

論文の内容の要旨

論文題目 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)
氏 名 賀 云野 (He Yunye)

Introduction

Peptic ulcer disease (PUD), referring to the acid-induced injury of digestive tract, is one of the most common gastrointestinal disorders, with a lifetime prevalence rate of 5-10% in the general population(1). 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). 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 Europeans(2). The most well-known causes of PUD include *H. pylori* (HP) infection and the use of non-steroidal anti-inflammatory drugs (NSAIDs)(3). It has been shown that genetic factors also play an important role in PUD etiology(4). 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 genetic etiology of PUD. The differences across PUD subtypes and the key cell types contributing to PUD etiology should be investigated given that GU and DU differ in various aspects. Additionally, although epidemiological studies have suggested that DU is a protective factor against gastric cancers (GC), whether genetic factors for PUD and GC are concordant and can explain the epidemiological findings is still unclear. 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.

Methods

Cohorts: Biobank Japan (BBJ) project enrolled approximately 200,000 participants of mainly Japanese ancestry from 2003 to 2007 as the first cohort (BBJ1)(5). From 2013 to 2017, BBJ additionally collected DNA and clinical information from 67,321 newly registered participants as the second cohort (BBJ2). Genotype data of the case and control individuals included in the discovery-stage GWAS were obtained from the primary dataset of Biobank Japan 1st cohort, including 181,927 individuals (denoted as BBJ1-180K). Replication was conducted in three Japanese studies: an additional and independent set of BBJ1 cohort of 11,715 individuals (denoted as BBJ1-12K), a cohort of 42,689 individuals from BBJ2 (denoted as BBJ2-42K), and a population-based cohort of 49,621 individuals from Tohoku University Tohoku Medical Megabank (TMM) Project (denoted as TMM-50K)(6).

Phenotypes: This study assessed PUD, which is a combination of the two major subtypes (DU and GU). PUD cases were obtained from the combination of individuals with any of the two major PUD subtypes.

Datasets: The 1000 Genomes (1KG) project datasets were used as the imputation and LD reference panel; European-specific summary statistics for PUD were obtained from previous studies conducted in UK Biobank (UKB; N = ~500K) and FinnGen (N = ~340K), which are two of the largest population-based biobanks of mainly European ancestry individuals; The tissue-specificity analysis and eQTL analysis used expression datasets from the Genotype-Tissue Expression (GTEx) project, which analyzed samples from 54 non-diseased tissue sites across nearly 1000 individuals; Cell-type specific gene-sets were obtained from a previous single-cell RNA sequencing (scRNA-seq) study(7), which analyzed human biopsies of the healthy stomach and duodenum.

Association analysis: After sample and genotype quality control, imputation was performed using the 1KG Phase 3 version 5 ALL panel. Single nucleotide variant association tests were performed with SAIGE (v0.44), which implements a generalized mixed model with the saddlepoint approximation (SPA) correction controlling for case-

control imbalance and cryptic relatedness. The regression model included age, sex, and top 10 PCs as covariates. The results of GWAS at the discovery and replication stages were combined using the fixed-effect inverse-variance method implemented in METAL (v2011-03-25). The fixed-effect inverse-variance weighted (IVW) method was used to conduct meta-analyses integrating GWAS results in EAS and EUR populations with METAL.

Post-GWAS analysis: LD score regression and genetic correlation analysis was performed using LDSC (v1.0.0); cross-ancestry genetic correlation was estimated by popcorn (v1.0); fine-mapping was performed using SuSiE (v0.11.92); colocalization analysis was performed using coloc and coloc-susie (v5.1.0); associations were characterized integrating eQTL and pQTL datasets. SBasyesS was employed to estimate the polygenicity; annotation was conducted using ANNOVAR (v2020-06-07); Mendelian randomization was performed using TwoSampleMR (v0.5.6); MAGMA (v1.08) was used to conduct gene-based, gene-set-based and tissue/cell-type specificity analysis; LDSC was additionally employed for cell-type specificity analysis.

