

Genome-wide association meta-analysis identifies novel *GP2* gene risk variants for pancreatic cancer in the Japanese population

Yingsong Lin^{*1,46}, Masahiro Nakatochi^{*2,46}, Hidemi Ito^{3,4,46}, Yoichiro Kamatani^{5,6}, Akihito Inoko⁷, Hiromi Sakamoto⁸, Fumie Kinoshita², Yumiko Kobayashi², Hiroshi Ishii⁹, Masato Ozaka¹⁰, Takashi Sasaki¹⁰, Masato Matsuyama¹⁰, Naoki Sasahira¹⁰, Manabu Morimoto¹¹, Satoshi Kobayashi¹¹, Taito Fukushima¹¹, Makoto Ueno¹¹, Shinichi Ohkawa¹¹, Naoto Egawa¹², Sawako Kuruma¹³, Mitsuru Mori¹⁴, Haruhisa Nakao¹⁵, Yasushi Adachi¹⁶, Masumi Okuda¹⁷, Takako Osaki¹⁸, Shigeru Kamiya¹⁸, Chaochen Wang¹, Kazuo Hara¹⁹, Yasuhiro Shimizu²⁰, Tatsuo Miyamoto²¹, Yuko Hayashi²², Yasuyuki Hosono²², Hiromichi Ebi²², Tomohiro Kohmoto^{23,24}, Issei Imoto²⁴, Yoshinori Murakami²⁵, Masato Akiyama^{5,26}, Kazuyoshi Ishigaki⁵, Koichi Matsuda²⁷, Makoto Hirata²⁵, Katsuaki Shimada²⁸, Takuji Okusaka²⁹, Takahisa Kawaguchi³⁰, Meiko Takahashi³⁰, Yoshiyuki Watanabe³¹, Kiyonori Kuriki³², Aya Kadota³³, Kenji Wakai³⁴, Taiki Yamaji³⁵, Motoki Iwasaki³⁵, Norie Sawada³⁵, Shoichiro Tsugane³⁶, Kengo Kinoshita³⁷, Nobuo Fuse³⁷, Fumiki Katsuoka³⁷, Atsushi Shimizu³⁸, Satoshi S. Nishizuka³⁸, Kozo Tanno^{38, 39}, Ken Suzuki^{40, 5, 41, 42}, Yukinori Okada^{5, 42,43}, Momoko Horikoshi⁴¹, Toshimasa Yamauchi⁴⁰, Takashi Kadowaki⁴⁰, Teruhiko Yoshida⁸, Fumihiko Matsuda³⁰, Michiaki Kubo⁴⁴, Shogo Kikuchi¹, Keitaro Matsuo^{* 7,45}

¹Department of Public Health, Aichi Medical University School of Medicine, Nagakute, Japan

²Data Science Division, Data Coordinating Center, Department of Advanced Medicine, Nagoya University Hospital, Nagoya, Japan

³Division of Cancer Information and Control, Aichi Cancer Center Research Institute, Nagoya, Japan

⁴Department of Descriptive Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan

⁵Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

⁶Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University

Graduate School of Medicine, Kyoto, Japan

⁷Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan

⁸Genetics Division, National Cancer Center Research Institute, Tokyo, Japan,

⁹Chiba Cancer Center, Chiba, Japan

¹⁰Department of Hepato-biliary-pancreatic Medicine, The Cancer Institute Hospital of Japanese Foundation for Cancer Research, Tokyo, Japan

¹¹Department of Gastroenterology, Hepatobiliary and Pancreatic Medical Oncology Division, Kanagawa Cancer Center, Yokohama, Japan

¹²Department of Gastroenterology, Tokyo Metropolitan Hiroo Hospital, Tokyo, Japan

¹³Department of Gastroenterology, Tokyo Metropolitan Komagome Hospital, Tokyo, Japan

¹⁴Hokkaido Chitose College of Rehabilitation, Hokkaido, Japan

¹⁵Division of Hepatology and Pancreatology, Aichi Medical University School of Medicine, Nagakute, Japan

¹⁶Sapporo Shirakabadai Hospital, Sapporo, Japan

¹⁷Department of Pediatrics, Aichi Medical University School of Medicine, Nagakute, Japan

¹⁸Department of Infectious Diseases, Kyorin University School of Medicine, Tokyo, Japan

¹⁹Department of Gastroenterology, Aichi Cancer Center Hospital, Nagoya, Japan

²⁰Department of Gastroenterological Surgery, Aichi Cancer Center Hospital, Nagoya, Japan

²¹Department of Genetics and Cell Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan

²²Division of Molecular Therapeutics, Aichi Cancer Center Research Institute, Nagoya, Japan

²³Department of Human Genetics, Tokushima University Graduate School of Medicine, Tokushima, Japan

²⁴Division of Molecular Genetics, Aichi Cancer Center Research Institute, Nagoya, Japan

²⁵Institute of Medical Science, The University of Tokyo, Tokyo, Japan

²⁶Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.

²⁷Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

²⁸Department of Hepatobiliary and Pancreatic Surgery, National Cancer Center Hospital, Tokyo, Japan

²⁹Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital, Tokyo, Japan

³⁰Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

³¹Department of Epidemiology for Community Health and Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan

³²Laboratory of Public Health, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan

³³Department of Health Science, Shiga University of Medical Science, Otsu, Japan

³⁴Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan

³⁵Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan

³⁶Center for Public Health Sciences, National Cancer Center, Tokyo, Japan

³⁷Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan

³⁸Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan.

³⁹Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University, Iwate, Japan.

⁴⁰Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The

University of Tokyo, Tokyo, Japan.

⁴¹Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan.

⁴²Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan.

⁴³Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan.

⁴⁴RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

⁴⁵Department of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan

⁴⁶These authors contributed equally to this work

*Correspondence:

Yingsong Lin (linys@aichi-med-u.ac.jp), Masahiro Nakatochi (mnakatochi@med.nagoya-u.ac.jp), and Keituro Matsuo (kmatsuo@aichi-cc.jp)

Abstract

The etiology of pancreatic cancer remains largely unknown. Here, we report the results of a meta-analysis of three genome-wide association studies (GWASs) comprising 2,039 pancreatic cancer cases and 32,592 controls, the largest sample size in the Japanese population. We identified 3 (13q12.2, 13q22.1, and 16p12.3) genome-wide significant loci ($P < 5.0 \times 10^{-8}$) and 4 suggestive loci ($P < 1.0 \times 10^{-6}$) for pancreatic cancer. Of these risk loci, 16p12.3 is novel; the lead SNP maps to rs78193826 (odds ratio (OR)=1.46, 95% CI=1.29-1.66, $P=4.28 \times 10^{-9}$), an Asian-specific, nonsynonymous *glycoprotein 2 (GP2)* gene variant predicted to be highly deleterious. Additionally, the gene-based GWAS identified a novel

gene, *KRT8*, which is linked to exocrine pancreatic and liver diseases. The identified *GP2* gene variants were pleiotropic for multiple traits, including type 2 diabetes, hemoglobin A1c (HbA1c) levels, and pancreatic cancer. Mendelian randomization analyses corroborated causality between HbA1c and pancreatic cancer. These findings suggest that *GP2* gene variants are associated with pancreatic cancer susceptibility in the Japanese population, prompting further functional characterization of this locus.

