ORIGINAL ARTICLE

Associations between pancreatic expression quantitative traits and risk of pancreatic ductal adenocarcinoma

Laura Pistoni[1,†], Manuel Gentiluomo[1,†,©], Ye Lu[2], Evangelina López de Maturana[3,4], Viktor Hlavac[5], Giuseppe Vanella[6,7], Erika Darvasi[8], Anna Caterina Milanetto[9], Martin Oliverius[10], Yogesh Vashist[11], Milena Di Leo[12], Beatrice Mohelnikova-Duchonova[13], Renata Talar-Wojnarowska[14], Cristian Gheorghe[15], Maria Chiara Petrone[6], Oliver Strobel[16], Paolo Giorgio Arcidiacono[6], Ludmila Vodickova[17], Andrea Szentesi[8,18], Gabriele Capurso[6,7], László Gajdán[19], Giuseppe Malleo[20], George E.Theodoropoulos[21], Daniela Basso[22], Pavel Soucek[5], Hermann Brenner[23,24,25,©], Rita T.Lawlor[26], Luca Morelli[27], Audrius Ivanauskas[28]; PanGenEU Study Investigators[‡]; Emanuele Federico Kauffmann[29], Angelica Macauda[1,2], Maria Gazouli[30,©], Livia Archibugi[6,7], Michael Nentwich[11], Martin Loveček[13], Giulia Martina Cavestro[31], Pavel Vodicka[17], Stefano Landi[1], Francesca Tavano[32], Cosimo Sperti[9], Thilo Hackert[16,©], Juozas Kupcinskas[28], Raffaele Pezzilli[33], Angelo Andriulli[32], Luca Pollina[34], Edita Kreivenaite[28], Domenica Gioffreda[32], Krzysztof Jamroziak[35], Péter Hegyi[8,18], Jakob R.Izbicki[11], Sabrina Gloria Giulia Testoni[6], Raffaella Alessia Zuppardo[31], Dania Bozzato[9], John P.Neoptolemos[16], Núria Malats[3,4], Federico Canzian[2,§,©] and Daniele Campa[1,*,§]

[1]Department of Biology, University of Pisa, Pisa, Italy, [2]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, [3]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain, [4]CIBERONC, Madrid, Spain, [5]Biomedical Center, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czech Republic, [6]Pancreato-Biliary Endoscopy and Endosonography Division, Pancreas Translational and Clinical Research Center, Vita-Salute San Raffaele University, IRCCS San Raffaele Scientific Institute, Milan, Italy, [7]Sant'Andrea Hospital, Faculty of Medicine and Psychology, Sapienza University of Rome, Rome, Italy, [8]First Department of Medicine, University of Szeged, Szeged, Hungary, [9]Department of Surgery, Oncology and Gastroenterology-DiSCOG, University of Padova, Padua, Italy, [10]Department of Surgery, Faculty Hospital Kralovske Vinohrady and Third Faculty of Medicine, Charles University, Prague, Czech Republic, [11]Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [12]Digestive Endoscopy Unit, Division of Gastroenterology, Humanitas Research Hospital, Milan, Italy, [13]Department of Surgery I, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc, Olomouc, Czech Republic, [14]Department of Digestive Tract Diseases, Medical University of Łodz, Łodz, Poland, [15]Fundeni Clinical Institute, Bucharest, Romania, [16]Department of General Surgery, University of Heidelberg, Heidelberg, Germany, [17]Institute of Biology and Medical Genetics, First Medical Faculty, Prague, Czech Republic; Institute of Experimental Medicine, Czech Academy of Sciences, Prague, Czech Republic, [18]Institute for Translational Medicine, Medical School, University of Pécs, Pécs, Hungary, [19]Szent György University Teaching Hospital of Fejér County,

Székesfehérvár, Hungary, [20]Department of Surgery, The Pancreas Institute, University and Hospital Trust of Verona, Verona, Italy, [21]Colorectal Unit, First Department of Propaedeutic Surgery, Athens Medical School, National and Kapodistrian University of Athens, Athens, Greece, [22]Department of Laboratory Medicine, University Hospital of Padova, Padua, Italy, [23]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, [24]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, [25]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, [26]ARC-NET: Centre for Applied Research on Cancer, University and Hospital Trust of Verona, Verona, Italy, [27]General Surgery Unit, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy, [28]Department of Gastroenterology and Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania, [29]Unità Operativa Chirurgia Generale e dei Trapianti, University of Pisa, Pisa, Italy, [30]Department of Basic Medical Sciences, Laboratory of Biology, Medical School, National and Kapodistrian University of Athens, Athens, Greece, [31]Division of Experimental Oncology, Gastroenterology and Gastrointestinal Endoscopy Unit, Vita-Salute San Raffaele University, IRCCS Ospedale San Raffaele Scientific Institute, Milan, Italy, [32]Division of Gastroenterology and Research Laboratory, IRCCS Scientific Institute and Regional General Hospital 'Casa Sollievo della Sofferenza', San Giovanni Rotondo, Italy, [33]Division of Gastroenterology, San Carlo Hospital, Potenza, Italy, [34]Division of Surgical Pathology, Department of Surgical, Medical, Molecular Pathology and Critical Area, University of Pisa, Pisa, Italy and [35]Department of Hematology, Transplantation and Internal Medicine, Medical University of Warsaw, Warsaw, Poland

*To whom correspondence should be addressed. Department of Biology, University of Pisa, Via Derna 1, 56126 Pisa, Italy. Tel: +39 050 2211510; Email: daniele.campa@unipi.it
†These authors share the first position.
‡Members of the PanGenEu consortium are listed at Supplementary Material: Annex S1.
§These authors share the last position.