Results

1. Association analysis of PUD and its subtypes

A three-stage genome-wide analysis of PUD and its subtypes was performed in this study. An overview of the workflow is provided in **Fig. 1a**. In the discovery stage GWAS on the BBJ1-180K dataset (19,713 cases), ten genome-wide significant loci ($P < 5.0 \times 10^{-8}$) for PUD were identified, five of which had not been reported as genome-wide significant loci in previous GWASs of PUD or any subtype. Moreover, 14 significant loci were identified for DU, including seven novel loci (three of which overlapped with novel PUD loci). Replication was then conducted in individuals from three independent studies: BBJ1-12K (1,001 cases), BBJ2-42K (3,637 cases), and TMM-50K (5,388 cases); four novel lead variants (4/9) were nominally associated ($P < 0.05$) with PUD or its subtypes in the same direction in at least two replication datasets, and five novel loci (5/9) were replicated in the population-based TMM-50K. East Asian-specific meta-analysis combining the discovery GWAS and three replication GWAS was performed, 25 non-overlapping risk loci associated with PUD or any subtype, including 11 additional novel loci. Finally, a fixed-effect IVW cross-ancestry meta-analysis (52,032 PUD cases and 905,344 controls) was performed (**Fig. 1b**), combining the Japanese and European studies (summary statistics from FinnGen(8) and UKB(9)); six additional loci for PUD and DU reached the genome-wide significance level.

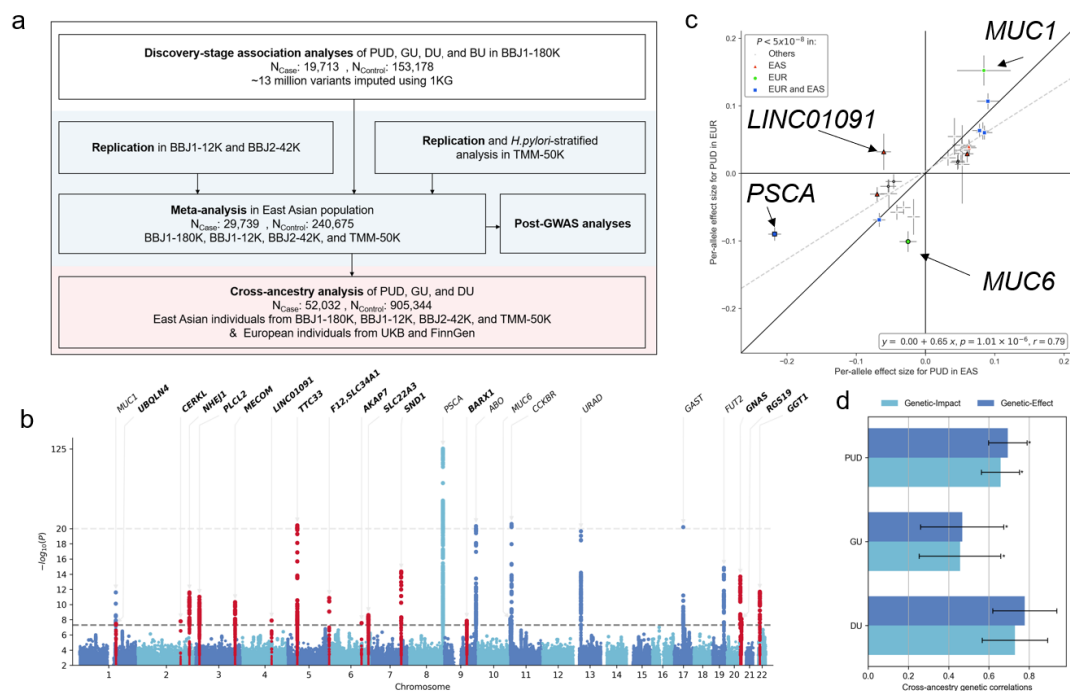


Figure 1. Study design and cross-ancestry analysis. a, overview of the three-stage study design. **b**, Manhattan plot of the

cross-ancestry meta-analysis for peptic ulcer diseases. **c**, cross-ancestry effect size comparison. **d**, cross-ancestry genetic correlation analysis.

