Carcinogenesis, 2023, **44**, 642–649
https://doi.org/10.1093/carcin/bgad056
Advance access publication 6 September 2023
Original Article

OXFORD

Original Article

A scan of all coding region variants of the human genome, identifies 13q12.2-rs9579139 and 15q24.1-rs2277598 as novel risk loci for pancreatic ductal adenocarcinoma

Matteo Giaccherini[1,#], Leonardo Gori[1,#], Manuel Gentiluomo[1], Riccardo Farinella[1], Klara Cervena[2,3], Jurgita Skieceviciene[4], Frederike Dijk[5], Gabriele Capurso[6,7], Antonis Vezakis[8], Livia Archibugi[6,7], Roger Chammas[9], Tamás Hussein[10,11], Francesca Tavano[12], Péter Hegyi[10,11,13], Martin Lovecek[14], Jakob R. Izbicki[15], Hermann Brenner[16,17,18], Beatrice Mohelnikova-Duchonova[19], Giuseppe Dell'Anna[20], Juozas Kupcinskas[4], Stefano Ermini[21], Mateus Nóbrega Aoki[22], John P. Neoptolemos[23], Maria Gazouli[24], Claudio Pasquali[25], Raffaele Pezzilli[26], Renata Talar-Wojnarowska[27], Martin Oliverius[28], Mohammed Al-Saeedi[23], Maurizio Lucchesi[29], Niccolò Furbetta[30], Silvia Carrara[31], Casper H.J. van Eijck[32], Almantas Maleckas[33], Anna Caterina Milanetto[25], Rita T. Lawlor[34], Ben Schöttker[16], Ugo Boggi[35], Luca Morelli[30], Laura Ginocchi[29], Ruggero Ponz de Leon Pisani[20], Cosimo Sperti[25], Alessandro Zerbi[36,37], Paolo Giorgio Arcidiacono[20], Faik G. Uzunoglu[15], Stefania Bunduc[10,38,39], Bernd Holleczek[40], Domenica Gioffreda[12], Ewa Małecka-Wojciesko[27], Mindaugas Kiudelis[33], Andrea Szentesi[13,41], Hanneke W.M. van Laarhoven[42], Pavel Soucek[43], Mara Götz[15], Bálint Erőss[10,11,13], Giulia Martina Cavestro[44], Daniela Basso[25], Francesco Perri[12], Stefano Landi[1], Federico Canzian[45] and Daniele Campa[1,*]

[1]Department of Biology, University of Pisa, Pisa, Italy
[2]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic
[3]First Faculty of Medicine, Institute of Biology and Medical Genetics, Charles University, Prague, Czech Republic
[4]Department of Gastroenterology, Institute for Digestive Research, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania
[5]Department of Pathology, Cancer Center Amsterdam, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands
[6]Digestive and Liver Disease Unit, S. Andrea Hospital, "Sapienza" University of Rome, Rome, Italy
[7]Pancreato-Biliary Endoscopy and Endosonography Division, IRCCS San Raffaele Scientific Institute, Pancreas Translational and Clinical Research Center, Vita-Salute San Raffaele University, Milan, Italy
[8]Department of Surgery, Aretaieio Hospital, Medical School, National and Kapodistrian University of Athens, Athens 11528, Greece
[9]Departamento de Radiologia e Oncologia, Instituto Do Câncer Do Estado de São Paulo (ICESP), Center for Translational Research in Oncology (LIM24), Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HCFMUSP), São Paulo, Brazil
[10]Center for Translational Medicine, Semmelweis University, Budapest, Hungary
[11]Division of Pancreatic Diseases, Heart and Vascular Center, Semmelweis University, Budapest, Hungary
[12]Division of Gastroenterology and Research Laboratory, Fondazione IRCCS "Casa Sollievo della Sofferenza" Hospital, San Giovanni Rotondo, Foggia, Italy
[13]Institute for Translational Medicine, Medical School, University of Pécs, Pécs, Hungary
[14]Department of Surgery I, University Hospital Olomouc, Olomouc, Czech Republic
[15]Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[16]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[17]Division of Preventive Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[18]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[19]Department of Oncology and Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University and University Hospital, Olomouc, Czech
[20]Pancreatico/Biliary Endoscopy and Endosonography Division, Pancreas Translational and Clinical Research Center, San Raffaele Scientific Institute, Milan, Italy