Introduction

With approximately 33,000 related deaths every year, pancreatic cancer is the fourth leading cause of cancer deaths in Japan, after lung, colorectal, and stomach cancers¹. The incidence and mortality rates of pancreatic cancer have increased steadily over the past decades, while those of other gastrointestinal cancers have shown a decreasing trend¹. Despite the increasing burden levied by pancreatic cancer, few modifiable risk factors other than smoking and type 2 diabetes mellitus (T2D) have been identified, and the 5-year survival rates remain the worst (<10%) among major malignancies.

Genome-wide association studies (GWASs) have increasingly revealed the role of inherited genetic variations in pancreatic cancer susceptibility. Since the first GWAS, conducted by the PanScan consortium, identified common variants in the gene coding for the ABO blood group system in 2009,² at least 23 genome-wide significant susceptibility loci have been linked to pancreatic cancer risk³. However, fewer loci have been identified for pancreatic cancer than for other common cancers, including breast and colorectal cancers^{4,5}. Furthermore, the identified risk variants explained approximately 13% of the total heritability on the basis of GWAS-identified SNPs in a population of European ancestry⁶. These observations suggest that additional risk loci can be identified by increasing the sample size, as evidenced by the trend in the numbers of novel variants reported by PanScan. It is also important to expand the

GWAS to populations of non-European ancestry because of differences in minor allele frequencies (MAFs) and patterns of linkage disequilibrium (LD) across diverse populations⁷. In fact, previous GWASs focusing exclusively on populations of Eastern Asian ancestry led to the identification of new susceptibility loci for breast and colorectal cancers^{8, 9}.

The majority of the risk loci for pancreatic cancer were discovered in the PanScan GWASs, which included populations of European ancestry. Only two GWASs have been conducted in East Asian populations: one in China¹⁰ and one in Japan¹¹. A total of 8 risk loci (5 genome-wide significant loci and 3 loci with suggestive evidence of association) have been identified for pancreatic cancer, but these loci were not replicated in a previous study using samples from European populations¹². Therefore, the role of common susceptibility loci in East Asian populations remains uncertain and needs further exploration. To detect additional susceptibility loci for pancreatic cancer, we conducted another GWAS in the Japanese population and then performed a meta-analysis combining all published and unpublished GWAS data in Japan.

Results

After imputation and quality control, we performed a meta-analysis of three Japanese GWASs comprising 2,039 cases and 32,592 controls as well as 7,914,378 SNPs (**Supplementary Table 1 and Table 2**). Genomic control adjustment was not applied because there was little evidence of genomic inflation ($\lambda=1.03$, **Supplementary Figure 1**). We observed genome-wide significant ($P<5.0\times 10^{-8}$) association signals at 3 loci (13q12.2, 13q22.1, and 16p12.3) (**Figure 1 and Table 1**), for which the genes nearest the lead SNP were *PLUT* (*PDX1-AS1*), *KLF5*, and *GP2*, respectively. In addition, 4 loci (1p13.2 (*WNT2B*), 2p12 (*CTNNA2*), 3p12.3 (*ROBO2*), and 9q34.2 (*ABO*)) showed suggestive evidence of association ($P<1.0\times 10^{-6}$) (**Figure 1 and Table 1**). The association results of each study are

shown in **Supplementary Table 3**. Among these risk loci, 16p12.3 is a novel locus, with genome-wide significant associations observed for 10 SNPs in this region (rs78193826, rs117267808, rs73541251, rs4609857, rs4544248, rs4632135, rs4420538, rs73541271, rs4383154, and rs4383153) (**Supplementary Table 4**). The ORs for these variants ranged from 1.43 to 1.47, indicating stronger associations in this region than the associations indicated by the odds ratios (ORs) for variants identified in previously published GWASs. The lead SNP maps to rs78193826, a nonsynonymous variant of the *GP2* (**Figure 2**) gene. The risk increased by 46% per copy of the minor T allele (OR=1.46, 95% CI=1.29-1.66, $P=4.28 \times 10^{-9}$) (**Table 1**). Regional association plots for the other loci are shown in **Supplementary Figure 2**. According to the 1000 Genomes Project Phase 3 database, rs78193826 is polymorphic, with a MAF ranging from 3.9% to 6.6% in Asian populations, compared with the much lower MAF (<0.1%) in other human populations (**Supplementary Table 4**). LD maps of these 10 SNPs at 16p12.3 are shown in **Supplementary Figure 3**. Complete LD between 9 of these SNPs (all except rs4420538) was observed in the Japanese population. Among the 10 SNPs in this region, only rs4383153 had available association summary statistics in the previous PanScan publications^{2,13}, but this SNP was not significantly associated with pancreatic cancer risk (**Supplementary Table 5**). The functional annotation results for the 10 SNPs at 16p12.3 are shown in **Supplementary Table 4**. The lead SNP rs78193826 was classified as “damaging” according to the Sifting Intolerant from Tolerant (SIFT) algorithm and as “possibly damaging” by Polymorphism Phenotyping v2 (PolyPhen-2). Moreover, the estimated combined annotation-dependent depletion (CADD) score was 20.3. For replication, we selected 4 SNPs (rs78193826, rs73541251, rs117267808, rs4632135) that met either of the following criteria: 1) exonic SNP or 2) intronic SNP with a score of 3 or less according to the Regulome DB database.

Analysis of another independent replication cohort comprising 507 cases and 879 controls

showed that rs4632135 (an intronic variant) was nominally significantly associated with pancreatic cancer risk ($P < 0.05$), whereas the other 3 SNPs did not show nominal significance (**Table 2**). However, the direction and magnitude of the effects for all 4 SNPs was consistent with those observed in the GWAS meta-analysis. Furthermore, the combined analysis of the three GWASs and the replication dataset yielded lower P values than those of the GWAS meta-analysis for each SNP (**Table 2**), suggesting that the novel risk locus at 16p12.3 discovered in our GWAS meta-analysis was unlikely to be false positive.

We also examined the previously published pancreatic cancer risk loci from PanScan¹⁴, noting that 11 of those 19 SNPs were nominally statistically significant (rs13303010 at 1p36.33, rs3790844 at 1q32.1, rs9854771 at 3q28, rs2736098 at 5p15.33, rs401681 at 5p15.33, rs10094872 at 8q24.21, rs505922 at 9q34.2, rs9581943 at 13q12.2, rs7214041 at 17q24.3, and rs16986825 at 22q12.1) or genome-wide significant (rs9543325 at 13q22.1) in our GWAS meta-analysis (**Supplementary Table 6**). Notably, we confirmed the nominally significant association ($P < 3.84 \times 10^{-5}$) between rs505922 of the ABO locus and pancreatic cancer risk.

