Title: A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer

Authors: Evangelina López de Maturana (1)*, Juan Antonio Rodríguez (2)*, Lola Alonso (1), Oscar Lao (2), Esther Molina-Montes (1), Isabel Adoración Martín-Antoniano (1), Paulina Gómez-Rubio (1), Rita Lawlor (3), Alfredo Carrato (4), Manuel Hidalgo (5), Mar Iglesias (6), Xavier Molero (7), Matthias Löhr (8), Christopher Michalski (9), José Perea (10), Michael O'Rorke (11), Victor Manuel Barberà (12), Adonina Tardón (13), Antoni Farré (14), Luís Muñoz-Bellvís (15), Tanja Crnogorac-Jurcevic (16), Enrique Domínguez-Muñoz (17), Thomas Gress (18), William Greenhalf (19), Linda Sharp (20), Luís Arnes (21), Lluís Cecchini (6), Joaquim Balsells (7) , Eithne Costello (19), Lucas Ilzarbe (6), Jörg Kleeff (9), Bo Kong (22), Mirari Márquez (1), Josefina Mora (14), Damian O'Driscoll (23), Aldo Scarpa (3), Weimin Ye (24), Jingru Yu (24), PanGenEU Investigators (25), Montserrat García-Closas (26), Manolis Kogevinas (27), Nathaniel Rothman (26), Debra Silverman (26), SBC/EPICURO Investigators (28), Demetrius Albanes (26), Alan Arslan (29), Laura Beane-Freeman (26), Paige Bracci (30), Paul Brennan (31), Bas Bueno-de-Mesquita (32), Julie Buring (33), Federico Canzian (34), Margaret Du (35), Steve Gallinger (36), Michael Gaziano (37), Phyllis Goodman (38), Marc Gunter (31), Loic LeMarchand (39), Donghui Li (40), Rachael Neale (41), Ulrika Peters (42), Gloria Petersen (43), Harvey Risch (44), Maria José Sánchez (45), Xiao-Ou Shu (46), Mark Thornquist (42), Kala Visvanathan (43), Wei Zheng (46), Stephen Chanock (26), Douglas Easton (47), Brian Wolpin (48), Rachael Stolzenberg-Solomon (26), Alison Klein (49), Laufey Amundadottir (50), Marc Marti-Renom (51), Francisco Xavier Real (52), Núria Malats (1).

* Equal contributions

Authors' affiliations:

(1) Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), Madrid, and CIBERONC, Spain.
(2) CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.
(3) ARC-Net Centre for Applied Research on Cancer and Department of Pathology and Diagnostics, University and Hospital trust of Verona, Verona, Italy.
(4) Department of Oncology, Ramón y Cajal University Hospital, IRYCIS, Alcala University, Madrid, and CIBERONC, Spain.
(5) Madrid-Norte-Sanchinarro Hospital, Madrid, Spain; and Weill Cornell Medicine, New York, USA.
(6) Hospital del Mar—Parc de Salut Mar, Barcelona, and CIBERONC, Spain.
(7) Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute (VHIR), Barcelona, Universitat Autònoma de Barcelona, and CIBEREHD, Spain.
(8) Gastrocentrum, Karolinska Institutet and University Hospital, Stockholm, Sweden.
(9) Department of Surgery, Technical University of Munich, Munich; and Department of Visceral, Vascular and Endocrine Surgery, Martin-Luther-University Halle-WittenberHalle (Saale), Germany.
(10) Department of Surgery, Hospital 12 de Octubre; and Department of Surgery and Health Research Institute, Fundación Jiménez Díaz, Madrid, Spain.
(11) Centre for Public Health, Belfast, Queen's University Belfast, UK; and College of Public Health, The University of Iowa, Iowa City, IA, US.
(12) Molecular Genetics Laboratory, General University Hospital of Elche, Spain.
(13) Department of Medicine, Instituto Universitario de Oncología del Principado de Asturias (IUOPA), Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, and CIBERESP, Spain.
(14) Department of Gastroenterology and Clinical Biochemistry, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.
(15) Department of Surgery, Hospital Universitario de Salamanca – IBSAL. Universidad de Salamanca and CIBERONC, Spain.
(16) Barts Cancer Institute, Centre for Molecular Oncology, Queen Mary University of London, London, UK.