The per-allele effect sizes of lead variants associated with PUD or any of the subtypes available for both ancestries were compared (**Fig. 1c**); The effect sizes for PUD showed a relatively high correlation (27 variants with MAF > 0.01 in both populations; $r = 0.79$) between the two ancestries. A cross-ancestry genetic correlation analysis was performed using popcorn. For PUD, the genetic impact was significantly different from one (null hypothesis: $P_{gi} = 1$) (**Fig. 1d**, genetic impact correlation $P_{gi} = 0.65$, $P = 3.0 \times 10^{-4}$), indicating the difference in genetic architecture of PUD across ancestries.

2. Post-GWAS analyses in East Asians

A stepwise conditional analysis using COJO was conducted to explore the secondary signals at the identified loci, and independent signals were detected at *PDX1* (**Fig. 2a**) and *JUP2* (**Fig. 2b**) loci, near *CDX2* and *GAST* loci (two of the previously reported loci in European individuals). Fine-mapping analysis using SuSiE¹⁹ was performed to identify the causal variants; 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. Cross-trait LD score regression was conducted to evaluate the genetic correlation across PUD-related traits (**Fig. 2c**). DU and GU showed significantly high genetic correlations ($r_g = 0.79$, FDR < 5%) with each other. HP-stratified association tests were conducted for PUD in HP-positive and HP-negative individuals from TMM-50K to examine the differences in genetic architectures between HP-induced and HP-unrelated peptic ulcers. A lead SNP at *CCKBR* was found to be HP-positive-specific (**Fig. 2d**). To examine the similarities and differences of 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 were compared; the effect sizes for GU showed a strong correlation with those for DU; however, the effect sizes for GU were systematically smaller than those for DU (**Fig. 2e**). Polygenicity estimation by SBayesS and polygenic score analysis supported a higher heterogeneity of GU. Additionally, two-sample Mendelian randomization (MR) was conducted to evaluate the causality of PUD or its subtypes on GC; although PUD and its subtypes showed significant ($P < 0.05/15$) protective effects against GC using the IVW method, MR-Egger analysis suggested significant pleiotropy for the instruments; a following effect size comparison of distinct signals in EAS between PUD subtypes and GC identified that three lead variants showed relatively strong but opposite effects on DU and GC (**Fig. 2f**).