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is among the most lethal cancers. Its poor prognosis is predominantly due to the fact that most patients remain asymptomatic until the disease reaches an advanced stage, alongside the lack of early markers and screening strategies. A better understanding of PDAC risk factors is essential for the identification of groups at high risk in the population. Genome-wide association studies (GWAS) have been a powerful tool for detecting genetic variants associated with complex traits, including pancreatic cancer. By exploiting functional and GWAS data, we investigated the associations between polymorphisms affecting gene function in the pancreas (expression quantitative trait loci, eQTLs) and PDAC risk. In a two-phase approach, we analysed 13 713 PDAC cases and 43 784 controls and identified a genome-wide significant association between the A allele of the rs2035875 polymorphism and increased PDAC risk ($P = 7.14 \times 10^{-10}$). This allele is known to be associated with increased expression in the pancreas of the keratin genes *KRT8* and *KRT18*, whose increased levels have been reported to correlate with various tumour cell characteristics. Additionally, the A allele of the rs789744 variant was associated with decreased risk of developing PDAC ($P = 3.56 \times 10^{-6}$). This single nucleotide polymorphism is situated in the *SRGAP1* gene and the A allele is associated with higher expression of the gene, which in turn inactivates the cyclin-dependent protein 42 (*CDC42*) gene expression, thus decreasing the risk of PDAC. In conclusion, we present here a functional-based novel PDAC risk locus and an additional strong candidate supported by significant associations and plausible biological mechanisms.

Abbreviations

CI	confidence interval
eQTL	expression quantitative trait locus
GWAS	genome-wide association study
IRB	institutional review board
LD	linkage disequilibrium
MAF	minor allele frequency
OR	odds ratio
PANDoRA	PANcreatic Disease ReseArch
PanGenEU	European Study Digestive Diseases and Genetics
PDAC	pancreatic ductal adenocarcinoma

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the seventh most common cancer worldwide. It ranks as the fourth cause of cancer related deaths in the western world (1,2) and is predicted to become the second by 2030 (3).

The incidence varies significantly within regions, continents, populations and ethnicities (1) suggesting that genetics, lifestyle and environmental factors play an important role in the development of the pathology. Several epidemiologic PDAC risk factors have been identified, including cigarette smoking, heavy alcohol intake, type two diabetes mellitus, chronic pancreatitis and family history of PDAC (4). Recently, Mendelian randomization approaches have identified body mass index as a causative risk factor for PDAC (5,6). Moreover, rare high-penetrance mutations and high-frequency low-penetrance variants, discovered through genome-wide association studies (GWASs), contribute, alone or in combination, to the genetic susceptibility of PDAC (7–15). Additional susceptibility variants have been identified through large multicentric gene candidate approaches and through secondary analysis of published GWAS data (16–25). The number of identified loci is, however, very small if compared with breast, prostate or colorectal cancers and the fraction of the heritability they explain is limited. Moreover, in the vast

majority of the cases, the understanding of the underlying biological mechanisms and functions of GWAS-identified risk loci is still a challenge (26).

In addition, there are no established strategies for early detection and standard screening program for the general population (27). The implementation of new strategies to uncover new risk factors is of the utmost importance. Defining a complete set of PDAC risk factors, both genetic and non-genetic, would be a first step in order to perform risk stratification in the population and define groups of people at high risk who would then be subjected to more intensive surveillance and/or screening.

In parallel, it would be especially beneficial to identify SNPs with an effect on gene function or gene expression regulation to gain a deeper insight in the molecular biology of the disease.

In particular, certain germline polymorphisms, called 'expression quantitative trait loci' (eQTLs), are associated with gene expression in a wide range of human tissues. Considering that many human traits are affected by alterations in gene expression, eQTLs could represent an effective tool to explore the aetiology of human diseases (28–30). The involvement of eQTLs in the development of several human diseases, including cancer, has been already shown (31–34).

With these premises we hypothesized that pancreatic eQTLs are implicated in PDAC development and aimed to find new susceptibility risk loci. We examined the entire human genome for pancreatic eQTLs and analysed their involvement in PDAC susceptibility in a study on 13 713 PDAC cases and 43 784 controls.

Materials and methods

This study followed a two-phase approach, performing a discovery phase leveraging available data from three GWASs for PDAC risk and a validation phase for which we used three additional independent populations. In the validation phase, we performed *de novo* genotyping of the most significant SNPs, using additional cases and controls recruited by the PANcreatic Disease ReseArch (PANDoRA) consortium, and used summary statistics of a Japanese GWAS and of the European Study Digestive Diseases and Genetics (PanGenEU).

Study populations

For the discovery phase, data from PanScan I, PanScan II and PanC4 were downloaded from the database of Genotypes and Phenotypes (dbGaP) web site (study accession numbers: phs000206.v5.p3 and phs000648.v1.p1, project reference: #12644). All the individuals were genotyped using either Illumina Infinium HumanHap550v3 (PanScan I), Illumina Infinium Human610-Quad (PanScan II) or HumanOmniExpressExome-8v1 (PanC4) DNA Analysis Genotyping BeadChip. Each participating study collected informed consent from study participants and received approval from the responsible institutional review board (IRB) as described in the original papers (8,9,11). After downloading the genotypes, we performed imputation and quality controls. Briefly, the genotypes were phased using SHAPEIT v2 software (35,36). We discarded SNPs with a minor allele frequency (MAF) <0.01, completion rate and call rate <98%, cryptic relatedness (PI_HAT > 0.2), gender mismatches, low-quality imputation score (INFO score <0.7), evidence for violations of Hardy–Weinberg equilibrium ($P < 1 \times 10^{-5}$), leaving 7 509 345 SNPs in the final dataset. Principal component analysis was performed with PLINK 2.0 (37) including genotypes from all populations of the phase 3 of the 1000 Genomes Project. Individuals not clustering in the principal component analysis with the 1000 Genomes subjects of European descent ($N = 439$) were excluded from further analysis. For this phase, genotyping data of 14 269 individuals (7207 cases and 7062 controls) were used. The 'inflation factor' did not show evidence of systematic inflation ($\lambda = 1.000$ for PanScan I, $\lambda = 1.015$ for PanScan II, $\lambda = 1.000$ for PanC4 and $\lambda = 1.000$ for the aggregate dataset).

As a first replication phase, the summary statistics of a GWAS comprising 2039 pancreatic cancer patients and 32 592 controls in the Japanese population was used to narrow the list of SNPs under consideration (15).