Received: February 16, 2023; Revised: June 8 2023; Accepted: August 28, 2023

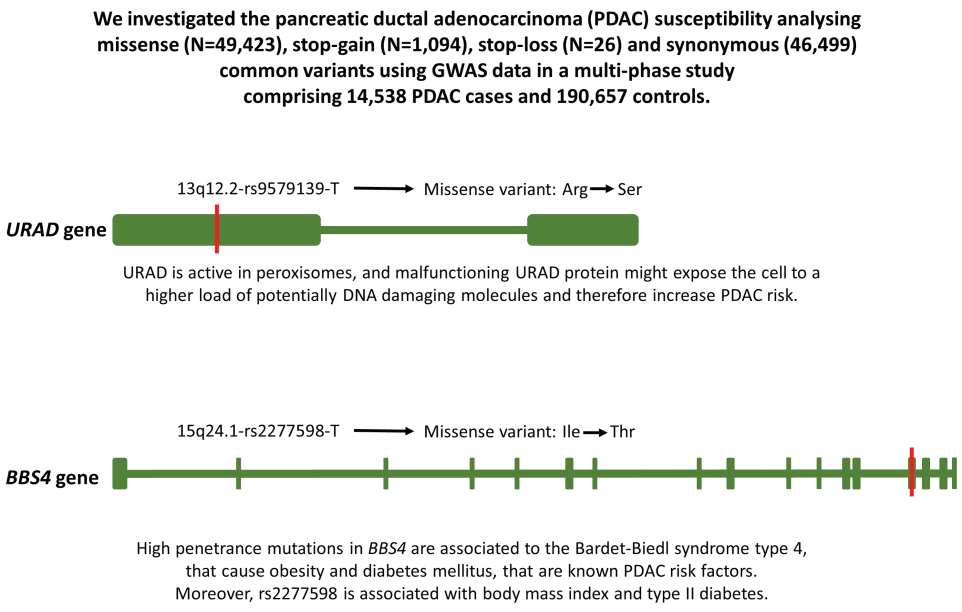
[21]Blood Transfusion Service, Azienda Ospedaliero-Universitaria Meyer, Children's Hospital, Florence, Italy
[22]Laboratory for Applied Science and Technology in Health, Carlos Chagas Institute, Oswaldo Cruz Foundation (Fiocruz), Curitiba, Brazil
[23]Department of General Surgery, University of Heidelberg, Heidelberg, Germany
[24]Department of Basic Medical Sciences, Laboratory of Biology, Medical School, National and Kapodistrian University of Athens, Athens, Greece
[25]Department of Surgery, Oncology and Gastroenterology-DiSCOG, University of Padova, Padua, Italy
[26]Potenza County Medical Association, Potenza, Italy
[27]Department of Digestive Tract Diseases, Medical University of Lodz, Lodz, Poland
[28]Department of Surgery, University Hospital Kralovske Vinohrady, Third Faculty of Medicine, Charles University, Prague, Czech Republic
[29]Department of Medical Oncology, Oncology of Massa Carrara, Azienda USL Toscana Nord Ovest, Carrara, Italy
[30]General Surgery Unit, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
[31]Endoscopic Unit, Department of Gastroenterology, IRCCS Humanitas Research, Milan, Italy
[32]Department of Surgery, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[33]Department of Surgery, Institute for Digestive Research, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania
[34]ARC-Net Centre for Applied Research on Cancer and Department of Diagnostics and Public Health, University of Verona, Verona, Italy
[35]Division of General and Transplant Surgery, Pisa University Hospital, Pisa, Italy
[36]Pancreatic Unit, IRCCS Humanitas Research Hospital, Milan, Italy
[37]Department of Biomedical Sciences, Humanitas University, Milan, Italy
[38]Carol Davila University of Medicine and Pharmacy, Bucharest, Romania
[39]Center for Gastroenterology, Hepatology and Liver Transplant, Fundeni Clinical Institute, Bucharest, Romania
[40]Saarland Cancer Registry, Saarbrücken, Germany
[41]Department of Medicine, Centre for Translational Medicine, University of Szeged, Szeged, Hungary
[42]Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands
[43]Biomedical Center, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czech Republic
[44]Gastroenterology and Gastrointestinal Endoscopy Unit, Vita-Salute San Raffaele University, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy
[45]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany

[#]These authors share the first position
[*]Corresponding author: Department of Biology, University of Pisa, Via Derna 1, 56126 Pisa (PI), Italy. Tel: +39-050 2211510 Email: daniele.campa@unipi.it

Abstract

Coding sequence variants comprise a small fraction of the germline genetic variability of the human genome. However, they often cause deleterious change in protein function and are therefore associated with pathogenic phenotypes. To identify novel pancreatic ductal adenocarcinoma (PDAC) risk loci, we carried out a complete scan of all common missense and synonymous SNPs and analysed them in a case–control study comprising four different populations, for a total of 14 538 PDAC cases and 190 657 controls. We observed a statistically significant association between 13q12.2-rs9581957-T and PDAC risk ($P = 2.46 \times 10^{-9}$), that is in linkage disequilibrium (LD) with a deleterious missense variant (rs9579139) of the *URAD* gene. Recent findings suggest that this gene is active in peroxisomes. Considering that peroxisomes have a key role as molecular scavengers, especially in eliminating reactive oxygen species, a malfunctioning URAD protein might expose the cell to a higher load of potentially DNA damaging molecules and therefore increase PDAC risk. The association was observed in individuals of European and Asian ethnicity. We also observed the association of the missense variant 15q24.1-rs2277598-T, that belongs to *BBS4* gene, with increased PDAC risk ($P = 1.53 \times 10^{-6}$). rs2277598 is associated with body mass index and is in LD with diabetes susceptibility loci. In conclusion, we identified two missense variants associated with the risk of developing PDAC independently from the ethnicity highlighting the importance of conducting re-analysis of genome-wide association studies (GWASs) in light of functional data.