Epidemiological studies have consistently shown that longstanding T2D is associated with an increased risk of pancreatic cancer¹⁵. Recently, 88 genetic variants, including both novel and established variants, were reported in a GWAS meta-analysis of T2D in the Japanese population¹⁶. We found that the top 3 SNPs at 16p12.3 (rs78193826, rs117267808, and rs73541251) were also genome-wide significantly associated with the risk of T2D in the latest GWAS comprising 191,764 Japanese subjects (**Supplementary Table 7**). In addition, these 3 SNPs were nominally significantly associated with hemoglobin A1c (HbA1c) ($P < 1 \times 10^{-4}$) and blood glucose levels ($P < 0.01$), which were included in another GWAS of quantitative traits in 42,790 and 93,146 Japanese subjects, respectively¹⁷. In addition to the *GP2* SNPs, 4 of the 82 T2D-associated SNPs (rs838720 at 2q37.1 (*DGKD*), rs2233580 at 7q32.1 (*PAX4*), rs2290203 at 15q26.1 (*PRCI-ASI*), and rs663129 at 18q21.32 (*MC4R*)) were nominally significant in

our GWAS meta-analysis (**Supplementary Table 8**). In addition, one of the 25 HbA1c-associated SNPs (rs4728092 at 7q32.1 (*SND1*)) was nominally significant in our GWAS meta-analysis (**Supplementary Table 9**).

To examine whether the associations of T2D and T2D-related quantitative traits with pancreatic cancer are consistent with a causal effect, we performed a Mendelian randomization (MR) analysis with the inverse variance-weighted (IVW) and MR-Egger methods. No significant associations were observed between SNP-modulated T2D and pancreatic cancer based on the IVW method (**Figure 3a**). Instead, genetically increased HbA1c levels appeared to be causally related to an increased risk of pancreatic cancer, on the basis of the significant results with both the IVW and MR-Egger methods (**Figure 3b and Supplementary Figure 4b**).

To complement the SNP-based GWAS, we performed a gene-based GWAS using MAGMA¹⁸ (**Supplementary Figure 5**). We confirmed the significant associations for *GP2* and *WNT2B* identified by the SNP-based GWAS. Notably, a novel significant association (Bonferroni-corrected $P < 2.84 \times 10^{-6}$) for the gene *KRT8* was observed (**Supplementary Table 10 and Supplementary Figures 5 and 6**), and this association was further replicated in the PanScan 1 and PanScan 2 datasets ($P=0.024$)¹³.

Discussion

The role of inherited common genetic variations in pancreatic cancer susceptibility remains incompletely understood. We identified and replicated a novel risk locus at 16p12.3 for pancreatic cancer through combining three GWAS datasets in the Japanese population. Furthermore, we provided evidence that the identification of this new locus can be attributed to the observed differences in the MAF of the lead SNP (rs78193826) at 16p12.3 and the LD structure in this region across ethnic populations.

Little overlap has been observed when risk loci reported from previous Chinese or Japanese GWASs are compared with those reported in the PanScan GWASs². By including more than twice the number of cases than were included in previous Japanese or Chinese GWASs as well as imputed SNP data, we replicated the majority of the significant risk loci discovered in the PanScan GWASs (**Supplementary Table 6**). Moreover, for most variants, the direction and magnitude of the effects in our GWAS meta-analysis of Japanese subjects were consistent with those in populations of European ancestry. These findings suggested that GWAS-identified causal variants at many loci are shared across ancestral groups and that the lack of replication may be due to an insufficient sample size in previous Chinese or Japanese GWASs.

Several lines of evidence indicate that rs78193826 is most likely a causal variant at 16p12.3, which harbors the *GP2* gene. First, this variant is nonsynonymous; the nucleotide mutation from C to T causes an amino acid change from valine to methionine, which could affect protein structure and function. Second, functional annotations in several databases consistently indicate that this variant is highly pathogenic. Third, the observed differences in the MAF of rs78193826 as well as the LD structure across different ethnic populations provide indirect evidence supporting its role as a causal variant in the Japanese population. The frequency of the minor T allele of rs78193826 is 0.1% in populations of European ancestry but 7% in the Japanese population. Given this apparent difference in the MAF, rs78193826 could not have been identified in the PanScan GWASs, although the PanScan GWASs included a much larger sample size than our GWAS. Of the 10 SNPs in this region, only rs4383153 has association summary statistics available in PanScan publications; however, no significant associations were observed between this SNP and pancreatic cancer risk (**Supplementary Table 5**). While complete LD between rs4383153 and rs78193826 was evident in the Japanese population, no LD data were available for these two SNPs in the European ancestry populations (1000 Genomes Project Phase 3 CEU). As the frequency of

the minor T allele of rs78193826 ranges from 3.9% to 6.6% in other Asian populations, rs78193826 is likely to be a causal variant for pancreatic cancer that is specific to Asian populations. However, further transethnicity replicability and fine mapping are necessary to establish the causal role of this variant.

Genetic variations in the *GP2* gene have been linked to several phenotypes in addition to pancreatic cancer. The SNP rs12597579, located in the upstream region of the *GP2* gene, has been associated with body mass index (BMI) in a GWAS including East Asians¹⁹. However, rs12597579 was not in LD with rs78193826 ($r^2=0.003$, calculated from Japanese samples in the 1000 Genomes Project Phase 3), suggesting that rs12597579 may have functions different from those of rs78193826. Coincidentally, the lead variant (rs117267808) in the *GP2* gene identified in the latest GWAS meta-analysis of T2D in the Japanese population is the same variant that we identified in our GWAS meta-analysis of pancreatic cancer (**Supplementary Table 7**). Of the 82 T2D-related SNPs, 5 showed significant associations ($P<0.05$) with pancreatic cancer, suggesting that pancreatic cancer and T2D may share specific genetic susceptibility factors. Furthermore, the risk alleles of rs78193826 and rs117267808 were identical for pancreatic cancer and T2D. Together, these findings indicate that *GP2* variants may exert pleiotropic effects on multiple traits.

The newly identified lead SNP (rs78193826) encodes the GP2 protein, which is present on the inner surface of zymogen granules in pancreatic acinar cells²⁰. GP2 is a glycosylated protein of ~90 kDa that contains multiple sites, such as an asparagine-linked glycosylation site, a zona pellucida (ZP) domain, and a glycosylphosphatidylinositol (GPI) linkage to the membrane²¹. During the secretory process, GP2 is cleaved from the membrane and secreted into the pancreatic duct along with other digestive enzymes. The expression of GP2 is extremely high in normal pancreatic tissues compared to that in other tissues (**Supplementary Figure 7**)²². However, according to the GEPIA database, pancreatic tumor tissues have a

decreased level of GP2 expression compared with that in nontumor tissues (**Supplementary Figure 8**)²³. The functional consequence of the lead SNP can be, in part, inferred from its location in the *GP2* gene: the ZP domain. In addition to GP2, an increasing number of proteins, such as uromodulin²⁴ and the transforming growth factor (TGF)- β receptor III²⁵, have been found to contain a ZP domain, the functions of which involve extracellular matrix formation and polymerization²⁶. In particular, uromodulin, the most abundant protein secreted by the kidney, shares 55% amino acid sequence identity with GP2 in the ZP-C subdomain²⁴. Pathogenic mutations in the *UMOD* gene have been shown to affect the structure and polymerization of uromodulin, resulting in kidney disease²⁴. Given the similarity in the amino acid sequence as well as the expected functionality between these two proteins, genetic variations in the ZP domain of *GP2* might also be associated with disease risk. Specifically, our GWAS finding that genetic variations in the *GP2* gene are associated with pancreatic cancer risk elucidates a possible role of bacterial infection in the pathogenesis of pancreatic cancer. The biological function of GP2 in the pancreas remains unclear; previous experiments did not show changes in exocrine secretion and morphology in *GP2* knockout mice²⁷. However, accumulating evidence has demonstrated the effects of GP2 on the innate immune response^{28,29}. The GP2 is also expressed in the membranous (M) cells of the intestinal epithelium in humans and mice, where it acts as an uptake receptor for a subset of commensal and pathogenic bacteria³⁰. For example, GP2 and its closest homolog, uromodulin, have been shown to bind to *Escherichia coli* (*E. coli*) that express type 1 fimbriae³¹. In particular, uromodulin null mice showed increased sensitivity to urinary tract infections²⁴. These findings suggest that GP2 may also play a role in host defense in the pancreas, given that proteobacteria have been detected in pancreatic ductal adenocarcinoma samples as well as in the normal human pancreas³².