A total of 6580 individuals (3430 PDAC cases and 3150 controls) from the PANDoRA consortium were genotyped to validate the previously selected polymorphisms. PANDoRA has been described in detail elsewhere (38). It is a multicentric consortium consisting of 11 European countries (Greece, Italy, Germany, the Netherland, Denmark, Czech Republic, Hungary, Poland, Ukraine, Lithuania and UK), Brazil and Japan. PDAC cases were defined by an established diagnosis of PDAC and controls were individuals from the general population without any pancreatic disease at recruitment, individuals who were hospitalized for non-tumour related causes, or blood donors. Data on sex, age (at diagnosis for cases and at recruitment for controls) and country of origin were collected. Controls were recruited in the same geographical regions as the cases. Controls from the Netherlands and UK were obtained from the European Prospective Investigation into Cancer and Nutrition (EPIC) (39). Only subjects recruited in Europe were used for this study. 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.

As an additional replication set, we used 1317 cases and 700 controls from the PanGenEU study with genotyped germline DNA. PanGenEU has been described in detail elsewhere (40,41). Briefly, this is a case–control study conducted in Spain, Italy, Sweden, Germany, UK and Ireland, between 2009 and 2014. IRB approval and written informed consent were obtained from all participating centres and study participants, respectively. DNA samples were genotyped using the Illumina Infinium OncoArray-500K. Imputation of missing genotypes was performed using IMPUTE2 with 1000 Genomes-phase 3 as reference panel (42).

The final sample size used in this study was therefore 13 713 PDAC cases and 43 784 controls as shown in Table 1.

Identification of eQTLs

As a starting point for this work, pancreatic eQTLs were retrieved from 305 GTEx non-tumour pancreatic samples. The data used for the analyses described in this manuscript were obtained from the version V8 GTEx analysis release (dbGaP Accession phs000424.v8.p2), data access on 6 January 2021. The initial downloaded raw file contained 1 340 448 million pancreatic tissue-specific cis eQTLs, reported with a nominal $P < 0.0005$ of association with the expression of at least one gene. We excluded variants on sex chromosomes and rare variants (MAF < 0.05). After filtering, 576 685 eQTLs were used in the discovery analysis (PanScan and PanC4). Next, all variants showing a discovery $P < 1 \times 10^{-4}$ were analysed in the Japanese GWAS dataset. Finally, the polymorphisms showing a $P < 0.05$ in Japanese were analysed in PanGenEU and in PANDoRA. We included in this last step all SNPs that were significant in the discovery phase but that were not present in the Japanese dataset and excluded all SNPs in known PDAC risk loci (±1Mb from a previously reported PDAC risk SNP).

Sample preparation and genotyping

DNA of cases and controls within PANDoRA was extracted from whole blood, using the QIamp® 96 DNA QIAcube® HT Kit (Qiagen, Hilden, Germany). Genotyping was performed using TaqMan technology (ThermoFisher Applied Biosystems, Waltham, MA) in 384-well plates according to the manufacturer recommendation. Similar numbers of cases and controls were put in each plate. In addition, duplicate samples (8%) were added for quality control purposes. Genotypes were called using the QuantStudio™ 5 Real-Time PCR system (Thermofisher, USA) and QuantStudio software.

Statistical analysis

In the discovery phase, an unconditional logistic regression analysis was carried out by computing odds ratio (OR) and 95% confidence interval (95% CI) and *P*-values to test the association between prioritized pancreatic eQTLs and PDAC risk. The analysis was performed in 14 269 individuals and was adjusted for sex, age and the top eight principal components to avoid confounding due to population stratification.

The novel associations were replicated, using an unconditional logistic regression, in the PANDoRA consortium and in the PanGenEU study populations. Additional information on the workflow is given in Results. Deviation from Hardy–Weinberg equilibrium was tested for the

Table 1. Description of study populations

	PanScan I–II and PanC4	Japanese GWAS[a]	PANDoRA	PanGenEu	Total
Study phase	Discovery	Replication	Replication	Replication	
Number of SNPs investigated	576 685	48	4	4	
Number of subjects					
Cases	7207	2039	3150	1317	13 713
Controls	7062	32 592	3430	700	43 784
Total	14 269	34 631	6580	2017	57 497
Median age (Q1–Q3)					
Cases	65 (55–75)	62.7|66.3	66 (58–73)	66 (57–73)	65 (55–75)
Controls	65 (55–75)	43.6|56.3	58 (48–67)	65 (55–75)	65 (55–75)
Sex					
Female	47%	33.6%|61.0%	45%	42%	45%
Male	53%	39.0%|66.4%	55%	58%	55%

[a]Numbers obtained from the study of Lin et al. 2020 performed on the Asian population (61), this study is a meta-analysis of three populations, we reported information on age and sex as minimum\|maximum value reported in the study.

SNPs genotyped in PANDoRA using the control subjects. Association analyses were adjusted for sex, age and country of origin (PANDoRA) or five principal components (PanGenEU). After the analysis conducted within PANDoRA and in PanGenEU, a fixed effect meta-analysis was conducted between the results of the two phases.

In order to take into account the number of independent tests, linkage disequilibrium (LD), with a threshold of $r^2 = 0.8$, was used to discard SNPs representing the same association, leaving a total of 48 456 independent variants, setting therefore the Bonferroni-corrected threshold for statistical significance to $0.05/48\ 456 = 1.03 \times 10^{-6}$.

Bioinformatics tools

In addition to GTEx, we used two databases to test the regulatory potential (i.e. possible changes in transcription factors affinity, regulation of chromatin state) of the SNPs showing the best associations: HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) and RegulomeDB (https://www.regulomedb.org/regulome-search/). We used LDlink (https://ldlink.nci.nih.gov/) (43–45) to explore the LD between the variants we identified and polymorphisms already reported in the literature. Finally, we used the Roadmap Epigenomics Mapping Consortium data (46) and the web-tool 3D-genome interaction viewer (47,48) to evaluate the chromatin interaction and the epigenetic annotation in the newly identified risk loci.

