between PUD and smoking/drinking. Additionally, nominally significant genetic correlations of PUD/GU with chronic obstructive pulmonary disease (COPD), asthma, and rheumatoid arthritis (RA) in EAS were detected (Supplementary Figure 12). A previous large-scale meta-analysis of asthma¹¹⁰ identified the significant genetic correlation between asthma and PUD in addition to the significant genetic correlation of asthma with COPD or RA. The consistent observations suggested a genetic link between PUD and immune-related diseases, which has not been well studied yet and warranted further investigations.

Effect size comparisons demonstrated that GU shared risk loci with DU, but had smaller effect sizes than DU. The polygenicity of GU was higher than that of DU. The SNP-based heritability estimate for DU (liability-scaled) was almost twice as high as for GU. Additionally, DU PRS showed a stronger association with GU than GU PRS in East Asians. The results revealed the genetic difference between GU and DU and reflected a higher heterogeneity of GU^{111–113}.

4.4 The correlation between DU and GC

Three variants (linked to *EFNA1*, *PTGER4*, and *PSCA*) were observed to have relatively large pleiotropic effects on DU and GC. *EFNA1* suppresses tumor growth while *PGE2* supports tumor growth by promoting angiogenesis^{84,114}. PUD risk alleles resulted in increased levels of *EFNA1* and reduced levels of *PTGER4*, whereas GC risk alleles were associated with a decreased level of *EFNA1* and an increased level of *PTGER4*. This suggested that the risk alleles of *EFNA1* and *PTGER4* for GC (non-risk alleles for PUD) potentially benefited peptic ulcer healing while imposing an increased risk for GC through up-regulated cell proliferation and angiogenesis. Additionally, multiple PUD-risk cancer-related genes were detected for the risk of PUD and its subtypes (e.g., *IHH*, *GNAS*, *NHEJ1*, *JUP*, and *MECOM*), which

provided potential targets contributing to the different outcomes of PUD or GC. No causal effects were identified in the outlier-corrected MR in this study, which may suffer from the power loss caused by sample split and removal of variants. Further research is warranted on the protective role of DU against GC.

4. 5 Limitations of this study

Although we identified multiple associations, the current study has several potential limitations. First, the phenotypic information of PUD and subtypes was obtained via interviews and reviews of medical records. However, the prevalence rate of PUD was consistent with that in previous epidemiological studies; our study replicated most of the previously identified loci, and the novel biological findings are feasible, which suggested the relatively high reliability of the results. Second, due to the lack of information regarding the chronological order of PUD onset and anti-inflammatory drug use at PUD onset in this study, the specific interaction of NSAIDs with host genetic factors was under-explored. Third, detailed information on the anatomic site of the ulcers or the strains of HP was not available, and the phenotyping of HP infection status based on anti-HP IgG level might be affected by factors including the stage of *H.pylori* infection since anti-HP IgG could disappear in around 25% of patients with advanced atrophic corpus gastritis¹¹⁵, which may lead to false negative infection status; the HP infection prevalence estimated by anti-HP IgG level in TMM-50K (39.1%) was notably lower than previous epidemiological observations in Japanese populations² (44.7%–58.7%); taking the HP infection history into account could potentially improve the analysis. Fourth, the variants identified by fine-mapping and the overlaps in association signals identified by lookup approaches might result from tagging distinct causal variants, and the meta-analysis fine-mapping using BBJ1-180K as LD reference might be miscalibrated¹¹⁶, which should be interpreted cautiously. Despite the biobank-scale cohort for HP-stratified analysis, the statistical power is still limited for certain analyses. Additionally, despite significant signals identified in the MHC region, we did not further investigate the signals due to the high complexity of the MHC region; since HP infection is one of the most important risk factors for peptic ulcer diseases, further investigation is warranted for the associations identified in the MHC locus. Even though our subtype analysis revealed the overall similarities and differences in genetic architecture, a large sample size, and more detailed classifications are still warranted to elucidate the potential heterogeneity further.