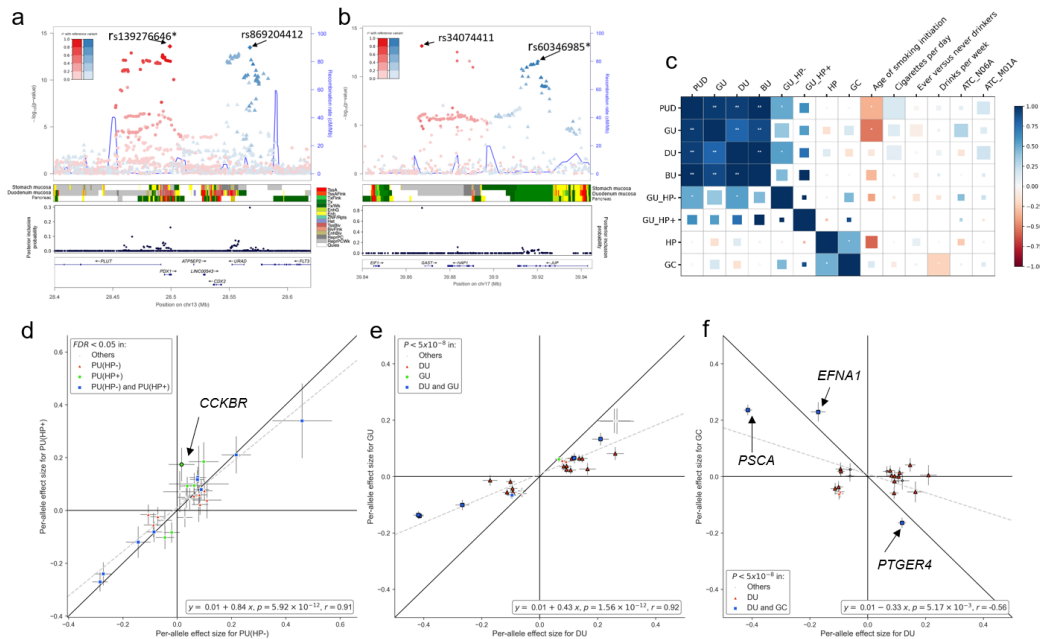


Figure 2. Post-GWAS analysis in East Asians. a,b, EAS-specific secondary signals at *PDX1* and *JUP2*. c, Genetic correlation among PUD, PUD-related phenotypes, and risk factors. d, Per-allele effect size (logarithm of odds ratios) comparison for PUD using summary statistics from *H. pylori* stratified analysis. e, Per-allele effect size comparison using EAS-specific summary

statistics for DU and GU. **f**, Per-allele effect size comparison between DU and gastric cancer (GC).

3. Tissue- and cell-type specificity analysis

We tested the tissue-level specificity using MAGMA with GTEx v8 datasets in EAS individuals to investigate the tissue types related to PUD and its subtypes; significant genetic enrichments ($FDR < 5\%$) were observed in the stomach, pancreas, small intestine, and kidney for PUD. To further characterize specific cell types associated with PUD, publicly available scRNA-seq datasets of the human stomach and duodenum(7) were obtained; the top 10% of the most specific genes for each cell type were selected as the cell-specific gene sets; cell specificity analysis using LDSC and MAGMA were performed in EAS and EUR individuals, which is followed by a fixed-effect meta-analysis combining EAS and EUR results for each method to increase statistical power. For PUD, stomach

D cells reached the significance threshold ($FDR < 5\%$) in both the analyses of MAGMA and LDSC (Fig. 3b-c). Moreover, duodenal enterochromaffin cells (EC cells), stomach antral ECs, and stomach tuft cells were significantly associated with PUD in the analysis of MAGMA.

Figure 3. 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.

Discussion

This study identified 25 novel, independent risk loci that were highly concordant across ancestries. Downstream analyses suggested that GU shared the same risk loci with DU but showed higher polygenicity and smaller genetic effect sizes than DU, indicating higher heterogeneity of GU. The cross-ancestry analysis also suggested the heterogeneity of GU across ancestries. The gene-level analysis showed that genetic factors are enriched in highly expressed genes in stomach tissue, especially in somatostatin-producing D cells and serotonin-secreting EC cells. The enrichment in D cells is consistent with the identification of missense variants in *PAX4* (transcriptional repressor for somatostatin; regulates somatostatin-producing cells of the distal stomach) and risk loci in regulatory regions of *PDX1* (activates somatostatin transcription); the enrichment in EC cells suggested the potential involvement of serotonin in PUD etiology, considering the bidirectional effects of the brain-gut axis and the causal role of major depression in PUD(9). HP-stratified analysis found an HP-positive ulcer-specific locus at *CCKBR* (receptor for gastrin). In summary, this study provided genetic evidence of gastrointestinal hormone regulation being critical in PUD etiology.

References

1. X. Xie, K. Ren, Z. Zhou, C. Dang, H. Zhang, *BMC Gastroenterol.* **22**, 58 (2022).
2. C. Tanikawa *et al.*, *Nat Genet.* **44**, 430–434, S1–2 (2012).
3. A. Lanas, F. K. L. Chan, *Lancet.* **390**, 613–624 (2017).
4. H. M. Malaty, D. Y. Graham, I. Isaksson, L. Engstrand, N. L. Pedersen, *Arch Intern Med.* **160**, 105–109 (2000).
5. A. Nagai *et al.*, *J Epidemiol.* **27**, S2–S8 (2017).
6. A. Hozawa *et al.*, *J Epidemiol.* **31**, 65–76 (2021).
7. G. A. Busslinger *et al.*, *Cell Rep.* **34**, 108819 (2021).
8. M. I. Kurki *et al.*, *Nature.* **613**, 508–518 (2023).
9. Y. Wu *et al.*, *Nat Commun.* **12**, 1146 (2021).