Ethics statement

Each participating study obtained approval from the responsible 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.

Results

In the discovery phase, we analysed 576 685 pancreatic eQTLs and observed 36 917 with a statistically significant association with PDAC risk, considering a nominal $P < 0.05$. LD (with a threshold of $r^2 = 0.8$) was then used to discard SNPs representing the same locus, leaving a total of 4553 independent variants.

We then filtered for a threshold of significance of $P < 10^{-4}$, resulting in 48 independent SNPs. A summary of GTEx data on associations between the 48 SNPs and gene expression or splicing (when SNPs act as splicing quantitative traits) in normal pancreatic and other tissues is reported in Supplementary Table I.

Among the 48 top SNPs, 34 map to known/putative PDAC risk loci. The remaining 14 were analysed in the Japanese

dataset and one (*KRT8*-rs2035875) showed a statistically significant association (OR = 1.18, 95% CI 1.09–1.27, $P = 2.18 \times 10^{-5}$), 3 (*ATP6AP1L*-rs2407156, *LOC105377095*-rs55843545 and *SRGAP1*-rs789744) were not present in the dataset, whereas 10 did not show a statistically significant association ($P < 0.05$) (Table 2). The first four SNPs were therefore replicated in PANDoRA and PanGenEU.

For *LOC105377095*-rs55843545 a TaqMan probe was not available and therefore a proxy (*LOC105377095*-rs12185922, $r^2 = 0.96$ in 1000G Europeans) was used to genotype PANDoRA cases and controls. We did not observe significant deviation from Hardy–Weinberg equilibrium ($P > 0.001$) for the four SNPs genotyped in PANDoRA. The average call rate was 98% and the concordance rate with the duplicate samples was >99%.

Among the four eQTLs that were tested within the PANDoRA and PanGenEU populations, we observed an association between the allele A of *KRT8*-rs2035875 and an increased risk of PDAC in the additive mode of inheritance: OR = 1.08, 95% CI 1.00–1.17, $P = 0.046$ in the PANDoRA population and OR = 1.10, 95% CI 0.95–1.26, $P = 0.21$ in the PanGenEU population. The results of the meta-analysis of all the studies (PanScan-PanC4, Japanese GWAS, PANDoRA and PanGenEU) for *KRT8*-rs2035875 showed an OR = 1.11, 95%CI 1.08–1.15, $P = 7.14 \times 10^{-10}$ and no heterogeneity, confirming the association between the A allele and PDAC risk.

Additionally, the A allele of *SRGAP1*-*rs789744* was associated with a decreased risk of PDAC in the additive model of inheritance: OR = 0.91, 95% CI 0.83–0.99, $P = 0.03$ in the PANDoRA population and OR = 0.91, 95% CI 0.77–1.08, $P = 0.274$ in the PanGenEU population. The results of the meta-analysis of all the studies (PanScan-PanC4, PANDoRA and PanGenEU) for *SRGAP1*-rs789744 showed an OR = 0.90, 95% CI 0.86–0.94, $P = 3.56 \times 10^{-6}$ and no heterogeneity, confirming an association between the A allele and PDAC risk.

Finally, *ATP6AP1L*-rs2407156 was significant in PanGenEU (OR = 1.23, 95% CI 1.04–1.45, $P = 0.018$) but not in PANDoRA, and the meta-analysis result showed OR = 1.10, 95% CI 1.05–1.15, $P = 0.016$, and heterogeneity $I^2 = 52.2\%$, whereas rs55843545 did not show any association in either replication population. The results are shown in Table 3.

Functional relevance

The main hypothesis behind the study is the functional effect of the SNPs on gene expression in the pancreatic tissue. To obtain this information, we used the GTEx web site that for *KRT8*-rs2035875 predicted the A allele to be associated with higher expression of the *KRT8 and KRT18* genes and for *SRGAP1*-rs789744

Table 2. List of SNPs associated with PDAC risk in PanScan I–II and PanC4 and their replication in the Japanese GWAS