Graphical Abstract



Abbreviations: BBJ, BioBank Japan; eQTL, expression quantitative trait loci; GWASs, genome-wide association studies; JaPAN, Japan Pancreatic Cancer Research; LD, linkage disequilibrium; NCC, National Cancer Center; PanC4, Pancreatic Cancer Case-Control Consortium; PCA, Principal component analysis; PDAC, pancreatic ductal adenocarcinoma; pQTL, protein quantitative trait loci; SNPs, single nucleotide polymorphisms.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) represents the majority of all pancreatic cancers. It is the 7th cause of cancer related deaths and it has been estimated to become the second by 2030 (1,2). Although the survival of the patients has increased in the last years, PDAC is characterized by comparable incidence and mortality rates, with a 5-year survival after diagnosis around 9% (3). The low survival rate is due to the difficult diagnosis at early stages, resulting in only one fifth of the patients amenable to surgical treatment (4,5). To date, a handful of high penetrance mutations have been associated with increased PDAC risk through family-based studies (3,6). Moreover, various common low-risk variants have been identified, individually or grouped in polygenic risk scores through candidate region and genome-wide association studies (GWASs) (6–18). However, the number of single nucleotide polymorphisms (SNPs) associated with PDAC susceptibility is considerably lower compared to other common solid tumours and a large proportion of the genetic heritability for this disease remains to be identified. Although GWASs have been successful in identifying susceptibility variants, their functional involvement in PDAC development remains elusive. In addition, GWAS are also prone to false negatives since only the top findings ($P < 5 \times 10^{-8}$) are usually reported. A possible solution to overcome these limitations, and identify new risk variants, is to perform secondary analysis of GWAS data (i.e. re-analysing a list of SNPs with a higher a priori probability to be associated because, for example, of their functionality) followed by a replication in one or more independent case–control sets. This strategy has been successfully used to identify functional variants, especially regulatory ones, associated with PDAC risk (18–22). A particularly interesting class of SNPs are those that affect the sequence and the biochemical properties of the encoded protein (i.e. missense and stop-gain or stop-loss) or that may alter the codon usage (synonymous variants). Missense and truncating variants are associated, for example, with breast (23,24), gastric (25), prostate (26), and ovarian cancer risk (27). Additionally, several common missense variants have also been reported to be associated with multiple cancers (28–30).

Common synonymous germline variants have also been associated with several tumours. One of the most studied, rs1045642 in the *ABCB1* gene, is associated with many human phenotypes including risk of several cancer types (31).

With these premises, the aim of this study was to identify novel PDAC risk loci by analysing all missense, stop-gain/stop-loss and synonymous SNPs in the human genome using 14 538 PDAC cases and 190 657 controls.

Materials and methods

The study was designed in a discovery phase, in which the SNPs of interest were selected and analysed in four studies consisting of 11 296 cases and 186 908 controls and in a replication phase, where the SNPs that showed a statistically significant association in at least three or the four discovery datasets were genotyped and analysed in an independent case–control set consisting of 3242 PDAC cases and 3749 controls belonging to the PANDoRA consortium.

Discovery phase

A complete list of all common (minor allele frequency, Global MAF > 1% in the 1000 Genomes project) germline missense ($N = 49\ 423$), stop-gain ($N = 1094$), stop-loss ($N = 26$) and synonymous (46 499) SNPs, was compiled using the NCBI Single Nucleotide Polymorphism (dbSNP) public database. The selected SNPs were analysed in the discovery phase using four datasets: the Pancreatic Cancer Cohort Consortium (PanScan I, II, III) the Pancreatic Cancer Case-Control Consortium (PanC4) GWASs, the summary statistics of a 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 (this dataset will be referred to as JaPAN), and the FinnGen study]. The genotypes of PanScan and PanC4 were downloaded from the database of Genotypes and Phenotypes (dbGaP; study accession nos. phs000206.v5.p3 and phs000648.v1.p1; project reference # 12644). The summary statistics of the JaPAN consortium and the FinnGen study are available at www.aichi-med-u.ac.jp/JaPAN and www.finngen.fi respectively.

Genotyping and quality control details of PanScan and PanC4 have been described in the original publications (7,11,15,16). More details on data filtering and quality control procedures are given in the original publications (13) or in the respective websites (www.aichi-med-u.ac.jp/JaPAN and www.finngen.fi). The genotypes of PanScan and PanC4 were imputed using the Michigan Imputation Server (https://imputationserver.sph.umich.edu), and the Haplotype Reference Consortium (HRC, V.r1.1) as reference panel.