4. 6 Summary and prospects

In summary, the current study quadrupled the number of risk loci for PUD and its subtypes and improved our understanding of the genetic architecture of PUD. The findings provided insight into the biological pathways involved in PUD pathogenesis and potential links between PUD and GC. We demonstrated that besides *H. pylori*-related loci, host genetic factors potentially involved in gastric hormone regulation, cell differentiation, and proliferation might play important roles in PUD pathogenesis. Our single-cell analysis further revealed the association of serotonin-secreting EC cells, somatostatin-secreting stomach D cells, and stomach tuft cells with PUD, indicating their key role in PUD etiology. Given the findings of preliminary polygenic risk score analysis in this study, more fine-tuned polygenic risk score models can be developed using the association summary statistics generated in this study or integrating other polygenic risk models, which is promising for the personalized medicine and PUD risk screening.

References

1. Xie, X., Ren, K., Zhou, Z., Dang, C. & Zhang, H. The global, regional and national burden of peptic ulcer disease from 1990 to 2019: a population-based study. *BMC Gastroenterol* **22**, 58 (2022).
2. Hooi, J. K. Y. *et al.* Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. *Gastroenterology* **153**, 420–429 (2017).
3. Tanikawa, C. *et al.* A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. *Nat Genet* **44**, 430–434, S1–2 (2012).
4. Lanas, A. & Chan, F. K. L. Peptic ulcer disease. *Lancet* **390**, 613–624 (2017).
5. Malik, T. F., Gnanapandithan, K. & Singh, K. Peptic Ulcer Disease. in *StatPearls* (StatPearls Publishing, 2023).
6. R  ih  , I., Kempainen, H., Kaprio, J., Koskenvuo, M. & Sourander, L. Lifestyle, Stress, and Genes in Peptic Ulcer Disease: A Nationwide Twin Cohort Study. *Archives of Internal Medicine* **158**, 698–704 (1998).
7. Li, L. F. *et al.* Cigarette smoking and gastrointestinal diseases: the causal relationship and underlying molecular mechanisms (review). *Int J Mol Med* **34**, 372–380 (2014).
8. Strate, L. L. *et al.* A Prospective Study of Alcohol Consumption and Smoking and the Risk of Major Gastrointestinal Bleeding in Men. *PLOS ONE* **11**, e0165278 (2016).
9. Hsu, C.-C. *et al.* Depression and the Risk of Peptic Ulcer Disease: A Nationwide Population-Based Study. *Medicine (Baltimore)* **94**, e2333 (2015).
10. Hansson, L. E. *et al.* The risk of stomach cancer in patients with gastric or duodenal ulcer disease. *N Engl J Med* **335**, 242–249 (1996).
11. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *The American Journal of Human Genetics* **110**, 179–194 (2023).
12. Wu, Y. *et al.* GWAS of peptic ulcer disease implicates Helicobacter pylori infection, other gastrointestinal disorders and depression. *Nat Commun* **12**, 1146 (2021).
13. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415–1424 (2021).
14. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol* **27**, S2–S8 (2017).
15. Hozawa, A. *et al.* Study Profile of the Tohoku Medical Megabank Community-Based Cohort Study. *J Epidemiol* **31**, 65–76 (2021).
16. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
17. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat Commun* **10**, 4393 (2019).
18. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443–1448 (2016).
19. Byrsk  -Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022).
20. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335–1341 (2018).

21. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
22. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
23. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
24. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief Bioinform* **14**, 144–161 (2013).
25. Tan, A., Abecasis, G. R. & Kang, H. M. Unified representation of genetic variants. *Bioinformatics* **31**, 2202–2204 (2015).
26. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639–3650 (2017).
27. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
28. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241 (2015).
29. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum Genet* **99**, 76–88 (2016).
30. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369–375, S1–3 (2012).
31. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. *J R Stat Soc Series B Stat Methodol* **82**, 1273–1300 (2020).
32. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
33. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539–551 (2017).
34. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
35. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
36. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
37. Koda, Y., Soejima, M., Liu, Y. & Kimura, H. Molecular basis for secretor type alpha(1,2)-fucosyltransferase gene deficiency in a Japanese population: a fusion gene generated by unequal crossover responsible for the enzyme deficiency. *Am J Hum Genet* **59**, 343–350 (1996).
38. Viechtbauer, W. Conducting Meta-Analyses in R with the **metafor** Package. *J. Stat. Soft.* **36**, (2010).
39. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic

- association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
40. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet* **17**, e1009440 (2021).
 41. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
 42. Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M. & He, X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet* **52**, 740–747 (2020).
 43. Ishigaki, K. *et al.* Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet* **52**, 669–679 (2020).
 44. Burgess, S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* **43**, 922–929 (2014).
 45. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1776 (2019).
 46. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* **36**, 214–224 (2012).
 47. Zhong, H. & Prentice, R. L. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **9**, 621–634 (2008).
 48. Palmer, C. & Pe'er, I. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet* **13**, e1006916 (2017).
 49. Zeng, J. *et al.* Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nat Commun* **12**, 1164 (2021).
 50. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
 51. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
 52. Birney, E. *et al.* An overview of Ensembl. *Genome Res* **14**, 925–928 (2004).
 53. Busslinger, G. A. *et al.* Human gastrointestinal epithelia of the esophagus, stomach, and duodenum resolved at single-cell resolution. *Cell Rep* **34**, 108819 (2021).
 54. Suzuki, K. *et al.* Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet* **51**, 379–386 (2019).
 55. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
 56. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
 57. Gudjonsson, A. *et al.* A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat Commun* **13**, 480 (2022).
 58. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* **53**, 1712–1721 (2021).
 59. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005–D1012 (2019).
 60. Reales, G. & Wallace, C. Sharing GWAS summary statistics results in more citations: evidence from the GWAS catalog. 2022.09.27.509657 Preprint at

- <https://doi.org/10.1101/2022.09.27.509657> (2022).
61. Hayhurst, J. *et al.* A community driven GWAS summary statistics standard. 2022.07.15.500230 Preprint at <https://doi.org/10.1101/2022.07.15.500230> (2022).
 62. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* **37**, 4593–4596 (2021).
 63. Yin, L. *et al.* rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics Proteomics Bioinformatics* **19**, 619–628 (2021).
 64. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software* **3**, 731 (2018).
 65. Mbatchou, J. *et al.* Computationally efficient whole genome regression for quantitative and binary traits. *Nature Genetics* (2020) doi:10.1101/2020.06.19.162354.
 66. MacArthur, J. A. L. *et al.* Workshop proceedings: GWAS summary statistics standards and sharing. *Cell Genomics* **1**, 100004 (2021).
 67. Lyon, M. S. *et al.* The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biology* **22**, 32 (2021).
 68. Matushyn, M. *et al.* SumStatsRehab: an efficient algorithm for GWAS summary statistics assessment and restoration. *BMC Bioinformatics* **23**, 443 (2022).
 69. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308–311 (2001).
 70. Winkler, T. W. *et al.* EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
 71. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).
 72. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genom* **2**, 100192 (2022).
 73. Malone, J. *et al.* Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* **26**, 1112–1118 (2010).
 74. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894 (2019).
 75. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
 76. Sun, B. B. *et al.* Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. 2022.06.17.496443 Preprint at <https://doi.org/10.1101/2022.06.17.496443> (2022).
 77. Kanaji, T. *et al.* A Common Genetic Polymorphism (46 C to T Substitution) in the 5'-Untranslated Region of the Coagulation Factor XII Gene Is Associated With Low Translation Efficiency and Decrease in Plasma Factor XII Level. *Blood* **91**, 2010–2014 (1998).
 78. Chattopadhyay, R., Sengupta, T. & Majumder, R. Inhibition of intrinsic Xase by protein S: a novel regulatory role of protein S independent of activated protein C. *Arterioscler Thromb Vasc Biol* **32**, 2387–2393 (2012).