SNP	POS	m	M	MAF (%)	Gene	Risk locus	PanScan + PanC4 P-value GWAS	Japanese P-value GWAS
rs3827749	1:230841559	A	G	18	[O] AGT	rs1326889; Walsh N, 2018	7.34E-05	0.60
rs5051	1:230849872	T	C	42	[O] AGT	rs1326889; Walsh N, 2018	7.66E-05	0.80
rs4908325	1:99989064	T	A	30	[O] LINC01708		4.36E-05	0.75
rs148608463	12:121413027	A	G	37	[O] HNF1A-AS1	rs1182933; Klein, 2018	8.69E-06	0.06
rs11065385	12:121423386	A	G	33	[O] HNF1A	rs7310409; Childs, 2015	5.04E-06	0.12
rs2035875	12:53295917	A	G	50	[O] KRT8		3.41E-05	2.18E-05
rs789744	12:64485360	A	G	25	[O] SRGAP1		7.09E-05	NA
rs11618832	13:28477111	A	G	43	[O] PLUT	rs9554197; Childs, 2015	4.11E-08	1.66E-05
rs7994281	13:28534916	G	C	27	[N] CDX2	rs9554197; Childs, 2015	4.59E-05	6.48E-04
rs7146643	14:105757392	C	T	31	[O] BRF1		8.53E-05	0.99
rs11248878	16:1511758	G	A	13	[O] CLCN7		4.53E-05	0.90
rs8054218	16:75252161	G	A	12	[N] CTRB1	rs7190458; Wolpin, 2014	2.07E-06	0.59
rs1035539	16:75276775	G	A	35	[O] BCAR1	rs7190458; Wolpin, 2014	7.98E-06	0.48
rs8046743	16:75294122	A	T	45	[O] BCAR1	rs7190458; Wolpin, 2014	1.99E-05	0.96
rs10871308	16:75342846	C	T	24	[O] CFDP1	rs7190458; Wolpin, 2014	2.17E-05	0.63
rs37596	16:75498320	A	C	19	[N] CFDP1	rs7190458; Wolpin, 2014	3.91E-06	0.58
rs2325748	17:45974251	T	C	30	[O] SP2		6.28E-05	0.48
rs12601827	17:46016613	G	A	44	[O] SP2-AS1		9.98E-05	0.72
rs2236141	22:29137870	T	C	14	[O] CHEK2	rs16986825; Wolpin, 2014	3.26E-05	0.02
rs5762799	22:29187380	T	G	19	[N] XBP1	rs16986825; Wolpin, 2014	2.34E-05	0.01
rs5762829	22:29228130	C	T	17	[O] ZNRF3	rs16986825; Wolpin, 2014	1.04E-05	7.40E-04
rs13058460	22:29258509	G	C	22	[O] ZNRF3	rs16986825; Wolpin, 2014	1.60E-05	1.73E-03
rs5762870	22:29261159	T	C	31	[O] ZNRF3	rs16986825; Wolpin, 2014	6.59E-05	1.77E-05
rs5762879	22:29286180	T	C	19	[O] ZNRF3	rs16986825; Wolpin, 2014	1.21E-05	1.38E-03
rs16986825	22:29300306	T	C	16	[O] ZNRF3	rs16986825; Wolpin, 2014	5.95E-05	0.01
rs55843545	3:54035492	A	G	29	[O] LOC105377095		4.73E-05	NA
rs2407156	5:81636368	G	A	25	[O] ATP6AP1L		9.45E-05	NA
rs2343590	6:166759622	C	G	26	[O] LOC100289495		3.53E-05	0.68
rs1122553	6:33798524	A	C	18	[N] MLN		8.14E-05	0.54
rs6905353	6:33813043	G	T	31	[N] LINC01016		5.02E-05	0.22
rs6462989	7:40869754	C	A	33	[O] SUGCT	rs17688601; Childs, 2015	6.00E-05	0.03
rs1733483	7:40873221	A	G	50	[O] SUGCT	rs17688601; Childs, 2015	4.71E-05	0.02
rs12701838	7:40877473	G	A	26	[O] SUGCT	rs17688601; Childs, 2015	3.26E-09	0.75
rs2410517	8:17151259	C	G	39	[O] VPS37A		1.64E-05	0.83
rs6985312	8:17168304	G	C	34	[O] MTMR7		9.58E-05	0.75
rs2943547	8:76451098	A	G	46	[O] HNF4G	rs2941471; Klein, 2018	1.87E-05	0.83
rs2941469	8:76469031	T	C	42	[O] HNF4G	rs2941471; Klein, 2018	1.01E-06	0.62
rs7855466	9:136121303	T	C	16	[N] ABO	rs505922; Amundadottir, 2009	5.66E-05	2.38E-07
rs8176715	9:136133148	T	C	41	[O] ABO	rs687289; Wolpin, 2014	1.75E-05	1.10E-03
rs512770	9:136133506	A	G	20	[O] ABO	rs687289; Wolpin, 2014	2.53E-05	0.12
rs8176645	9:136149098	A	T	39	[O] ABO	rs505922; Amundadottir, 2009	8.60E-15	3.81E-05
rs616154	9:136150466	C	T	45	[O] ABO	rs505922; Amundadottir, 2009	9.27E-12	0.59
rs8176632	9:136152547	T	C	16	[N] ABO	rs687289; Wolpin, 2014	9.96E-05	0.80
rs633862	9:136155444	C	T	43	[N] ABO	rs505922; Amundadottir, 2009	3.96E-06	0.02
rs9650778	9:136184798	T	C	13	[O] LCN1P2	rs687289; Wolpin, 2014	2.42E-05	0.56
rs11789139	9:136185324	G	C	18	[N] LCN1P2	rs505922; Amundadottir, 2009	2.66E-06	0.65
rs567493	9:136185528	A	G	26	[N] LCN1P2	rs687289; Wolpin, 2014	9.80E-05	1.93E-03
rs4962153	9:136323754	A	G	15	[O] ADAMTS13	rs687289; Wolpin, 2014	5.31E-06	0.53

SNP, rs-ID; POS, chromosome and variant position Genome Reference Consortium Human Build 37 (GRCh37); m, minor allele; M, major allele; MAF, minor allele frequency in PanScan I, II and PanC4 data; Gene: [O] 'overlapped' if the SNP is located inside the gene region, [N] the 'nearest' gene; NES, normalized effect size of SNPs on gene expression level in the pancreatic tissue samples from GTEx. NES > 0 indicates that the minor allele of the SNP is associated with increased expression of the gene, NES < 0 indicates association with decreased expression; eQTL, gene whose expression level is associated with SNP in the pancreatic tissue samples from GTEx; risk locus, polymorphisms already reported in association with PDAC risk by European ancestry GWAS in the region ±1Mb from the 48 SNPs selected.

predicted the A allele to be associated with higher expression of the *SRGAP1* gene.

The *KRT8*-rs2035875 SNP is located in a regulatory region and enhancer for the *KTR8* and *KRT18* genes in pancreatic tissue. The tool 3D-Genome Interaction Viewer and database showed that the *KRT8*-rs2035875 SNP interacts with the *KTR8* and *KRT18* genes and suggested also a long-range interaction with the *NR4A1* locus that is located 880 000 bp away.

RegulomeDB gave for *SRGAP1*-rs789744 a score of 0.18 and a rank of 7, indicating a very low functional potential, whereas HaploReg confirmed the SNP as a multi-tissue eQTL. The moderate potential functionality of the locus *SRGAP1*-rs789744 was confirmed also by roadmap epigenomics consortium data that report a weak transcription functionality and no epigenetic function.