(17) Department of Gastroenterology, University Clinical Hospital of Santiago de Compostela, Spain.

(18) Department of Gastroenterology, University Hospital of Giessen and Marburg, Marburg, Germany.

(19) Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK.

(20) National Cancer Registry Ireland and HRB Clinical Research Facility, University College Cork, Cork, Ireland; and Newcastle University, Institute of Health & Society, Newcastle, UK.

(21) Centre for Stem Cell Research and Developmental Biology, University of Copenhagen, Denmark; and Department of Genetics and Development, Columbia University Medical Center, New York, NY, US; Department of Systems Biology, Columbia University Medical Center, New York, NY, US.

(22) Department of Surgery, Technical University of Munich, Munich, Germany.

(23) National Cancer Registry Ireland and HRB Clinical Research Facility, University College Cork, Cork, Ireland.

(24) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stokholm, Sweden.

(25) PanGenEU Study investigators (Supplementary Annex 1)

(26) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, US.

(27) Institut Municipal d'Investigació Mèdica – Hospital del Mar, Centre de Recerca en Epidemiologia Ambiental (CREAL), Barcelona, and CIBERESP, Spain.

(28) SBC/EPICURO Investigators (Supplementary Annex 2)

(29) Department of Obstetrics and Gynecology, New York University School of Medicine, New York, NY; Department of Environmental Medicine, New York University School of Medicine, New York, NY; Department of Population Health, New York University School of Medicine, New York, NY, US.

(30) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, US.

(31) International Agency for Research on Cancer (IARC), Lyon, France.

(32) Dept. for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.

(33) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, US.

(34) Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany.

(35) Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, US.

(36) Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada.

(37) Departments of Medicine, Brigham and Women's Hospital, VA Boston and Harvard Medical School, Boston, MA, US.

(38) SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, US.

(39) Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, US.

(40) University of Texas MD Anderson Cancer Center, Houston, TX, US.

(41) Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

(42) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, US.

(43) Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, US.

(44) Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, US.

(45) Escuela Andaluza de Salud Pública (EASP), Granada, Spain; Instituto de Investigación Biosanitaria Granada, Spain; Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; Universidad de Granada, Granada, Spain.

(46) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, US.

(47) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, UK.

(48) Department Medical Oncology, Dana-Farber Cancer Institute, Boston, US.

(49) Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD, US.
(50) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
(51) National Centre for Genomic Analysis (CNAG), Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST); Universitat Pompeu Fabra (UPF); ICREA, Barcelona, Spain.
(52) Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid; Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona; and CIBERONC, Spain.

Correspondence to: Núria Malats, Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), C/Melchor Fernandez Almagro 3, 28029-Madrid, Spain, Email: nmalats@cnio.es, Phone: +34 917328000; and Marc A. Martí-Renom, Structural Genomics Group, Centre Nacional d'Anàlisi Genòmica - Centre de Regulació Genòmica (CNAG-CRG), Baldiri Reixac 4, 08028-Barcelona, Email: martirenom@cnag.crg.eu, Phone: +34 9340 33743.

Keywords: pancreatic cancer risk, genome wide association analysis, genetic susceptibility, 3D genomic structure, Local Indices of Genome Spatial Autocorrelation

Word count – Abstract: **254**
Word count – Text: **6,567**
Number of References: **86**
Number of Tables: **1**
Number of Figures: **5**
Number of Supplementary Tables: **8**
Number of Supplementary Figures: **9**

ABSTRACT

Background. Pancreatic cancer (PC) is a complex disease in which both non-genetic and genetic factors interplay. To-date, 40 GWAS hits have been associated with PC risk in individuals of European descent, explaining 4.1% of the phenotypic variance.