Before imputation, the following quality control filters were applied to the datasets: removal of individuals with gender mismatches, call rate <0.98 and minimal or excessive heterozygosity (>3 SDs from the mean). Additionally, the SNPs with a MAF < 0.01, call-rate < 98%, cryptic relatedness (PI_HAT > 0.2), low-quality imputation score (information score <0.7), and evidence for violations of Hardy-Weinberg equilibrium ($P < 1 \times 10^{-5}$) were discarded. Principal component analysis (PCA) was performed with PLINK 2.0, including the genotypes of phase 3 of the 1000 Genomes Project as reference panel (32). Individuals not clustering in the PCA with the 1000 Genomes subjects of European descent were excluded from further analysis. After QCs the discovery dataset consisted of 4857 PDAC cases and 3418 controls for PanScan and 3881 PDAC cases and 3616 controls for PanC4 (Table 1) which were analysed for 7 509 345 SNPs. The 'inflation factor' calculated in each dataset, did not show evidence of systematic inflation ($\lambda = 1.000$ for PanScan $\lambda = 1.000$ for PanC4, and $\lambda = 1.000$ for the aggregate dataset).

The available summary statistics of the JaPAN consortium and the FinnGen studies were obtained from the analysis of 2039 PDAC cases and 32 592 controls for JaPAN, and 519 PDAC cases and 147 282 controls for FinnGen (R5 release) (Table 1). The discovery phase consisted therefore, of a total of 11 296 PDAC cases and 186 908 controls.

Replication phase

In the replication phase, the SNPs that showed an association with PDAC risk ($P < 0.05$) in at least three datasets, that did

Table 1. Details of study populations

	Cases	Controls
Discovery phase		
PanScan I–II–III	4857	3418
PanC4	3881	3616
JaPAN	2039	32 592
FinnGen	519	147 282
Replication phase		
PANDoRA	3242	3749
Brazil	69	259
Czech Republic	382	176
Germany	452	1131
Greece	115	16
Hungary	319	367
Italy	1448	1520
Lithuania	249	181
Netherlands	154	62
Poland	54	37
Sex		
Female	1491 (46%)	1722 (46%)
Male	1752 (54%)	2027 (54%)
Median age (25–75%)	65.1 (58–73)	57.8 (50–67)
Combined analysis		
Total	14 538	190 657

not show evidence of heterogeneity ($P_{Het} > 0.05$), that were independent to SNPs already known to be associated with PDAC risk through GWAS ($r^2 < 0.8$ in Europeans) and that had a statistical power to be replicated higher than 0.8F were genotyped in 3242 PDAC cases and 3749 from PANDoRA, which has been previously described in detail (33,34). Briefly, pancreatic cancer cases and controls, with information on sex, age of diagnosis for the cases and age of recruitment for the controls, were collected from ten European countries (Greece, Italy, Germany, Netherlands, Denmark, Czech Republic, Hungary, Poland, Lithuania, United Kingdom) and Brazil. The controls were enrolled among the general population, blood donors or hospitalized individuals not affected by cancer (Table 1). PANDoRA individuals were genotyped using TaqMan technology, using 384-well plates in which negative controls and duplicate samples (approximately 8%) were included for quality control purpose. QuantStudio™ 5 Real-Time PCR system (Thermofisher, USA) was used to determine the genotypes. The intronic variant *URAD*-rs9581957-T, was selected as a proxy for the missense variant rs9579139 ($r^2 = 0.97$, $D' = 0.99$) since the probe was not available as TaqMan assay.

Statistical analysis

The number of available SNPs to be analysed in the discovery phase (i.e. SNPs for which genotyping, or imputation data was available) were 45 200 for PanScan and PanC4 23 979 SNPs for JaPAN and 44 378 SNPs for FinnGen. A logistic regression, adjusting for sex, age, and the first eight principal components was used in the discovery phase. In the replication phase, the selected SNPs were analysed using logistic regression adjusting for sex, age (at recruitment for controls, at diagnosis for cases) and country of origin. Finally, a meta-analysis considering all the populations together was performed. To calculate the threshold for statistical significance corrected for multiple testing we computed the number of independent SNPs ($r^2 < 0.8$) that was of 13 164, resulting in $P = 3.80 \times 10^{-6}$ (0.05/13 164).

Functional characterization

All SNPs that showed a statistically significant association after multiple testing correction were investigated for their effect on the protein using PolyPhen-2 (35) and SIFT (36). These two tools predict the possible impact on the structure and function of a human protein due to an amino acidic substitution caused by allelic change of missense SNPs. In addition, data from the Genotype-Tissue Expression (GTEx) project (37) were used to analyse the SNPs in relation to gene expression to determine if they are expression quantitative trait loci (eQTL). Open Target Genetics data were used to collect data on protein quantitative trait loci (pQTL) (38).

Results

After analysing the discovery set eight SNPs fulfilled the criteria described in the material and methods and were genotyped in PANDoRA.

The association analysis of these eight SNPs in PanScan, PanC4, JaPAN, and FinnGen is reported in Table 2. In PANDoRA an association between VIPR2-rs3793232-C and an increased PDAC risk (OR = 1.13, 95% CI = 1.04–1.23, $P = 3.19 \times 10^{-3}$), and *URAD*-rs9581957-T and decreased risk (OR = 0.88, 95% CI = 0.80–0.98, $P = 0.017$) was observed (Table 2).