79. Grover, S. P. & Mackman, N. Intrinsic Pathway of Coagulation and Thrombosis. *Arterioscler Thromb Vasc Biol* **39**, 331–338 (2019).
80. Matoba, N. *et al.* GWAS of 165,084 Japanese individuals identified nine loci associated with dietary habits. *Nat Hum Behav* **4**, 308–316 (2020).
81. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390–400 (2018).
82. International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
83. Manchia, M. *et al.* The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One* **8**, e76295 (2013).
84. Take, Y., Koizumi, S. & Nagahisa, A. Prostaglandin E Receptor 4 Antagonist in Cancer Immunotherapy: Mechanisms of Action. *Front Immunol* **11**, 324 (2020).
85. Kim, Y.-D. *et al.* Risk of gastric cancer is associated with PRKAA1 gene polymorphisms in Koreans. *World J Gastroenterol* **20**, 8592–8598 (2014).
86. Wang, Y. *et al.* G protein subunit α q regulates gastric cancer growth via the p53/p21 and MEK/ERK pathways. *Oncol Rep* **37**, 1998–2006 (2017).
87. Matsubara, A. *et al.* Activating GNAS and KRAS mutations in gastric foveolar metaplasia, gastric heterotopia, and adenocarcinoma of the duodenum. *Br J Cancer* **112**, 1398–1404 (2015).
88. Sishc, B. J. & Davis, A. J. The Role of the Core Non-Homologous End Joining Factors in Carcinogenesis and Cancer. *Cancers (Basel)* **9**, 81 (2017).
89. Fattahi, S., Nikbakhsh, N., Ranaei, M., Sabour, D. & Akhavan-Niaki, H. Association of sonic hedgehog signaling pathway genes IHH, BOC, RAB23a and MIR195-5p, MIR509-3-5p, MIR6738-3p with gastric cancer stage. *Sci Rep* **11**, 7471 (2021).
90. Chen, Y. *et al.* Effects of differential distributed-JUP on the malignancy of gastric cancer. *J Adv Res* **28**, 195–208 (2021).
91. Nikolsky, Y. *et al.* Genome-wide functional synergy between amplified and mutated genes in human breast cancer. *Cancer Res* **68**, 9532–9540 (2008).
92. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621–629 (2018).
93. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
94. Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G. & Gruss, P. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature* **386**, 399–402 (1997).
95. Zhu, Y., Liu, Q., Zhou, Z. & Ikeda, Y. PDX1, Neurogenin-3, and MAFA: critical transcription regulators for beta cell development and regeneration. *Stem Cell Res Ther* **8**, 240 (2017).
96. Kosinski, C. *et al.* Indian hedgehog regulates intestinal stem cell fate through epithelial-mesenchymal interactions during development. *Gastroenterology* **139**, 893–903 (2010).
97. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophoblast differentiation. *Cell* **123**, 917–929 (2005).
98. Rorsman, P. & Huising, M. O. The somatostatin-secreting pancreatic δ -cell in health and disease. *Nat Rev Endocrinol* **14**, 404–414 (2018).

99. Ampofo, E., Nalbach, L., Menger, M. D. & Laschke, M. W. Regulatory Mechanisms of Somatostatin Expression. *Int J Mol Sci* **21**, 4170 (2020).
100. Smith, S. B., Ee, H. C., Connors, J. R. & German, M. S. Paired-Homeodomain Transcription Factor PAX4 Acts as a Transcriptional Repressor in Early Pancreatic Development. *Molecular and Cellular Biology* **19**, 8272–8280 (1999).
101. Larsson, L. I., St-Onge, L., Hougaard, D. M., Sosa-Pineda, B. & Gruss, P. Pax 4 and 6 regulate gastrointestinal endocrine cell development. *Mech Dev* **79**, 153–159 (1998).
102. Zhang, T. *et al.* Pax4 synergistically acts with Pdx1, Ngn3 and MafA to induce HuMSCs to differentiate into functional pancreatic β -cells. *Experimental and Therapeutic Medicine* **18**, 2592–2598 (2019).
103. Latorre, R., Sternini, C., De Giorgio, R. & Greenwood-Van Meerveld, B. Enteroendocrine cells: a review of their role in brain-gut communication. *Neurogastroenterol Motil* **28**, 620–630 (2016).
104. Deding, U. *et al.* Perceived stress as a risk factor for peptic ulcers: a register-based cohort study. *BMC Gastroenterol* **16**, 140 (2016).
105. Duan, S., Rico, K. & Merchant, J. L. Gastrin: From Physiology to Gastrointestinal Malignancies. *Function* **3**, zqab062 (2022).
106. Dufresne, M., Seva, C. & Fourmy, D. Cholecystokinin and gastrin receptors. *Physiol Rev* **86**, 805–847 (2006).
107. Przemeck, S. M. C. *et al.* Hypergastrinemia increases gastric epithelial susceptibility to apoptosis. *Regul Pept* **146**, 147–156 (2008).
108. Jang, E. S. *et al.* Effects of coffee, smoking, and alcohol on liver function tests: a comprehensive cross-sectional study. *BMC Gastroenterol* **12**, 145 (2012).
109. Wannamethee, S. G. & Shaper, A. G. Cigarette smoking and serum liver enzymes: the role of alcohol and inflammation. *Ann Clin Biochem* **47**, 321–326 (2010).
110. Tsuo, K. *et al.* Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity. *Cell Genom* **2**, 100212 (2022).
111. Rotter, J. I., Rimoin, D. L. & Samloff, I. M. Genetic heterogeneity in peptic ulcer. *Lancet* **1**, 1088–1089 (1979).
112. Fleshler, B. Genetics and Heterogeneity of Common Gastrointestinal Disorders. *JAMA* **246**, 390–391 (1981).
113. Rotter, J. I. Gastric and duodenal ulcer are each many different diseases. *Dig Dis Sci* **26**, 154–160 (1981).
114. Hao, Y. & Li, G. Role of EFNA1 in tumorigenesis and prospects for cancer therapy. *Biomed Pharmacother* **130**, 110567 (2020).
115. Kokkola, A. *et al.* Spontaneous disappearance of *Helicobacter pylori* antibodies in patients with advanced atrophic corpus gastritis. *APMIS* **111**, 619–624 (2003).
116. Kanai, M. *et al.* Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genom* **2**, 100210 (2022).