Previous epidemiological studies have suggested that HbA1c levels, even in nondiabetic

ranges, or changes in HbA1c levels in new-onset T2D are associated with pancreatic cancer risk^{33,34}. Our MR analysis provided corroborating evidence that genetically increased HbA1c levels may be causally associated with pancreatic cancer risk. This result was also partially consistent with a previous MR analysis, in which T2D was not causally implicated but BMI and fasting insulin were causally associated with pancreatic cancer³⁵. Because known HbA1c-related genetic variants explain little of the variance in HbA1c levels, further studies are needed to strengthen the causal inference by incorporating more variants. The null findings on the causal role for T2D may reflect the phenotypic and genetic heterogeneity of T2D, but T2D may also be both a cause and consequence of pancreatic cancer³⁵.

Three genome-wide significant genes (*GP2*, *WNT2B*, and *KRT8*) emerged in the gene-based GWAS. Among these genes, *WNT2B* was reported in the previous PanScan GWAS as a novel gene at 1p13.1 with suggestive evidence of association⁶, and *KRT8* is a novel finding. *KRT8* belongs to a group of intermediate-filament cytoskeletal proteins involved in maintaining epithelial structural integrity³⁶. *KRT8* is expressed in both ductal and acinar single-layer epithelia, and mutations in the *KRT8* gene have been linked to exocrine pancreatic disorders and liver disease^{37, 38}.

In conclusion, our GWAS meta-analysis identified a novel risk locus at chromosome 16p12.3, which harbors the *GP2* gene, for pancreatic cancer in the Japanese population. Further fine mapping and functional characterization are required to elucidate the effects of common *GP2* gene variants on pancreatic cancer susceptibility. Moreover, our findings highlight genetic susceptibility factors shared between T2D and pancreatic cancer.

Methods

Study samples. We performed a GWAS meta-analysis based on three Japanese studies: the Japan Pancreatic Cancer Research (JaPAN) consortium GWAS, the National Cancer Center

(NCC) GWAS, and the BioBank Japan (BBJ) GWAS. An overview of the characteristics of the study populations is provided in **Supplementary Table 1**. Information on the study-specific genotyping, imputation, and analysis tools is provided in **Supplementary Table 2**. All studies were imputed based on the 1000 Genomes Project reference panel (Phase 3).

JaPAN consortium GWAS. Participants in this GWAS were drawn from the JaPAN consortium³⁹. Two case-control datasets were combined, resulting in a total of 945 pancreatic cancer cases and 3134 controls. The vast majority of cases were diagnosed as primary adenocarcinoma of the exocrine pancreas (ICD-O-3 codes C250–C259). The first dataset included 622 pancreatic cancer patients who were recruited from January 2010 to July 2014 at five participating hospitals in the Central Japan, Kanto, and Hokkaido regions. This multi-institutional case-control study collected questionnaire data on demographic and lifestyle factors, as well as 7-ml blood samples, from the study participants. The second dataset included 323 patients with newly diagnosed pancreatic cancer and 3134 control subjects recruited to an epidemiological research program at Aichi Cancer Center (HERPACC) between 2005 and 2012. All new outpatients on their first visit to Aichi Cancer Center were invited to participate in HERPACC. Those who agreed to participate filled out a self-administered questionnaire and provided a 7-ml blood sample. After quality control, 943 cases and 3057 controls remained for the subsequent analysis (**Supplementary Table 2**). None of the control subjects had a diagnosis of cancer at the time of recruitment. Written informed consent was obtained from all study participants, and the study protocol was approved by the Ethical Review Board of Aichi Medical University, the Institutional Ethics Committee of Aichi Cancer Center, the Human Genome and Gene Analysis Research Ethics Committee of Nagoya University, and the ethics committees of all participating hospitals.

BBJ GWAS. Pancreatic cancer cases were obtained from the BBJ GWAS, which was launched in 2003 and collected DNA and clinical information from approximately 200,000

patients, including those with pancreatic cancer⁴⁰. Overall, 422 pancreatic cancer cases with available genotype data were recruited from 2003 to 2008. Clinical information was collected using a standardized questionnaire. This study was approved by the ethics committees of the RIKEN Center for Integrative Medical Sciences. The controls were drawn from the participants in four population-based cohort studies in Japan: the Japan Multi-Institutional Collaborative Cohort Study (J-MICC), the Japan Public Health Center-based Prospective Study (JPHC), the Tohoku Medical Megabank Project Organization (ToMMo), and the Iwate Tohoku Medical Megabank Organization (IMM). A total of 28,870 controls who passed genotype data quality control assessments were included in the study. In all participating cohort studies, informed consent was obtained from the participants by following the protocols approved by the corresponding institutional ethics committees. The detailed descriptions of the BBJ and each cohort study are provided in **the Supplementary Note**.

NCC GWAS. The case and control samples were derived from a previous pancreatic cancer GWAS¹¹. The cases were 677 patients diagnosed with invasive pancreatic ductal adenocarcinoma at the NCC Hospital, Tokyo, Japan. The control population consisted of 677 Japanese volunteers who participated in a health checkup program in Tokyo. After preimputation quality control, 674 cases and 674 controls remained for the subsequent analysis (**Supplementary Table 2**). This project was approved by the ethics committee of the NCC.

Quality control after genotype imputation

After genotype imputation, quality control was applied to each study. SNPs with an imputation quality of $r^2 < 0.5$ or a MAF of < 0.01 were excluded. SNPs that passed quality control in at least two cohorts were included in the meta-analysis.

Association analysis for SNPs and pancreatic cancer

The association of pancreatic cancer with SNP allele dose was tested using logistic

regression analysis with adjustment for the top 2 principal components. Other known covariates, such as age, sex, and cigarette smoking, were not included in the analysis because the inclusion of covariates has been shown to substantially reduce the power for the identification of disease-associated variants when the disease prevalence is less than 2%⁴¹. The effect sizes and standard errors were used in the subsequent meta-analysis.