Discussion

PDAC is a polygenic and complex multifactorial disease for which aetiology remains still not completely understood. At present, there are no molecular signatures to detect PDAC at early stages and genetic susceptibility has been less studied compared with other more common cancers. Nonetheless, several risk SNPs have been identified and, as for several other cancer types, many are not in gene coding regions. Instead, they are situated in intergenic and intronic regions and probably are involved in gene expression regulation. SNPs involved in regulating gene expression are called eQTLs and assessing their association with PDAC risk will not only further increase our knowledge on the genetic susceptibility but also shed light on their possible role in the etiopathogenesis of the disease and in the biological mechanisms involved in it. With these premises, we combined a GWAS approach and a genome-wide investigation of pancreatic eQTLs. We used a two-phase strategy. In the discovery phase, we analysed 14 269 individuals and we

identified 48 eQTLs associated with PDAC risk (P < 10⁻⁴). Checking and filtering for possible association and proximity with loci discovered in previous studies, we obtained a list of four novel variants to be tested in a validation phase and a list of 34 SNPs in loci already identified as PDAC risk regions.

Among the latter, a group of 11 SNPs are located on chromosome 9 in proximity of the *ABO* locus. Most are in moderate to high LD with rs505922, the index SNP that was identified in the first GWAS on PDAC risk ([8]). They are reported by GTEx to be strong eQTLs for *ABO* itself, as well as for nearby genes *SURF1* and *SURF6*. Although the functional link between ABO and PDAC risk is still unclear, it is worth noting that expression level of *ABO* was reported to be associated with PDAC risk in transcriptome-wide association studies ([49],[50]). To the best of our knowledge, no relation between SURF1 or SURF6 and PDAC has been reported.

Additionally, rs8054218, rs1035539, rs8046743, rs1087 1308 and rs37596 are located in the region of the *BCAR1/ CTRB1/CTRB2* locus on chromosome 16 ([10]), although they are in low LD (r^2 = 0.07) with rs7190458, the top SNP at the locus. These SNPs are strong eQTLs and splicing QTLs for a number of genes in multiple tissues, including the pancreas. Among the genes whose expression is associated with the eQTLs on chromosome 16, there are *CTRB1* and *CTRB2* ([51]), two chymotripsinogens that are precursors of pancreatic proteolytic enzymes. Genetic variation at this locus is associated also with susceptibility to chronic pancreatitis ([52]), which is a strong risk factor for PDAC.

The observation that some of the SNPs identified with this approach are located in established PDAC risk loci represents an implicit validation of this strategy. Additionally, it provides hints to the possible function of loci that were identified previously in the context of epidemiological studies but lacking mechanistic explanation of their role in disease susceptibility.

From a statistical point of view, the strongest association we observed was between the A allele of *KRT8*-rs2035875 SNP

Table 3. Associations with PDAC risk of the four candidate eQTLs selected in the discovery phase of the study

SNP Chromosome Position (hg38) Gene name	M/m[a]	Statistic[b]	PanScan I–II and PanC4 7207|7062[c]	Japanese GWAS 2039|32 592[c]	PANDoRA 2774|2981[c]	PanGenEU 1317|700[c]	Meta-analysis
rs12185922[d] 3p21.1 53 999 759 *LOC105377095*	A/G	MAF OR (95% CI) P-value	0.29 0.90 (0.85–0.94) 4.73 × 10⁻⁵	NA	0.28 1.01 (0.92–1.10) 0.89	0.26 1.04 (0.88–1.22) 0.68	— 0.96 (0.88–1.06) 0.44[e]
rs2407156 5q14.2 82 340 549 *ATP6AP1L*	A/G	MAF OR (95% CI) P-value	0.25 1.11 (1.06–1.18) 9.45 × 10⁻⁵	NA	0.25 1.03 (0.94–1.12) 0.55	0.22 1.23 (1.04–1.45) 0.018	- 1.10 (1.02–1.19) 0.016[e]
rs789744 12q14.2 64 091 580 *SRGAP1*	G/A	MAF OR (95% CI) P-value	0.25 0.90 (0.85–0.95) 7.09 × 10⁻⁵	NA	0.23 0.91 (0.83–0.99) 0.03	0.25 0.91 (0.77–1.08) 0.27	— 0.90 (0.86–0.94) 3.56 × 10⁻⁶
rs2035875 12q13.13 52 902 133 *KRT8*	G/A	MAF OR (95% CI) P-value	0.50 1.10 (1.05–1.16) 3.41 × 10⁻⁵	0.39 1.18 (1.09–1.27) 2.18 × 10⁻⁵	0.49 1.08 (1.00–1.17) 0.05	0.48 1.10 (0.95–1.26) 0.21	— 1.11 (1.08–1.15) 7.14 × 10⁻¹⁰

[a]Major and minor allele.
[b]MAF, minor allele frequency; OR, allelic odds ratio and corresponding 95% confidence interval; P-value results.
[c]Number of cases and controls (cases|controls).
[d]For this SNP, a genotyping assay was not available at the time of genotyping within PANDoRA. A SNP in strong LD, having TaqMan assays available, was chosen instead (rs12185922 proxy of rs55843545).
[e]The random effects statistical model was used instead of fixed effect for the meta-analysis because heterogeneity was reported I^2 > 25%.

on chromosome 12 and increased risk of developing PDAC ($P = 7.14 \times 10^{-10}$ in the overall meta-analysis). According to GTEx, the A allele is associated with increased expression of the keratin 8 and 18 (*KRT8* and *KRT18*) genes in the pancreatic tissue. The role of *KRT8*-rs2035875 in regulating the *KRT8* and *KRT18* genes is also supported by the Epigenomics Mapping Consortium data, where this SNP is annotated as super-enhancer of these genes. Furthermore, the 3D-Genome Interaction Viewer tool confirmed the interaction of the *KRT8*-rs2035875 locus with the genomic regions where *KRT8* and *KRT18* are located. Keratins K8 and K18 are the major components of the intermediate-filament cytoskeleton of simple epithelia. Increased levels of these keratins have been correlated with various tumour cell characteristics, including progression to malignancy (53–55), invasive behaviour and drug sensitivity, although their exact role in tumorigenesis has not yet been clarified.