The meta-analysis of discovery and replication phase showed two associations with a *P*-value below the significance threshold corrected for multiple testing ($P = 3.80 \times 10^{-6}$). The missense variant rs9579139, analysed in the replication phase using its proxy *URAD*-rs9581957-T ($r^2 = 0.97$, $D' = 0.99$), was associated with reduction of PDAC risk (OR = 0.88, 95% CI = 0.85–0.92, $P = 2.46 \times 10^{-9}$), and the missense variant *BBS4*-rs2277598-T was associated with an increase PDAC risk (OR = 1.08, 95% CI = 1.05–1.12, $P = 1.53 \times 10^{-6}$) (Table 2). The forest plots obtained from the combined datasets (PanScan, PanC4, JaPAN, FinnGen and PANDoRA) are reported in Figure 1. PolyPhen-2 classifies rs9579139 as possibly damaging and rs2277598 as benign, while SIFT classifies rs9579139 as deleterious and rs2277598 as tolerated. The two SNPs are not associated with an altered gene expression and protein levels in pancreatic tissues according to the GTEx and Open Target Genetics databases. A visual representation of the regions around rs9579139 and rs2277598, using LocusZoom and divided by ancestry, is reported in Supplementary Figures 1 and 2.

Discussion

In this study, all common (MAF > 1%) germline missense and synonymous SNPs were analysed in relation to PDAC risk in a multi-phase study consisting of 14 538 PDAC cases and 190 657 controls.

The most statistically significant association was observed for 13q12.2-rs9581957 ($P = 2.46 \times 10^{-9}$). This SNP is an intronic variant of the *URAD* gene, and it was selected

Table 2. Results of discovery (PanScan, PanC4, JaPAN and FinnGen), replication phase (PANDoRA) and the combined analysis conducted using all the result of previous phases

SNP	rs3793232	rs3736583	rs2041028	rs9581957[c]	rs2277598	rs313841	rs7250850	rs2232079
Gene	VIPR2	PSTK	OTOG	URAD	BBS4	STRN4	ZC3H4	FERMT1
Locus	7q36.3	10q26.13	11p15.1	13q12.2	15q24.1	19q13.32	19q13.32	20p12.3
Function	Missense	Synonymous	Missense	Intron	Missense	Synonymous	Synonymous	Synonymous
M/m[a]	T/C	A/G	C/T	C/T	C/T	G/A	C/G	C/T
MAF[b]	0.27	0.27	0.38	0.31	0.34	0.14	0.31	0.11
PanScan								
OR (95% CI)	0.98 (0.91–1.05)	0.90 (0.84–0.96)	1.07 (1.01–1.15)	0.94 (0.88–1.01)	1.10 (1.03–1.18)	0.89 (0.81–0.97)	1.08 (1.01–1.16)	0.87 (0.79–0.97)
P value	0.535	2.64×10^{-3}	0.032	0.114	2.93×10^{-3}	0.011	0.022	0.013
PanC4								
OR (95% CI)	1.12 (1.04–1.20)	0.92 (0.86–0.99)	1.07 (1.00–1.15)	0.86 (0.80–0.93)	1.09 (1.02–1.17)	0.89 (0.81–0.98)	1.08 (1.00–1.16)	0.89 (0.80–0.99)
P value	3.62×10^{-3}	0.022	0.043	6.82×10^{-5}	0.013	0.020	0.040	0.029
JaPAN								
OR (95% CI)	1.13 (1.05–1.23)	0.92 (0.85–0.98)	1.11 (1.00–1.23)	0.82 (0.74–0.91)	1.03 (0.94–1.13)	1.10 (0.90–1.34)	1.12 (1.01–1.25)	0.83 (0.70–0.99)
P value	1.66×10^{-3}	0.017	0.047	1.55×10^{-4}	0.523	0.338	0.031	0.039
FinnGen								
OR (95% CI)	1.15 (1.01–1.31)	0.85 (0.74–0.97)	1.06 (0.93–1.21)	0.88 (0.77–1.00)	1.15 (1.01–1.31)	0.81 (0.68–0.96)	0.98 (0.86–1.12)	1.01 (0.82–1.25)
P value	0.038	0.016	0.350	0.045	0.031	0.017	0.802	0.918
PANDoRA								
OR (95% CI)	1.13 (1.04–1.23)	1.03 (0.94–1.13)	1.06 (0.98–1.15)	0.88 (0.80–0.98)	1.05 (0.97–1.13)	1.07 (0.89–1.28)	1.06 (0.97–1.15)	0.91 (0.75–1.10)
P value	3.19×10^{-3}	0.495	0.16	0.017	0.236	0.473	0.227	0.321
Meta-analysis								
OR (95% CI)	1.10 (1.03–1.17)	0.92 (0.89–0.96)	1.07 (1.04–1.11)	0.88 (0.85–0.92)	1.08 (1.05–1.12)	0.91 (0.86–0.97)	1.07 (1.03–1.12)	0.89 (0.84–0.94)
P value	3.30×10^{-3}	1.40×10^{-5}	3.36×10^{-5}	2.46×10^{-9}	1.53×10^{-6}	1.82×10^{-3}	5.56×10^{-4}	7.02×10^{-5}
P value Het.	0.042	0.107	0.977	0.317	0.570	0.054	0.594	0.715

The following individuals were removed from the PANDoRA analysis because the controls population were not in HWE ($P < 0.05$): individuals from Czech Republic for rs2041028, rs2232079, rs313841 and rs9581957; individuals from Germany for rs3736583, rs7250850 and rs9581957, individuals from Hungary for rs9581957; individuals from Italy for rs2232079 and rs313841; individuals from Poland for rs313841 and rs3736583.