Meta-analysis

We performed a meta-analysis of three pancreatic cancer GWASs (JaPAN, BBJ and NCC). The association results for each SNP across the studies were combined with METAL software in a fixed effects inverse variance-weighted meta-analysis. Heterogeneity in allelic effects was assessed using the I^2 index. The meta-analysis included 7,914,378 SNPs with genotype data available from at least two cohorts. A P value threshold of 5×10^{-8} was used to determine genome-wide significance. We assessed the inflation of test statistics using the genomic control lambda.

Replication analysis

The replication cohort comprised 507 cases and 879 controls who were recruited under the same framework as the multi-institutional case-control study of the JaPAN consortium (**Supplementary Table 1**). The cases and controls were recruited from August 2014 to March 2018 at five participating hospitals in the Central Japan, Kanto, and Hokkaido regions. The control subjects included outpatients as well as screening participants with no diagnosis of cancer. SNP genotyping was performed using a Fluidigm SNP assay at the Aichi Cancer Center Research Institute, with the laboratory staff blinded to the case-control status. The association of pancreatic cancer with SNP allele dose was tested using unadjusted logistic regression analysis.

Functional annotations

To prioritize the associated SNPs at the novel loci, we adopted a series of bioinformatic

approaches to collate functional annotations. We first used ANNOVAR⁴² to obtain an aggregate set of functional annotations—including the gene location and the impact of the amino acid substitution based on prediction tools such as SIFT, PolyPhen-2, and CADD—for SNPs with a P value of $<5 \times 10^{-8}$ for pancreatic cancer. We also explored potential effects on gene regulation by annotating these SNPs using the RegulomeDB database⁴³.

MR analysis

We performed MR analyses using independent, genome-wide significant T2D-associated or HbA1c-associated SNPs, which were available from two published GWAS meta-analyses in Japanese subjects, as instrumental variables^{16,17}. For the two-sample MR analysis of T2D and pancreatic cancer, we did not exclude the overlapping samples (15.5% found only in the controls) because retaining these samples was unlikely to introduce substantial bias⁴⁴. A total of 106 pancreatic cancer cases were excluded in the HbA1c GWAS, and the effect sizes for the HbA1c-associated SNPs were reestimated. After the exclusion of 6 SNPs on the X chromosome (5 SNPs for T2D and 1 SNP for HbA1c) and an SNP for T2D (rs35678078) without genotype data, the summary data for 82 T2D-related SNPs and 25 HbA1c-related SNPs and the associations of these SNPs with pancreatic cancer risk were analyzed using IVW and MR-Egger regression methods. MR analysis was performed with the MendelianRandomization package⁴⁵.

Gene-based analysis

SNP-based P values were combined into gene-based P values using MAGMA software version 1.06¹⁸. SNP summary statistics (P values) from the meta-analysis were used as input for MAGMA. In gene-based association tests, LD between SNPs was accounted for, and the P value threshold for genome-wide significant associations was set at 2.84×10^{-6} . The 1000 Genomes reference panel (Phase 3, East Asian) was used to control for LD. We did not include any upstream/downstream regions around the genes in this analysis; only variants

located between the first exon and the last exon of a gene were used to calculate the gene-based P values. The NCBI Gene database was used to define genomic intervals for protein-coding genes. To replicate the association between *KRT8* and pancreatic cancer, we applied SNP summary statistics from PanScan 1 and PanScan 2 (pha002889.1)¹³ to MAGMA. The MAF of the Haplotype Map (HapMap) project Phase 2 CEU samples for each SNP was added to the summary statistics because the pha002889.1 data did not include the MAFs. We excluded variants with a call rate (CR) of $> 95\%$ in the cases, a CR of $> 95\%$ in the controls, a Hardy-Weinberg equilibrium test P value of $< 1 \times 10^{-6}$ in the controls, and/or a MAF of < 0.01 . The 1000 Genomes reference panel (Phase 3, European) was used to control for LD. The significance level was set at $\alpha=0.05$.

Acknowledgments

We thank Mayuko Masuda, Kikuko Kaji, Kazue Ando, Etsuko Ohara, Sumiyo Asakura, and Keiko Hanai for assistance with data collection. We thank Yumiko Kasugai for assistance with SNP genotyping. We would such as to express our gratefulness to the staff of BioBank Japan for their outstanding assistance.

This work was supported by the Ministry of Health, Labor, and Welfare of Japan (H21-11-1) and the Ministry of Education, Culture, Sports, Science, and Technology of Japan (Nos. 16H06277, 17K09095, 26253041, and 17015018). HERPACC, a part of the JaPAN consortium, was supported by Grants-in-Aid for Scientific Research for Priority Areas of Cancer (No. 17015018) and Innovative Areas (No. 221S0001) and JSPS KAKENHI grants (Nos. 26253041, 15H02524, 16H06277, and 18H03045) from the Japanese Ministry of Education, Science, Sports, Culture and Technology; and a Grant-in-Aid for the Third Term Comprehensive 10-year Strategy for Cancer Control from the Ministry of Health, Labor and Welfare of Japan and the Cancer Biobank Aichi.

This study was partially supported by the BioBank Japan project and the Tohoku Medical Megabank project, which is supported by the Ministry of Education, Culture, Sports, Sciences and Technology of Japan and the Japan Agency for Medical Research and Development.

The JPHC Study has been supported by the National Cancer Center Research and Development Fund since 2011 and was supported by a Grant-in-Aid for Cancer Research from the Ministry of Health, Labor and Welfare of Japan from 1989 to 2010.

The J-MICC Study was supported by Grants-in-Aid for Scientific Research for Priority Areas of Cancer (No. 17015018) and Innovative Areas (No. 221S0001) and JSPS KAKENHI Grant (No. 16H06277) from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

Author contributions

Y.L., M.N., H.I., A.I., M.O., T.O., S.Kamiya., and C.W. designed the study. Y.L, M.N., and Y. Kamatani wrote the manuscript. M.N., F.K., Y. Kobayashi., Y.Kamatani, M.A. K.I. performed the statistical analysis. H.S., H.I., M.O., T.S., M.Matsuyama., N.S., M.Morimoto., S.Kobayashi., T.F., M.U., S.O., N.E., S.Kuruma. M.Mori., H.N., Y.A., K.H., Y.S., Y.M., K.M., M.Hirata., K.Shimada., T.O, Y.W., K.Kuriki., A.K., K.W., T.Yamaji., M.I., N.S., S.T., K.Kinoshita., N.F., F.K., A.S., S.N., K.T., K.Suzuki., Y.O., M.Horikoshi., T.Yamauchi., T.Kadowaki., T.Yoshida. contributed to the data acquisition. A.I., T.M., Y.Hayashi., Y.Hosono., H.E., T. Kohmoto., I.I., M.K. T.Kawaguchi., M.T., and F.M. contributed to the SNP genotyping. S.Kikuchi. and K.M. supervised the study. All authors approved the final version of the manuscript.

Completing financial interests: The authors declare no competing financial interests.

Data availability statement: Summary data will be available from a public data repository.

Restrictions apply to the availability of these data; however, data are available from the authors upon reasonable request.