Methods. We complemented a new conventional PC GWAS (1D) with genome spatial autocorrelation analysis (2D) permitting to prioritize low frequency variants not detected by GWAS. These were further expanded via Hi-C maps (3D) interactions to gain additional insight into the inherited basis of PC. *In-silico* functional analysis of public genomic information allowed prioritization of potentially relevant candidate variants.

Results. We identified several new variants located in genes for which there is experimental evidence of their implication in the biology and function of pancreatic acinar cells. Among them is a novel independent variant in *NR5A2* (rs3790840) with a meta-analysis *p*-value= 5.91E-06 in 1D-approach and a Local Moran's Index (LMI)= 7.76 in 2D-approach. We also identified a multi-hit region in *CASC8* - a lncRNA associated with pancreatic carcinogenesis - with a lowest p-value= 6.91E-05. Importantly, two new PC loci were identified both by 2D- and 3D-approaches: *SIAH3* (LMI= 18.24), *CTRB2/BCAR1* (LMI= 6.03), in addition to a chromatin interacting region in *XBP1* - a major regulator of the ER stress and unfolded protein responses in acinar cells - identified by 3D; all of them with a strong *in-silico* functional support.

Conclusions. This multi-step strategy, combined with an in-depth *in-silico* functional analysis, offers a comprehensive approach to advance the study of PC genetic susceptibility and could be applied to other diseases.

BACKGROUND

Pancreatic cancer (PC) has a relatively low incidence but it is one of the deadliest tumors. In Western countries, PC ranks fourth among cancer-related deaths with 5-year survival of 3-7% in Europe [1-3]. In the last decades, progress in the management of patients with PC has been meagre. In addition, mortality is rising [2] and it is estimated that PC will become the second cause of cancer-related deaths in the United States by 2030 [4].

PC is a complex disease in which both genetic and non-genetic factors participate. However, relatively little is known about its etiologic and genetic susceptibility background. In comparison with other major cancers, fewer genome-wide association studies (GWAS) have been carried out and the number of patients included in them is relatively small (N=9,040). According to the GWAS Catalog, (January 2019) [5], 40 common germline variants sited in 32 loci and associated with PC risk have been identified in individuals of European descent [6-11]. However, these variants only explain 4.1% of the phenotypic variance for PC [12]. More importantly, given the challenges in performing new PC case-control studies with adequate clinical, epidemiological, and genetic information, the field is far from reaching the statistical power that has been achieved in other more common cancers such as breast, colorectal, or prostate cancers with >100,000 subjects included in GWAS, yielding a much larger number of genetic variants associated with them [5].

Current GWAS methodology relies on establishing simple SNP-disease associations by setting a strict statistical threshold of significance (p-value=$5x10^{-8}$) and replicating them in independent studies. This approach has been successful in minimizing false positive hits at the expense of discarding variants that may be truly associated with the disease displaying association p-values not reaching genome-wide significance after multiple testing correction or not being replicated in independent populations. The false negatives can be the result of weak associations or of low prevalence of the variant SNP assessed, among others. The "simple"

solution to this problem is to increase the sample size. However, it will take considerable time for PC GWAS studies to reach the number of subjects achieved in other tumors and the funding climate for replication studies is extremely weak. While a meta-analysis based on available datasets provides an alternative strategy for novel variant identification, this approach introduces heterogeneity because studies differ regarding methods, data quality, testing strategies, genetic background of the included individuals (e.g., population substructure), and study design, factors that can lead to lack of replicability. Therefore, we are faced with the need of exploring alternative approaches to discover new putative genetic risk variants missed by conventional GWAS criteria.