In an experiment conducted using transgenic mice expressing the human *KRT8* gene, the role of this gene in pancreatic tissue development and health was highlighted, particularly on the exocrine component of the pancreatic tissue (55). In detail, the study showed that the amount of K8/K18 transcript was estimated to be ~3-fold higher in transgenic mice with anomalous pancreas than in non-transgenic or normal-pancreas transgenic mice. We observed that the A allele of *KRT8*-rs2035875 was associated with increased expression of *KRT8* and *KRT18* in pancreatic tissue, and the same allele is associated with increased risk of developing PDAC, suggesting a role of the *KRT8* and *KRT18* overexpression in pancreatic cancer.

The association of the *KRT8* gene considering a gene association approach (all the SNPs in the region together) was also shown by Lin *et al.* (15), however, the association between *KRT8*-rs2035875 and PDAC risk was not reported.

In the meta-analysis, the A allele of the Slit-Robo GTPase-Activating Protein 1 (*SRGAP1*)-rs789744 showed an association with decreased risk of developing PDAC; however, this association does not reach the Bonferroni threshold (1.03×10^{-6}) but it is very close (3.56×10^{-6}).

The decrease in risk associated with the A allele may have a biological explanation. According to GTEx, the A allele is associated with increased expression of the *SRGAP1* gene in the pancreatic tissue. The protein encoded by *SRGAP1*s gene is a GTPase activator that interacts with ROBO1 and CDC42 to negatively regulate various processes such as cell cycle progression, cell morphology, endocytosis cell adhesion and migration (56,57). *ROBO1* and *ROBO2* gene expression and regulation is key in pancreatic cancer, as reported by Pinho *et al.* (58), and therefore an eQTL that changes the expression of the *SRGAP1* in the pancreas could have an impact on the disease aetiology. Moreover, in a recent review, Xiao *et al.* reported that *CDC42* abnormal expression could be involved in various human tumours such as lung, breast, kidney, gastric, prostate, thyroid and PDAC, promoting neoplastic growth and metastasis (57). In addition, Feng *et al.* reported that *SRGAP1* plays a role in colorectal cancer, in particular its downregulation is associated with tumour progression and poor prognosis due to the diminished inhibition of *CDC42* (59). Moreover, He *et al.* reported that three genetic variants (rs781626187, rs797044990 and rs114817817) are associated with decreased SRGAP1 protein level and are associated with increased risk of developing papillary thyroid cancer (60). Two of the SNPs (rs114817817 and rs789744) share a modest degree of LD ($D' = 1$, $r^2 = 0.0011$ in the European population of the 1000 genomes database) and are close to each other (17 387 bp) and therefore could represent the same locus. There are no data in LDlink for the remaining two SNPs (rs781626187, rs797044990). The overall picture that could be gathered from these data is that the A allele of *SRGAP1*-rs789744 upregulates the expression of *SRGAP1* that, in turn, increases the inhibition exerted by SRGAP1 on CDC42 that could decrease the risk of developing PDAC. As suggested by He *et al.* for thyroid cancer, the locus around *SRGAP1* could be a low-penetrance modifying gene also for PDAC development. However, the locus where rs789744 is located is reported by the Roadmap Epigenomics Mapping Consortium data as a region with quiescent regulatory status and low and weak enhancer transcription role.

Clear strengths of this study are represented by the very large sample size and the two-phase approach that may decrease the chance of reporting spurious finding.

Possible limitations of this work are that we excluded rare variants and that we used a population of East Asian descent as a first validation step. However, considering that the success rate in validating SNPs across populations is limited, our choice has the advantage of identifying (when present in both datasets) risk loci for both ethnicities. We performed only individual eQTL analyses and did not combine them into gene-level predictions, because pancreatic cancer transcriptome-wide association studies have already been published, using the same PanScan I–II and PanC4 data we used (49,50).

In conclusion, we present here a potentially interesting association of two functional SNPs supported by a plausible biological mechanism influencing PDAC risk. The emergence of eQTLs as cancer risk markers highlights the importance of using functional SNPs to generate new insights into the understanding of the genetic basis of complex human traits.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Funding

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). The PANDoRA and PanGenEU 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, Germany, and the ISCIII, Spain, respectively. Data will be stripped from all information allowing identification of study participants.

Authors' contributions

D.C. conceived and designed the study. L.P. performed lab work. M.G. and E.L.d.M. performed data quality control and statistical analyses. L.P., M.G. and F.C. drafted the manuscript. All other authors provided samples and data. All authors critically read, commented and approved the manuscript.