Reference

1. Vital Statistics (1958-2016). Ministry of Health, Labour and Welfare, Japan. https://ganjoho.jp/reg_stat/statistics/dl/index.html
2. Amundadottir, L. et al. Genome-wide association study identifies variants in the *ABO* locus associated with susceptibility to pancreatic cancer. *Nat. Genet.* **41**, 986–990 (2009).
3. Amundadottir, L. Pancreatic Cancer Genetics. *Int. J. Biol. Sci.* **12**, 314-325 (2016).
4. Fachal, L. & Dunning, A. M. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Curr. Opin. Genet. Dev.* **30**, 32–41 (2015).
5. Schmit, S. L. et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. *J. Natl. Cancer Inst.* (2018). [Epub ahead of print]
6. Childs, E. J. et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat. Genet.* **47**, 911–916 (2015).
7. Park, S. L., Cheng, I. & Haiman, C. A. Genome-Wide Association Studies of Cancer in Diverse Populations. *Cancer Epidemiol. Prev. Biomark.* **27**, 405–417 (2018).
8. Zhang, B. et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat. Genet.* **46**, 533–542 (2014).
9. Cai, Q. et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat. Genet.* **46**, 886–890 (2014).
10. Wu, C. et al. Genome-wide association study identifies five loci associated with

- susceptibility to pancreatic cancer in Chinese populations. *Nat. Genet.* **44**, 62–66 (2012).
11. Low, S-K. et al. Genome-wide association study of pancreatic cancer in Japanese Population. *PLoS One* **5**, e11824 (2010).
 12. Campa, D. et al. Lack of replication of seven pancreatic cancer susceptibility loci identified in two Asian populations. *Cancer Epidemiol. Prev. Biomark.* **22**, 320–323 (2013).
 13. Li, D. et al. Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. *Carcinogenesis* **33**, 1384–1390 (2012)
 14. Klein, A. P., et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat. Commun.* **9**, 556 (2018).
 15. Li, D. Diabetes and pancreatic cancer. *Mol. Carcinog.* **51**:64-74 (2012)
 16. Suzuki, K. et al. Identification of 28 novel susceptibility loci for type 2 diabetes in the Japanese population. *Nat. Genet.* (in press)
 17. 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)
 18. 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).
 19. Wen, W. et al. Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat. Genet.* **44**, 307–311 (2012).
 20. Fukuoka, S., Freedman, S. D., & Scheele, G. A. A single gene encodes membrane-bound and free forms of GP-2, the major glycoprotein in pancreatic secretory (zymogen) granule membranes. *Proc. Natl. Acad. Sci.* **88**, 2898–2902 (1991).
 21. Ohno, H, & Hase, K. Glycoprotein 2 (GP2). *Gut Microbes.* **1**, 407-410 (2010)

22. Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580-585 (2013).
23. Tang, Z., et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic. Acids. Res.* **45**, W98-W102 (2017).
24. Devuyst, O., Olinger, E., & Rampoldi, L. Uromodulin: from physiology to rare and complex kidney disorders. *Nat. Rev. Nephrol.* **13**, 525–544 (2017).
25. Jovine, L., et al. The ZP domain is a conserved module for polymerization of extracellular proteins. *Nat. Cell. Biol.* **4**, 457-462 (2002).
26. Wassarman, P. M. Zona Pellucida Glycoproteins. *J. Biol. Chem.* **283**, 24285–24289 (2008).
27. Yu, S., Michie, S. A. & Lowe, A. W. Absence of the major zymogen granule membrane protein, GP2, does not affect pancreatic morphology or secretion. *J. Biol. Chem.* **279**, 50274–50279 (2004).
28. Tornai, T. et al. Loss of tolerance to gut immunity protein, glycoprotein 2 (GP2) is associated with progressive disease course in primary sclerosing cholangitis. *Sci. Rep.* **8**, 399 (2018).
29. Werner, L. et al. Identification of Pancreatic Glycoprotein 2 as an Endogenous Immunomodulator of Innate and Adaptive Immune Responses. *J. Immunol.* **189**, 2774–2783 (2012).
30. Hase, K. et al. Uptake through glycoprotein 2 of FimH⁺ bacteria by M cells initiates mucosal immune response. *Nature* **462**, 226–230 (2009).
31. Yu, S. & Lowe, A. W. The pancreatic zymogen granule membrane protein, GP2, binds Escherichia coli type 1 Fimbriae. *BMC. Gastroenterol.* **9**, 58 (2009).
32. Geller, L. T. et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* **357**, 1156–1160 (2017).

33. Grote, V.A. et al. Diabetes mellitus, glycated haemoglobin and C-peptide levels in relation to pancreatic cancer risk: a study within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Diabetologia* **54**, 3037-3046 (2011).
34. Lu, Y. et al. New-onset type 2 diabetes, elevated HbA1c, anti-diabetic medications, and risk of pancreatic cancer. *Br. J. Cancer* **113**, 1607-1614 (2015).
35. Carreras-Torres, R. et al. The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study. *J. Natl. Cancer Inst.* **109**(9), djx012 (2017).
36. Karantza, V. Keratins in health and cancer: more than epithelial cell markers. *Oncogen.* **30**, 127-138 (2011).
37. Casanova, M. L. et al. Exocrine pancreatic disorders in transgenic mice expressing human keratin 8. *J. Clin. Invest.* **103**, 1587-1595 (1999).
38. Ku, N. O., Gish, R., Wright, T. L., & Omary, M. B. Keratin 8 mutations in patients with cryptogenic liver disease. *N. Engl. J. Med.* **344**, 1580-1587 (2001).
39. Nakatochi, M. et al. Prediction model for pancreatic cancer risk in the general Japanese population. *PLoS One* **13**, e0203386 (2018).
40. Nagai, A. et al. Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27**, S2-S8 (2017).
41. Pirinen, M., Donnelly, P., & Spencer, C. C. A. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat. Genet.* **44**, 848-851 (2012).
42. Wang, K., et al. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
43. Boyle, A. P., et al. Annotation of functional variation in personal genomes using

- RegulomeDB. *Genome Res.* **22**, 1790-1797 (2012).
44. Burgess, S., Davies, N.M., Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597-608 (2016).
45. Yavorska, O. O., Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734-1739 (2017).

Figure legends

Figure 1. Manhattan plot for the meta-analysis. The horizontal red line represents the genome-wide significance level ($\alpha = 5 \times 10^{-8}$). The horizontal blue line represents the suggestive significance level ($\alpha = 1 \times 10^{-6}$).

Figure 2. Regional association plot for the 16p12.3 locus identified in the meta-analysis. The vertical axis indicates the $-\log_{10}(P$ value) for the assessment of the association of each SNP with pancreatic cancer. The colors indicate the LD (r^2) between each sentinel SNP and neighboring SNPs based on the JPT population in the 1000 Genomes Project Phase 3.

Figure 3. MR analysis with the IVW method for the relationship between T2D and pancreatic cancer in the Japanese population. (a) The results from 82 T2D-associated SNPs (b) 25 HbA1c-associated SNPs.