[a]M: major allele; m: minor allele.

[b]MAF: minor allele frequency in Europeans from 1000 Genomes.

[c]The TaqMan probe was not available for rs9579139, and the genotyping was performed using rs9581957 as proxy ($r^2 = 0.97$, $D' = 0.99$).

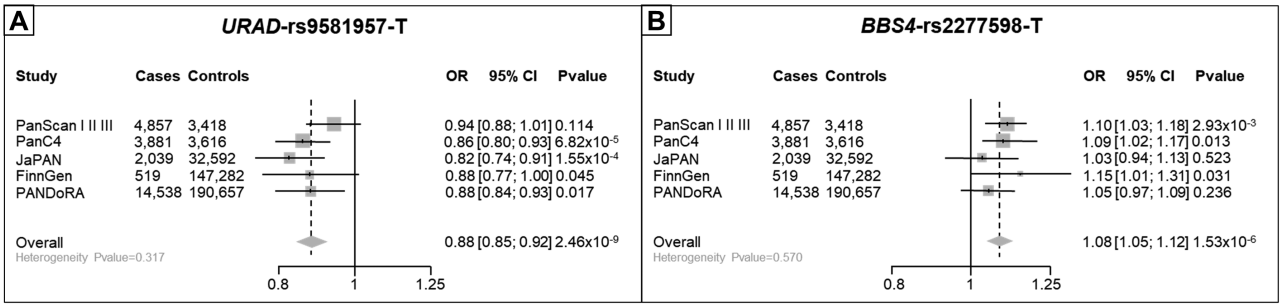


Figure 1. Forest plot of *URAD*-rs95959595-T (**A**) and *BBS4*-rs2277598-T (**B**).

as proxy ($r^2 = 0.97$, $D' = 0.99$) for the missense variant *URAD*-rs9579139. The minor allele (T) of 13q12.2-rs9579139 determines an aminoacidic change from the polar positively charged Arginine to the polar non charged Serine at position 114. PolyPhen classifies this change as possibly damaging and SIFT as deleterious. The URAD protein is a decarboxylase involved in the purine metabolism. However, the specific role of this protein in humans is uncertain (39). In mammals, the decarboxylase activity of the *URAD* gene is limited to the stereoselective decarboxylation of 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) to (S)-allantoin, which does not occur in humans. However, the *URAD* gene expression has recently been detected in several organs of the digestive system. The fact that the gene is active in humans suggests that it has another function that is still unknown. In fact, very recently, the Alliance of Genome Resources project suggested that the URAD protein is active in peroxisomes (39). Considering that peroxisomes have a key role as molecular scavengers, especially in eliminating reactive oxygen species, it is tempting to speculate that a malfunctioning URAD protein might expose the cell to a higher load of potentially DNA damaging molecules and therefore increase PDAC risk.

Another possible explanation for the association of 13q12.2-rs9581957 is that it is in LD ($r^2 = 0.71$) with rs9579135, a genetic locus of glycate haemoglobin level in blood that is a marker for diabetes (T2D), which in turn is an establish PDAC risk factor. Additionally, 13q12.2-rs9581957 is also in LD ($r^2 = 0.61$) with rs4581570, a T2D genetic risk locus (40,41).

We also observed an association of 15q24.1-rs2277598-T ($P = 1.53 \times 10^{-6}$) with increased PDAC risk. This SNP is a missense variant of Bardet-Biedl syndrome 4 (*BBS4*) gene, and the T allele is associated to the aminoacidic change from isoleucine to threonine which is classified as benign by Polyphen and tolerated by SIFT. *BBS4* is ubiquitously expressed in all tissues, and it is involved in intracellular trafficking via microtubule-related transport. High penetrance mutations in this gene are associated to the Bardet-Biedl syndrome type 4, that has a heterogeneous plethora of clinical manifestation and disorders (42), among which obesity and diabetes mellitus. 15q24.1-rs2277598 is also associated with body mass index (BMI) (43), while SNPs in LD with it are associated with various traits correlated to T2D and BMI (44–48).

It is, therefore, possible to speculate that while high penetrance mutations in this gene are causative of BBS4 syndrome, low penetrance SNPs contribute to complex and related traits, such as BMI, diabetes and PDAC. Given the relevance of BMI and diabetes as risk factors for PDAC occurrence, this finding deserve attention.

A clear strength of this study, besides being the largest investigation on missense and synonymous variants and PDAC risk, is represented by the multiple ethnicities analysed, considering that is uncommon for a SNP to be associated across multiple ethnic groups, especially in PDAC. A possible limitation of this study is that many SNPs that were identified as missense, synonymous, stop-gain and stop-loss, could not be analysed in the discovery phase since they were not genotyped or imputed in the arrays. Additionally, we focussed only on common SNPs since we had no power to investigate rarer variants, even though we used the largest possible genetic dataset on PDAC.