Here, we build upon one of the largest PC case-control studies with extensive standardized clinical and epidemiological data annotation and apply both a classical GWAS approach (1D strategy) and novel strategies for risk-variant discovery. We use, for the first time in genomics, the Local Moran's Index (LMI) [13], an approach that is widely followed in geospatial statistics. In its original application to geographic two-dimensional (2D) analysis, LMI identifies the existence of relevant autocorrelated clusters in the spatial arrangement of a variable, highlighting points closely surrounded by others with similar risk estimates values, allowing the identification of "hot spots". We computed LMI of (genomic) spatial autocorrelation to identify clusters of SNPs based on their similar risk estimates (odds ratio, OR) weighted by their genomic distance as measured by linkage disequilibrium (LD). By capturing LD structures of nearby SNPs, LMI leverages the values of SNPs with low minor allele frequencies (MAF) that conventional GWAS fail to assess properly. In this regard, LMI offers a novel opportunity to identify potentially relevant new sets of genomic candidates associated with PC genetic susceptibility.

In addition, we have taken advantage of recent advances in 3D genomic analyses providing insights into the spatial relationship of regulatory elements and their target genes.

Since GWAS have largely identified variants present in non-coding regions of the genome, a challenge has been to ascribe such variants to the corresponding regulated genes, which may lie far away in the genomic sequence. Chromosome Conformation Capture experiments (3C-related techniques) [14] can provide insight into the biology and function underlying previously "unexplained" hits, in addition to identify further genetic susceptibility loci [15,16].

The combined use of conventional GWAS (1D) analysis with LMI (2D) and 3D genomic approaches has allowed enhancing the discovery of novel candidate variants involved in PC genetic susceptibility (**Fig. 1**). As high-throughput technologies have produced large amounts of publicly-available data from cell types and tissues, these resources represent a valuable approach to perform an *in-silico* functional validation of prioritized variants using novel criteria, as well as for functional interpretation of genetic findings. Importantly, here we identified several new variants located in genes for which there is functional evidence of their implication in the biology of pancreatic acinar cells. Among them are a novel independent variant in *NR5A2,* a multi-hit region in *CASC8*, and three new PC loci in *SIAH3*, *CTRB2/BCAR1*, and *XBP1*, all of them with strong *in-silico* functional support.

METHODS

1D Approach: PanGenEU GWAS - *Single marker association analyses*

Study population. We used the resources from the PanGenEU case-control study conducted in Spain, Italy, Sweden, Germany, United Kingdom, and Ireland, between 2009-2014 [17,18]. Eligible PC patients, men and women ≥18 years of age, were invited to participate. Eligible controls were hospital in-patients with primary diagnoses not associated with known risk factors of PC. Controls from Ireland and Sweden were population-based. Institutional Review Board approval and written informed consent were obtained from all participating centers and study participants, respectively. To increase statistical power, we included controls from the

Spanish Bladder Cancer (SBC)/EPICURO study, carried out in the same geographical areas where PanGenEU Study was conducted. Characteristics of the study populations are detailed in **Additional file 1: Table S1**.

Genotyping and quality control of PanGenEU study subjects. DNA samples were genotyped using the Infinium OncoArray-500K [19] at the CEGEN facility (Spanish National Cancer Research Centre, CNIO, Madrid, Spain). Genotypes were called using the GenTrain 2.0 cluster algorithm in GenomeStudio software v.2011.1.0.24550 (Illumina, San Diego, CA). Genotyping quality control criteria considered the missing call rate, unexpected heterozygosity, discordance between reported and genotyped gender, unexpected relatedness, and estimated European ancestry <80%. After removing samples that did not pass the quality control filters, duplicated samples, and individuals with incomplete data regarding age of diagnosis/recruitment, 1,317 cases and 700 controls were available for the association analyses. SNPs in sex chromosomes and those that did not pass the Hardy-Weinberg equilibrium (*p*-value<10^{-6}) were also discarded. Overall, 451,883 SNPs passed the quality control filters conducted before the imputation.