Table 1. Genomic region and lead SNPs associated with pancreatic cancer susceptibility in the meta-analysis of three Japanese GWASs

SNP	Locus	Chr	Position	Gene	Alleles		OR (95% CI)	P value	I^2	HetP value
					Risk	Non-risk				
Genome-wide significant loci										
rs147905965	13q12.2	13	28474234	<i>PLUT</i>	AT	A	1.29 (1.18-1.41)	1.66×10^{-8}	28.3	0.248
rs9543325	13q22.1	13	73916628	<i>KLF5, LINC00392</i>	C	T	1.24 (1.15-1.33)	1.38×10^{-8}	24.9	0.264
rs78193826	16p12.3	16	20328666	<i>GP2</i>	T	C	1.46 (1.29-1.66)	4.28×10^{-9}	14.5	0.310
Suggestive loci										
rs3737136	1p13.2	1	113060432	<i>WNT2B</i>	G	A	1.24 (1.14-1.34)	1.35×10^{-7}	74.3	0.020
rs35067842	2p12	2	79568318	<i>REG3A, LOC101927987</i>	C	A	1.24 (1.14-1.34)	3.28×10^{-7}	0	0.883
rs6809193	3p12.3	3	76479914	<i>ZNF717, ROBO2</i>	G	A	1.23 (1.14-1.33)	4.33×10^{-7}	34.2	0.219
rs7855466	9q34.2	9	136121303	<i>OBP2B, ABO</i>	T	C	1.22 (1.13-1.31)	2.38×10^{-7}	56.7	0.100

OR values represent the increased risk of pancreatic cancer per risk allele copy for each SNP. Chr, chromosome. RAF, risk allele frequency.

Table 2. Associations between selected SNPs at 16p12.3 and pancreatic cancer risk in the replication and combined analyses

SNP	Chr	Position	Alleles		Replication analysis				Combined analysis			Het value
			Risk	Non-risk	RAF		OR (95% CI)	P value	OR (95% CI)	P value	I^2	
					Case	Control						
rs78193826	16	20328666	T	C	0.091	0.075	1.26 (0.94-1.68)	0.12	1.43 (1.27-1.60)	1.98×10^{-9}	7.7	0.35
rs117267808	16	20323168	A	G	0.091	0.075	1.25 (0.93-1.66)	0.134	1.43 (1.27-1.60)	3.26×10^{-9}	11.3	0.33
rs73541251	16	20331737	G	C	0.091	0.071	1.34 (1.00-1.79)	0.05	1.44 (1.28-1.62)	1.20×10^{-9}	3.8	0.37
rs4632135	16	20337884	C	T	0.092	0.071	1.34 (1.00-1.79)	0.046	1.42 (1.26-1.60)	6.17×10^{-9}	14.7	0.31

OR values represent the increased risk of pancreatic cancer per risk allele copy for each SNP. Chr, chromosome. RAF, risk allele frequency.