In conclusion, in the present study we identified two novel missense variants associated with PDAC risk in populations of different ethnicity, with a plausible function related with increased risk of developing cancer.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Funding

This work was supported by: Fondazione Arpa (to Daniele Campa); Fondazione Tizzi (to Daniele Campa); Ministry of Health of the Czech Republic NV19-08-00113 (to Pavel Soucek), NV19-03-00097 (to Beatrice Mohelnikova-Duchonova), AZVNU21-07-00247 (to Klara Cervena), NV19-03-00096 (to Martin Lovecek); Palacky University Olomuc IGA-LF-2022-003 (to Martin Lovecek); Italian Minister of Health, Ricerca Corrente program 2022-2024, to the Division of Gastroenterology—Fondazione IRCCS 'Casa Sollievo della Sofferenza' Hospital, San Giovanni Rotondo (FG). Gabriele Capurso received support from AIRC under IG 2021—ID.26201 project.

Acknowledgements

This article is based upon work from COST Action 'Identification of biological markers for prevention and translational medicine in pancreatic cancer (TRANSPAN)', CA21116, supported by COST (European Cooperation in Science and Technology). We acknowledge the contribution of the late Dr. Bas Bueno-de-Mesquita (National Institute for Public Health and the Environment, Bilthoven, the Netherlands). We would like to thank Prof. Vermeulen R.C.H. (University of Utrecht) for the EPIC genotyping data used in the PANDoRA replication. The biosamples from the Humanitas Research Hospital were obtained from the Center for Biological Resources.

Conflict of Interest Statement: The authors have no conflicts of interest to report.

Ethical approval

Each participating study obtained approval from the responsible institutional review board (IRB) and IRB certification permitting data sharing in accordance with the NIH Policy for sharing of Data Obtained in NIH-Supported or NIH-Conducted Genome Wide Association Studies. The PANDoRA study protocol was approved by the Ethics Commission of the Medical Faculty of the University of Heidelberg. In accordance with the Declaration of Helsinki, written informed consent was obtained from each participant. The FinnGen study was approved by the ethical Review Board of the Hospital District of Helsinki and Uusimaa. FinnGen participants provided written, informed consent. For JaPAN, 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.

Data availability

The PanScan and PanC4 genotyping data are available from the database of Genotypes and Phenotypes (dbGaP, study accession numbers phs000206.v5.p3 and phs000648.v1.p1). JaPAN data are available from the JaPAN consortium website (www.aichi-med-u.ac.jp/JaPAN). FinnGen summary statistics are available from the FinnGen study website (www.finngen.fi). The PANDoRA primary data for this work will be made available to researchers who submit a reasonable request to the corresponding author, conditional to approval by the PANDoRA Steering Committee and Ethics Commission of the Medical Faculty of the University of Heidelberg. Data will be stripped from all information allowing identification of study participants.