Genotyping and quality control of SBC/EPICURO controls. Genotyping of germline DNA was performed using the Illumina 1M Infinium array at the National Cancer Institute (NCI) Core Genotyping Facility as previously described [20], which provided calls for 1,072,820 SNP genotypes. We excluded SNPs in sex chromosomes, those with a low genotyping rate (<95%), and those that did not pass the Hardy-Weinberg equilibrium threshold. In addition, the exome of 36 controls was sequenced with the TruSeq DNA Exome and a standard quality control procedure both at the SNP and individual level was applied: SNPs with read depth <10 and those that did not pass the tests of base sequencing quality, strand bias or tail distance bias, were considered as missing and imputed (see *Imputation* section for further details). Overall, 1,122,335 SNPs were available for imputation from a total of 916 additional controls.

Imputation. Imputation was performed at the Wellcome Sanger Institute (Cambridge, UK) and at CNIO for the PanGenEU and the SBC/EPICURO studies, respectively. Imputation of missing genotypes was performed using IMPUTE v2 [21], and genotypes of SBC/EPICURO controls were pre-phased to produce best-guess haplotypes using SHAPEIT v2 software [22]. For both PanGenEU and EPICURO studies, the 1000 G (Phase 3, v1) reference data set was used [23].

Association analyses. A final set of 317,270 common SNPs (MAF>0.05) that passed quality control in both studies and showed comparable MAF across genotyping platforms was used. We ensured the inclusion of the 40 variants previously associated with PC risk in individuals of Caucasian origin compiled in GWAS Catalog [5]. Logistic regression models [24] were computed assuming an additive mode of inheritance for the SNPs, adjusted for age at PC diagnosis or at control recruitment, sex, the area of residence [Northern Europe (Germany and Sweden), European islands (UK and Ireland), and Southern Europe (Italy and Spain)], and the first 5 principal components (PCA) calculated with *prcomp* R function based on the genotypes of 32,651 independent SNPs, (J Tyrer, personal communication) to control for potential population substructure.

Validation of the novel GWAS hits. To replicate the top 20 associations identified in the Discovery phase, we performed a meta-analysis using risk estimates obtained in previous GWAS studies from the Pancreatic Cancer Cohort Consortium (PanScan: https://epi.grants.cancer.gov/PanScan/) and the Pancreatic Cancer Case-Control Consortium (PanC4: http://www.panc4.org/), based on 16 cohort and 13 case-control studies, respectively. Details on individual studies, namely PanScan I, PanScan II, PanScan III, and PanC4, have been described elsewhere [6-9]. Genotyping for PanScan studies was performed at the NCI Cancer Genomic Research Laboratory using HumanHap550v3.0, and Human 610-Quad genotyping platforms for PanScan I and II, respectively, and the Illumina Omni series arrays

for PanScan III. Genotyping for PanC4 was performed at the Johns Hopkins Center for Inherited Disease Research using the Illumina HumanOmniExpressExome-8v1 array. PanScan I/II datasets were imputed together while PanScan III and PanC4 were each imputed independently using the 1000 G (Phase3, v1) reference data set [23] and IMPUTE2 [21]. Association models were adjusted for study (PanScan I and II), geographical region (for PanScan III), age, sex, and PCA of population substructure (5 PCA for PanScan I+II, 6 for PanScan III) for PanScan models, and for study, age, sex and 7 PCA population substructure for PanC4 models. Summary statistics from PanScanI/II, PanScan III and PanC4 were used for a meta-analysis using a random-effects model based on effect estimates and standard errors with the metafor R package [25].

2D Approach: Local Moran Index.

***Local Moran's Index calculation*.** The LMI was obtained for each SNP considered in the GWAS (n=317,270) using the summary statistics resulting from the association analyses as follows. First, we standardized the OR of each SNP after referring it to the risk-increasing allele (i.e., OR>1) using the inverse of the normal distribution. Then, we calculated the weight matrix containing the linkage disequilibrium (r^2) as proxy for the distance between each SNP and each of its neighboring SNPs (+/- 500kb). SNPs present within this window were matched by MAF to maximize the chance that haplotypes match. Therefore, the LMI for *i*-th SNP was calculated as:

$$LMI_i = z_i * \sum \frac{z_j * r_{i,j}^2}{\sum r_{i,j}^2},$$

where LMI_i is the LMI value for the *i*-th SNP; z_i is the OR value for the *i*-th SNP, obtained from the inverse of the normal distribution of ORs for all SNPs; z_j is the OR for the *j*-th SNP

within the physical distance and MAF-matched defined bounds; and $r_{i,j}^2$ is the LD value, measured by r^2, between the *i*-th SNP and the *j*-th SNP [26].