References

1. Ferlay, J. *et al.* (2019) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer*, 144, 1941–1953.
2. Bray, F. *et al.* (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.*, 68, 394–424.
3. 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.
4. Maisonneuve, P. *et al.* (2015) Risk factors for pancreatic cancer: a summary review of meta-analytical studies. *Int. J. Epidemiol.*, 44, 186–198.
5. Lu, Y. *et al.* (2020) Mendelian randomisation study of the effects of known and putative risk factors on pancreatic cancer. *J. Med. Genet.*, 57, 820–828.
6. Langdon, R.J. *et al.* (2019) A phenome-wide mendelian randomization study of pancreatic cancer using summary genetic data. *Cancer Epidemiol. Biomarkers Prev.*, 28, 2070–2078.
7. Gentiluomo, M. *et al.* (2020) Germline genetic variability in pancreatic cancer risk and prognosis. *Semin. Cancer Biol.*, S1044-579X(20)30174-7.
8. 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.
9. 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.
10. Wolpin, B.M. *et al.* (2014) Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat. Genet.*, 46, 994–1000.
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. 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.
13. Klein, A.P. *et al.* (2018) Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat. Commun.*, 9, 556.
14. Galeotti, A.A. *et al.* (2021) Polygenic and multifactorial scores for pancreatic ductal adenocarcinoma risk prediction. *J. Med. Genet.*, 58, 369–377.
15. Lin, Y. *et al.* (2020) Genome-wide association meta-analysis identifies GP2 gene risk variants for pancreatic cancer. *Nat. Commun.*, 11, 3175.
16. Campa, D. *et al.* (2016) Functional single nucleotide polymorphisms within the cyclin-dependent kinase inhibitor 2A/2B region affect pancreatic cancer risk. *Oncotarget*, 7, 57011–57020.
17. Campa, D. *et al.* (2015) TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. *Int. J. Cancer*, 137, 2175–2183.
18. Campa, D. *et al.* (2013) Lack of replication of seven pancreatic cancer susceptibility loci identified in two Asian populations. *Cancer Epidemiol. Biomarkers Prev.*, 22, 320–323.
19. Gentiluomo, M. *et al.* (2019) Genetic polymorphisms in inflammatory genes and pancreatic cancer risk: a two-phase study on more than 14 000 individuals. *Mutagenesis*, 34, 395–401.
20. Gentiluomo, M. *et al.* (2019) Genetic variants in taste-related genes and risk of pancreatic cancer. *Mutagenesis*, 34, 391–394.
21. Xu, X. *et al.* (2019) Genetic variants in the liver kinase B1-AMP-activated protein kinase pathway genes and pancreatic cancer risk. *Mol. Carcinog.*, 58, 1338–1348.
22. Feng, Y. *et al.* (2019) Potential functional variants in SMC2 and TP53 in the AURORA pathway genes and risk of pancreatic cancer. *Carcinogenesis*, 40, 521–528.
23. Yang, W. *et al.* (2019) Three novel genetic variants in NRF2 signaling pathway genes are associated with pancreatic cancer risk. *Cancer Sci.*, 110, 2022–2032.
24. Campa, D. *et al.* (2020) Genome-wide association study identifies an early onset pancreatic cancer risk locus. *Int. J. Cancer*, 147, 2065–2074.
25. 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.
26. Barbeira, A.N. *et al.* (2019) Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.*, 15, e1007889.
27. Neoptolemos, J.P. *et al.* (2018) Pancreatic cancer. *Pancreat. Cancer*, 2, 1–1661.
28. Strunz, T. *et al.* (2018) A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. *Sci. Rep.*, 8, 5865.
29. Gilad, Y. *et al.* (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.*, 24, 408–415.
30. Nica, A.C. *et al.* (2013) Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 368, 20120362.
31. Dai, J.Y. *et al.* (2020) DNA methylation and cis-regulation of gene expression by prostate cancer risk SNPs. *PLoS Genet.*, 16, e1008667.
32. Ferreira, M.A. *et al.*; EMBRACE Collaborators, GC-HBOC Study Collaborators, GEMO Study Collaborators, ABCTB Investigators, HEBON Investigators, BCFR Investigators. (2019) Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. *Nat. Commun.*, 10, 1741.
33. Fung, J.N. *et al.* (2018) Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. *Sci. Rep.*, 8, 11424.
34. Loo, L.W.M. *et al.* (2017) In silico pathway analysis and tissue specific cis-eQTL for colorectal cancer GWAS risk variants. *BMC Genomics*, 18, 381.
35. 1000 Genomes Project Consortium *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061–1073.
36. Delaneau, O. *et al.* (2014) Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.*, 5, 3934.
37. Chang, C.C. *et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4, 7.
38. 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.
39. Riboli, E. *et al.* (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.*, 5, 1113–1124.
40. Gomez-Rubio, P. *et al.*; PanGenEU Study Investigators. (2017) Reduced risk of pancreatic cancer associated with asthma and nasal allergies. *Gut*, 66, 314–322.
41. Molina-Montes, E. *et al.*; PanGenEU Study Investigators. (2021) Deciphering the complex interplay between pancreatic cancer, diabetes mellitus subtypes and obesity/BMI through causal inference and mediation analyses. *Gut*, 70, 319–329.
42. Howie, B.N. *et al.* (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, 5, e1000529.
43. Machiela, M.J. *et al.* (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, 31, 3555–3557.
44. Ward, L.D. *et al.* (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, 40(Database issue), D930–D934.
45. Boyle, A.P. *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, 22, 1790–1797.
46. Roadmap Epigenomics Consortium *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, 518, 317–329.
47. Yang, D. *et al.* (2018) 3DIV: a 3D-genome Interaction Viewer and database. *Nucleic Acids Res.*, 46, D52–D57.
48. Jung, I. *et al.* (2019) A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.*, 51, 1442–1449.

49. Zhong, J. *et al.* (2020) A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer. *J. Natl. Cancer Inst.*, 112, 1003–1012.

50. Liu, D. *et al.* (2020) A transcriptome-wide association study identifies candidate susceptibility genes for pancreatic cancer risk. *Cancer Res.*, 80, 4346–4354.

51. López de Maturana, E. *et al.* (2021) A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. *Genome Med.,* 13, 15.

52. Rosendahl, J. *et al.*; all members of the PanEuropean Working group on ACP. (2018) Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut*, 67, 1855–1863.

53. Schneider, A. *et al.* (2006) Keratin 8 mutations are not associated with familial, sporadic and alcoholic pancreatitis in a population from the United States. *Pancreatology*, 6, 103–108.

54. Cavestro, G.M. *et al.* (2003) Association of keratin 8 gene mutation with chronic pancreatitis. *Dig. Liver Dis.*, 35, 416–420.

55. Casanova, M.L. *et al.* (1999) Exocrine pancreatic disorders in transgenic mice expressing human keratin 8. *J. Clin. Invest.*, 103, 1587–1595.

56. Bateman, A. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, 47, D506–D515.

57. Xiao, X.-H. *et al.* (2018) Regulating Cdc42 and its signaling pathways in cancer: small molecules and MicroRNA as new treatment candidates. *Molecules*, 23, 787.

58. Pinho, A.V. *et al.* (2018) ROBO2 is a stroma suppressor gene in the pancreas and acts via TGF-β signalling. *Nat. Commun.*, 9, 5083.

59. Feng, Y. *et al.* (2016) srGAP1 mediates the migration inhibition effect of Slit2-Robo1 in colorectal cancer. *J. Exp. Clin. Cancer Res.*, 35, 191.

60. He, H. *et al.* (2013) SRGAP1 is a candidate gene for papillary thyroid carcinoma susceptibility. *J. Clin. Endocrinol. Metab.*, 98, E973–E980.