Acknowledgments

First, I would like to express my deepest gratitude to my supervisor, Prof. Yoichiro Kamatani, for his invaluable mentorship, comprehensive academic supervision, and financial support throughout my doctoral course at the University of Tokyo. I am extremely grateful to Dr. Masaru Koido for his comprehensive and constructive guidance that greatly helped me develop academic skills and finish the thesis for my research. Their expertise, advice, and encouragement have played a critical role not only in this research but also in my personal development as a researcher.

I would like to express my appreciation to Prof. Matsuda Koichi as both collaborator and thesis committee member, and Prof. Chizu Tanikawa as collaborator for their valuable support and insightful suggestions during my doctoral course. I am also grateful for their valuable supervision as supervisors during my master's course. I express gratitude to Prof. Mark Lathrop for his valuable support. I would like to thank Prof. Yoichi Sutoh, Dr. Yayoi Otsuka-Yamasaki, Prof. Tsuyoshi Hachiya, and Prof. Atsushi Shimizu from the Iwate Tohoku Medical Megabank Organization, Dr. Hans Markus Munter from McGill University, and Prof. Takayuki Morisaki, Dr. Akiko Nagai, Prof. Yoshinori Murakami and Dr. Yuka Shimmori from the University of Tokyo as the collaborators of this project for their valuable contributions to this study.

I am grateful to Prof. Kenta Nakai, Prof. Yutaka Suzuki, and Prof. Shumpei Ishikawa, for their valuable feedback, constructive criticism, and insightful suggestions as advisers and thesis committee members who greatly contributed to the development and improvement of this thesis. Their diverse expertise and perspectives have greatly enriched and improved this study.

I want to acknowledge all the participants and investigators of BioBank Japan, Tohoku Medical Megabank, UK Biobank, and FinnGen. I thank the Human Genome Center, the Institute of Medical Science, University of Tokyo for providing the super-computing resource used in this study.

This research was supported by the Ministry of Education, Culture, Sports, Sciences and Technology (MEXT) of Japanese government and the Japan Agency for Medical Research and Development (AMED) under grant numbers JP18km0605001/JP23tm0624002 (the BioBank Japan project) and JP19km0405215 (to C.T., K.M., and Y.K.), and JP223fa627011 (to K.M. and Y.K.). This work was supported by the Tohoku Medical Megabank Project (Special Account for the Reconstruction of the Great East Japan Earthquake) from MEXT and AMED (Grant numbers JP15km0105004 and JP21tm0124006), including the supercomputer resource powered by the AMED research grant (Grant number JP20km0405001).

Lastly, I am grateful to my family for their unwavering love throughout this demanding and long academic journey. I would like to express my heartfelt gratitude to my beloved Chunyan Li for her tremendous support and encouragement that helped me through difficult times when I was overwhelmed and uncertain. Their constant patience and emotional support have been an unlimited source of strength and motivation for me.