The LMI score could be estimated for 98.8% of the SNPs in our dataset, as 1.2% of the SNPs were not genotyped in the 1000 G (Phase 3, v1) reference data set [23] or had a MAF<1% in the CEU European population (n=85 individuals, phase 1, version 3). We then discarded the SNPs that: (1) had a negative LMI, meaning either that surrounding SNPs and target SNP have largely different ORs or that they are in linkage equilibrium and, therefore, do not pertain to the same cluster; or (2) had a positive LMI, i.e. target and surrounding SNPs have similar ORs, but the SNP came from the bottom 50% tail of the distribution of the ordered transformed OR distribution.

To assess the usefulness of the LMI score for SNP prioritization, we ran two benchmarking tests. First, we evaluated whether the GWAS Catalog PC-associated SNPs known to be associated with PC in European populations (GWAS Catalog, n=40 [5]) had a LMI value higher than expected. Then, we assessed how many of the previously reported loci were also identified according to the LMI out of the 30 independent signals of ≥1 SNPs. Further details can be found in **Additional file 1: Supplementary methods**.

3D Approach: Hi-C pancreas interaction maps and interaction selection.

The 3D Hi-C interaction maps for both healthy pancreas tissue [27] and for a pancreatic cancer cell line (PANC-1) were generated using TADbit as previously described [28]. Briefly, Hi-C FASTQ files for 7 replicas of healthy pancreas tissue were downloaded from GEO repository (Accession number: GSE87112; Sequence Read Archive Run IDs: SRR4272011, SRR4272012, SRR4272013, SRR4272014, SRR4272015, SRR4272016, SRR4272017) and for PANC-1 FASTQ, files were available from ENCODE (Accession number: ENCSR440CTR). Merged FASTQ files of the 7 healthy samples and those of PANC-1 were

mapped against the human reference genome hg19, parsed and filtered with TADbit to get the final number of valid interacting read pairs (99,074,082 and 287,201,883 valid interaction pairs, respectively). From this set we built chromosome-wide interaction matrices at 40kb resolution. The HOMER package [29] was used to detect significant interactions between bins using the –center and --maxDist 2000000 parameters. Using HOMER's default parameters, the final number of nominally significant (p-value ≤ 0.001) interactions was 41,833 for the healthy dataset and 357,749 for the PANC-1 dataset. To further filter the interactions, we retained those that passed a Bonferroni corrected threshold $<1x10^{-5}$, resulting in 6,761 for the healthy sample, (16.2% top interactions from those originally selected by HOMER default parameters). To make it comparable, we also kept the top 16.2 % interactions identified in PANC-1, resulting in 57,813 significant interactions.

Functional *in-silico* analysis.

An exhaustive *in-silico* analysis was conducted for associations with p-values$<1\times10^{-4}$ in the PanGenEU GWAS (N=143) and for the top 0.5% loci according to their LMI (N=510) (**Additional file 1: Fig. S1**). Bioinformatics assessments included evidence of functional impact [30-32], annotation in overlapping genes and pathways [31], methylation quantitative trait locus in leukocyte DNA from a subset of the PanGenEU controls (mQTLs), expression QTL (eQTLs) in normal and tumoral pancreas (GTEx and TCGA, respectively) [33,34], annotation in PC-associated long non-coding RNA (lncRNAs) [35], protein quantitative trait locus analysis in plasma (pQTLs) [36], overlap with regulatory chromatin marks in pancreatic tissue obtained from ENCODE [37], association with relevant human diseases [38], and annotation in differentially open chromatin regions (DORs) in human pancreatic cells [39]. We also investigated whether prioritized variants had been previously associated with PC comorbidities or other types of cancers [5].