References

1. Ferlay, J. et al. (2021) Cancer statistics for the year 2020: an overview. *Int. J. Cancer*, 149, 778–789. doi:10.1002/ijc.33588
2. Rahib, L. et al. (2014) Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.*, 74, 2913–2921.
3. Klein, A.P. (2021) Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. *Nat. Rev.*, 18, 493–502.
4. Sakaguchi, T. et al. (2020) The past, present, and future status of multimodality treatment for resectable/borderline resectable pancreatic ductal adenocarcinoma. *Surg. Today*, 50, 335–343.
5. Huang, L. et al. (2019) Resection of pancreatic cancer in Europe and USA: an international large-scale study highlighting large variations. *Gut*, 68, 130–139.
6. Gentiluomo, M. et al. (2022) Germline genetic variability in pancreatic cancer risk and prognosis. *Semin. Cancer Biol.*, 79, 105–131.
7. Amundadottir, L. et al. (2009) Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat. Genet.*, 41, 986–990.
8. Galeotti, A.A. et al. (2021) Polygenic and multifactorial scores for pancreatic ductal adenocarcinoma risk prediction. *J. Med. Genet.*, 58, 369–377.
9. Campa, D. et al. (2015) TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. *Int. J. Cancer*, 137, 2175–2183.
10. Campa, D. et al. (2020) Genome-wide association study identifies an early onset pancreatic cancer risk locus. *Int. J. Cancer*, 147, 2065–2074.
11. Childs, E.J. et al. (2015) Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat. Genet.*, 47, 911–916.
12. Klein, A.P. et al. (2018) Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat. Commun.*, 9, 1–11.
13. Lin, Y. et al. (2020) Genome-wide association meta-analysis identifies GP2 gene risk variants for pancreatic cancer. *Nat. Commun.*, 11, 3175.
14. Low, S.-K. et al. (2010) Genome-wide association study of pancreatic cancer in Japanese population. *PLoS One*, 5, e11824.
15. Petersen, G.M. et al. (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.*, 42, 224–228.
16. Wolpin, B.M. et al. (2014) Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat. Genet.*, 46, 994–1000.
17. Zhang, M. et al. (2016) Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget*, 7, 66328–66343.
18. Lu, Y. et al. (2021) Identification of recessively inherited genetic variants potentially linked to pancreatic cancer risk. *Front. Oncol.*, 11, 771312.
19. Pistoni, L. et al.; PanGenEU Study Investigators. (2021) Associations between pancreatic expression quantitative traits and risk of pancreatic ductal adenocarcinoma. *Carcinogenesis*, 42, 1037–1045.
20. Lu, Y. et al. (2021) Association of genetic variants affecting microRNAs and pancreatic cancer risk. *Front. Genet.*, 12, 1–31.
21. Corradi, C. et al. (2021) Genome-wide scan of long noncoding RNA single nucleotide polymorphisms and pancreatic cancer susceptibility. *Int. J. Cancer*, 148, 2779–2788.
22. Corradi, C. et al. (2023) Polymorphic variants involved in methylation regulation: a strategy to discover risk loci for pancreatic ductal adenocarcinoma. *J. Med. Genet.* doi:10.1136/jmg-2022-108910
23. Babteen, N.A. et al. (2020) Signal peptide missense variant in cancer-brake gene CTLA4 and breast cancer outcomes. *Gene*, 737, 144435.
24. Wu, J. et al. (2021) Correlation between ZBRK1/ZNF350 gene polymorphism and breast cancer. *BMC Med. Genom.*, 14, 7.
25. Rubinstein, J.C. et al. (2019) APC mutational patterns in gastric adenocarcinoma are enriched for missense variants with associated decreased survival. *Genes Chromosomes Cancer*, 59, 64–68. doi:10.1002/gcc.22792
26. Schumacher, F.R. et al.; Profile Study. (2018) Association analyses of more than 140 000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.*, 50, 928–936.
27. Charbonneau, B. et al.; for AOCS/ACS group. (2014) Risk of ovarian cancer and the NF-κB pathway: genetic association with IL1A and TNFSF10. *Cancer Res.*, 74, 852–861.
28. Jiang, X. et al. (2022) Uncovering variable neoplasms between ATM protein-truncating and common missense variants using 394 694 UK Biobank exomes. *Genes Chromosomes Cancer*, 61, 523–529.
29. Moyer, C.L. et al. (2020) Rare BRIP1 missense alleles confer risk for ovarian and breast cancer. *Cancer Res.*, 80, 857–867.
30. Hall, M.J. et al. (2021) Germline pathogenic variants in the ataxia telangiectasia mutated (ATM) gene are associated with high and moderate risks for multiple cancers. *Cancer Prev. Res. (Phila)*, 14, 433–440.
31. Sheng, X. et al. (2012) MDR1 C3435T polymorphism and cancer risk: a meta-analysis based on 39 case-control studies. *Mol. Biol. Rep.*, 39, 7237–7249.

32. 1000 Genomes Project Consortium et al. (2010) A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061–1073.
33. Campa, D. et al. (2013) Genetic susceptibility to pancreatic cancer and its functional characterisation: the PANcreatic Disease ReseArch (PANDoRA) consortium. *Dig. Liver Dis.*, 45, 95–99.
34. Campa, D. et al. (2023) The PANcreatic Disease ReseArch (PANDoRA) consortium: ten years' experience of association studies to understand the genetic architecture of pancreatic cancer. *Crit. Rev. Oncol. Hematol.*, 186, 104020.
35. Adzhubei, I.A. et al. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, 7, 248–249.
36. Ng, P.C. et al. (2001) Predicting deleterious amino acid substitutions. *Genome Res.*, 11, 863–874.
37. GTEx Consortium. (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, 45, 580–585.
38. Ghoussaini, M. et al. (2021) Open targets genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.*, 49, D1311–D1320.
39. Alliance of Genome Resources Consortium. (2022) Harmonizing model organism data in the alliance of genome resources. *Genetics*, 220.
40. Kanai, M. et al. (2018) Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.*, 50, 390–400.
41. Sinnott-Armstrong, N. et al.; FinnGen. (2021) Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.*, 53, 185–194.
42. Forsythe, E. et al. (2013) Bardet-Biedl syndrome. *Eur. J. Hum. Genet.*, 21, 8–13.
43. Turcot, V. et al.; CHD Exome + Consortium. (2018) Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.*, 50, 26–41.
44. Mansour Aly, D. et al.; Regeneron Genetics Center. (2021) Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat. Genet.*, 53, 1534–1542.
45. Richardson, T.G. et al. (2020) Use of genetic variation to separate the effects of early and later life adiposity on disease risk: Mendelian randomisation study. *BMJ*, 369, m1203. doi:10.1136/bmj.m1203.
46. Locke, A.E. et al.; LifeLines Cohort Study. (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518, 197–206.
47. Justice, A.E. et al. (2017) Genome-wide meta-analysis of 241 258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.*, 8, 14977.
48. Graff, M. et al.; CHARGE Consortium. (2017) Genome-wide physical activity interactions in adiposity—a meta-analysis of 200 452 adults. *PLoS Genet.*, 13, e1006528.