Figure 1

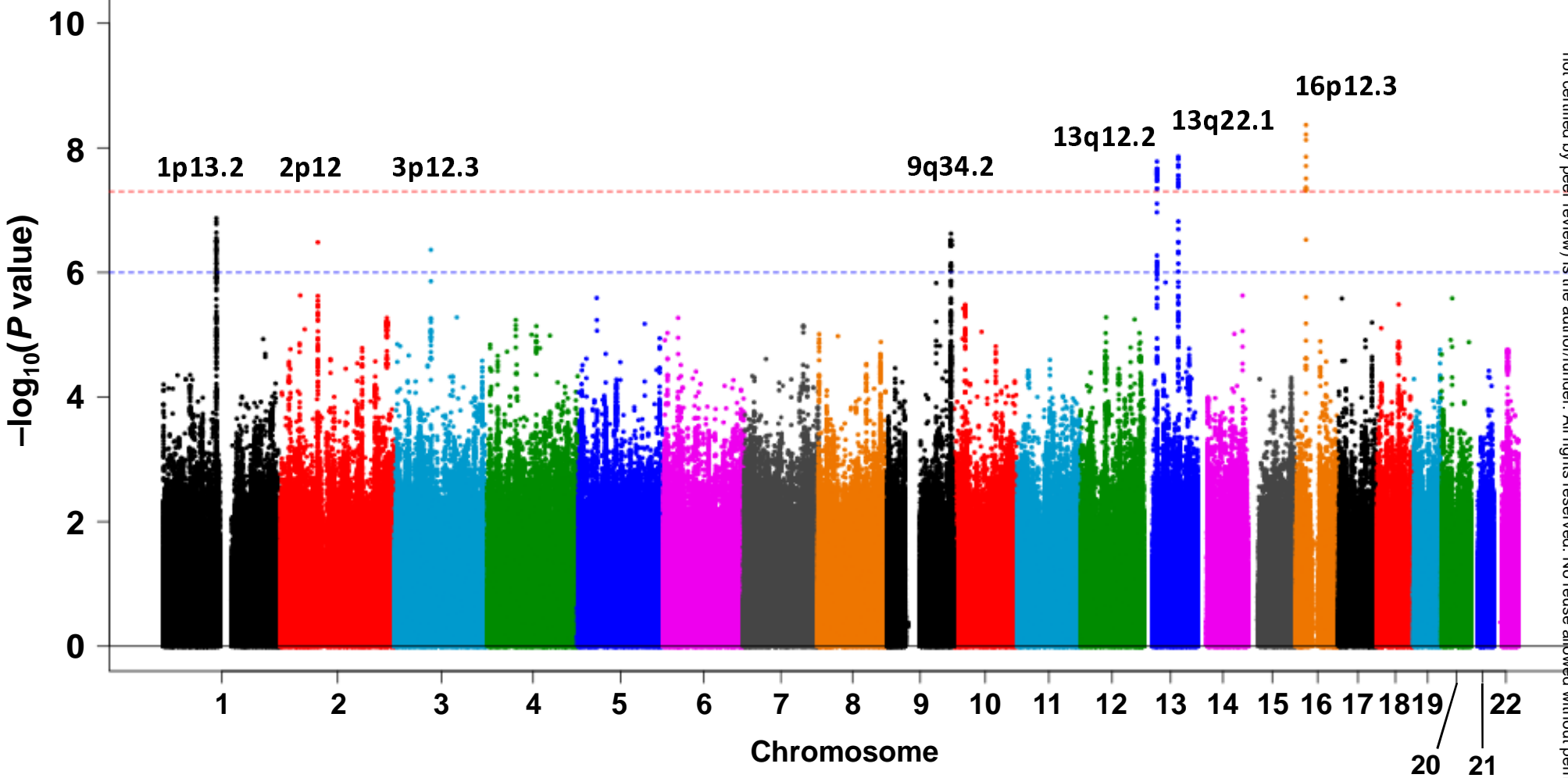


Figure 2

16p12.3

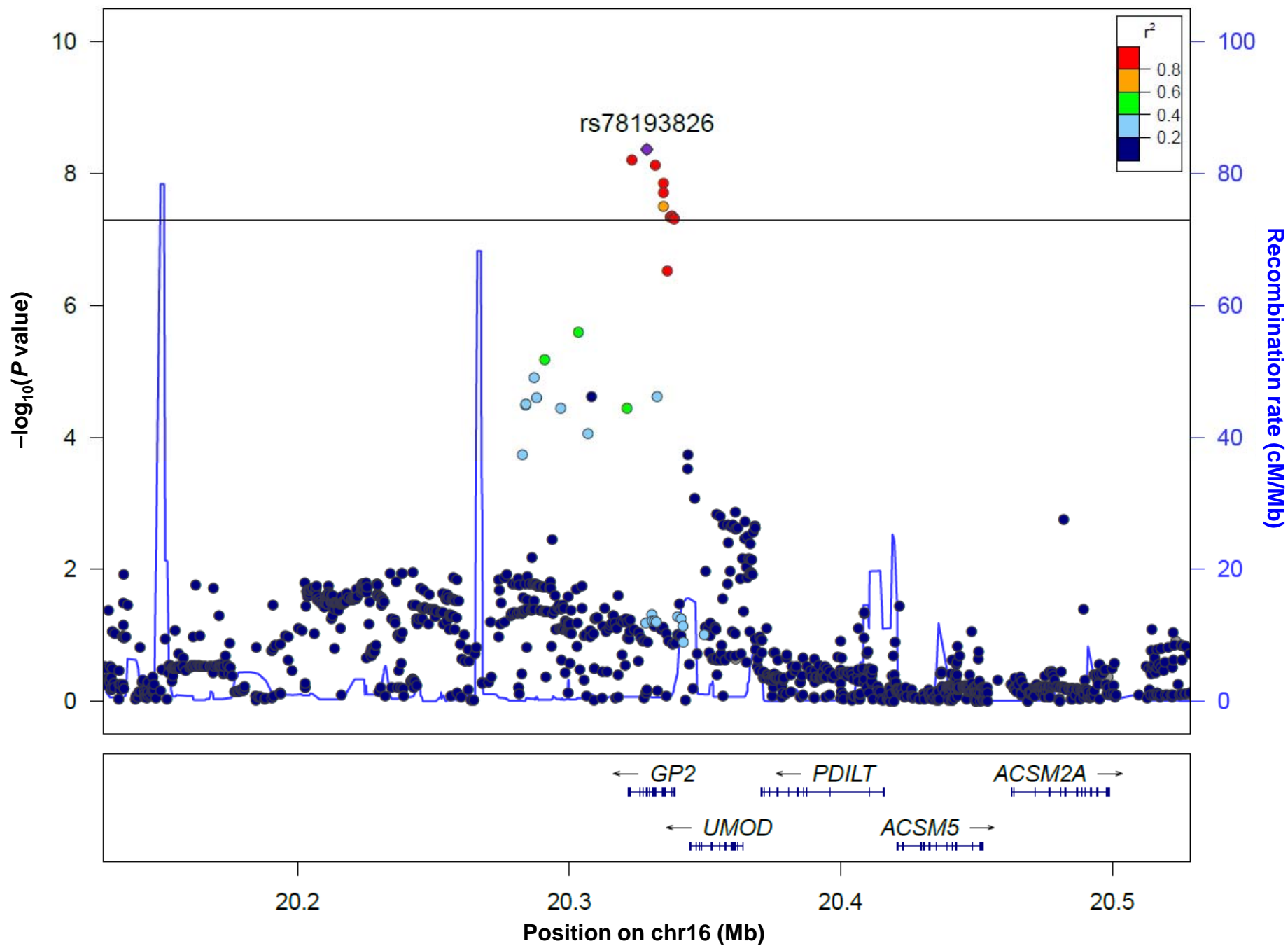


Figure 3a

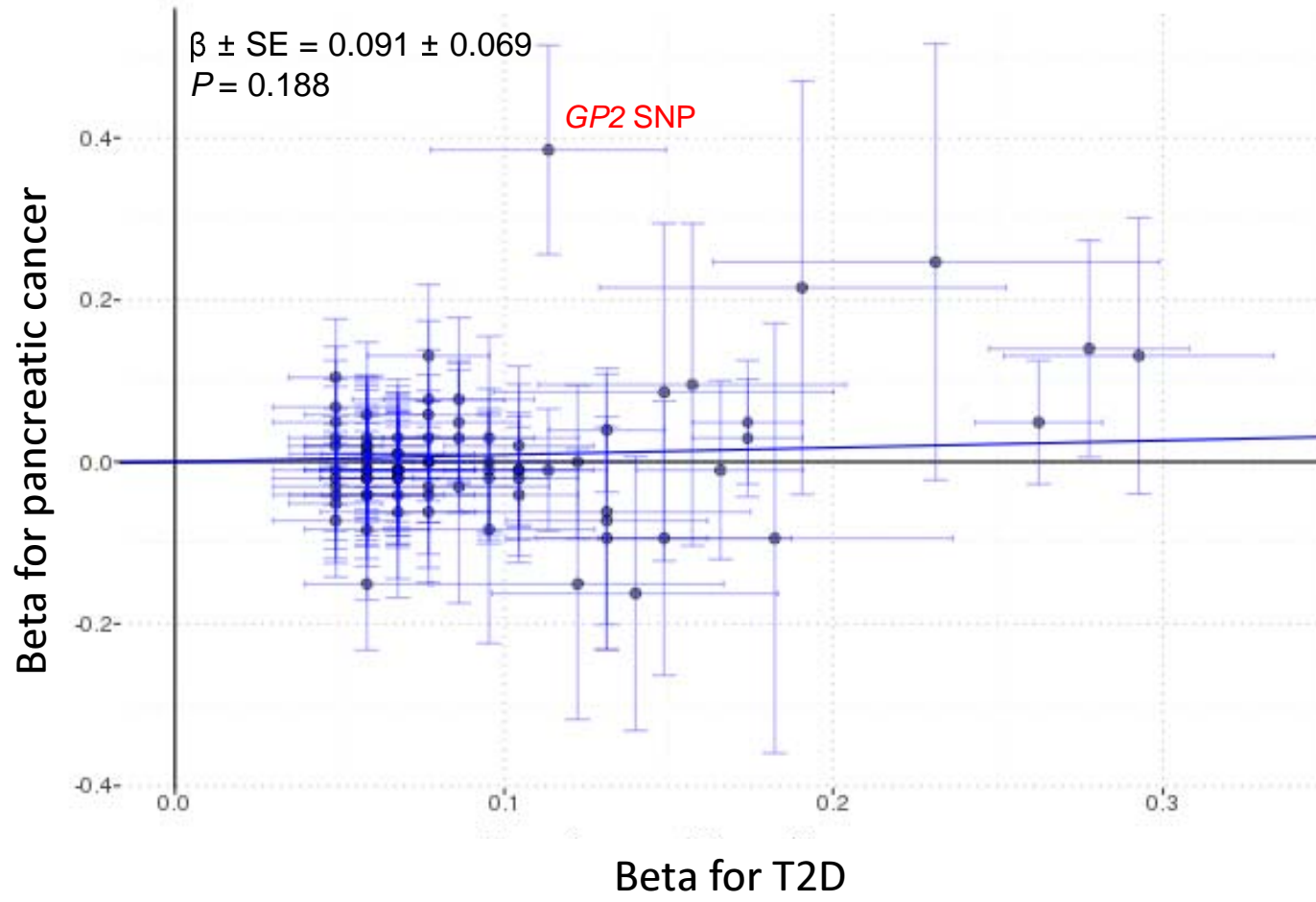


Figure 3b