We also computed the *credible sets* (calculated following the procedure in [40]; code at https://github.com/hailianghuang/FM-summary), with an r^2 >0.1, physical distance +/-500 kb and up to a posterior probability of 0.99) for the variants prioritized by the 1D (N=143 SNPs) and the 2D (510 SNPs) approaches within a 1Mb window.

In addition to the *in-silico* functional analyses at the variant level, we conducted enrichment analyses at the gene level using the FUMAGWAS web tool [38] and investigated whether our prioritized set of genes appeared altered at the tumor level in a collection of pancreatic tumor samples [41]. Methodological details of all bioinformatics analyses conducted are described in detail in **Additional file 1: Supplementary methods**.

RESULTS

1D Approach: PanGenEU GWAS - Single marker association analyses

We performed a GWAS including data from 1,317 patients diagnosed with PC (cases) and 1,616 control individuals from European countries. In addition to the genotyped SNPs that passed the quality control, we considered the imputed genotypes for previously reported PC-associated hits not included in the OncoArray-500K (19 SNPs with info score ≥0.91). In all, 317,270 SNPs were tested (**Additional file 1: Fig. S2**) with little evidence of genomic inflation (**Additional file 1: Fig. S3**).

Replication of previously reported GWAS hits. Of the 40 previously GWAS-discovered variants associated with PC risk in European ancestry populations [5], 17 (42.5%) were replicated with nominal *p*-values<0.05. For all 17, the associations were in the same direction as in the primary reports (**Additional file 1: Table S2**). Among them, we replicated *NR5A2*-rs2816938 and *NR5A2*-rs3790844, a gene for which extensive experimental evidence supporting a role in PC has been acquired. We also observed significant associations for seven variants tagging *NR5A2* previously reported in the literature [7-10,42]. Replicated GWAS hits

included *LINC00673*-rs7214041, reported to be in complete LD with *LINC00673*-rs11655237 [11], previously shown to be a PC-associated variant [9]. At the GWAS significance level, we also replicated *TERT*-rs2736098 [8,11].

The top 20 PanGenEU GWAS hits: Validation in independent populations. The risk estimates of the top 20 variants in the PanGenEU GWAS were included in a meta-analysis with those derived from PanScanI+II, PanScan III, and PanC4 consortia GWAS, representing a total of 10,357 cases and 14,112 controls (**Additional file 1: Table S3**). PanGenEU identified a new variant in *NR5A2* associated with PC (*NR5A2*-rs3790840, metaOR=1.23, *p*-value=5.91x10^{-6}) which is in moderate LD with *NR5A2*-rs4465241 (r^2=0.45, metaOR=0.81, *p*-value=3.27x10^{-10}) and had previously been reported in a GWAS pathway analysis [42]. *NR5A2*-rs3790840 remained significant (*p*-value<0.05) when conditioned on *NR5A2*-rs4465241, on *NR5A2*-rs3790844 plus *NR5A2*-rs2816938, and even on the 13 *NR5A2* GWAS hits reported in the literature, indicating that *NR5A2*-rs3790840 is a new, distinct, PC risk signal. The SKAT-O [43] (seqMeta R package: https://rdrr.io/cran/seqMeta/man/skatOMeta.html), a gene-based analysis considering all significant *NR5A2* hits plus *NR5A2*-rs3790840, yielded a significant association (*p*-value=8.9x10^{-4}). Furthermore, in a case-only analysis conducted within the PanGenEU study, the overall *NR5A2* variation was associated with diabetes (*p*-value=6.0x10^{-3}), suggesting an interaction between both factors in relation to PC risk.

While not replicated in the meta-analysis or not in the top 20 SNPs, other variants of interest identified by the 1D-approach are located in *SETDB1*, *FAM63A*, *SCTR, SEC63, CASC8,* and *RPH3AL* loci (**Table 1**). Their potential functionality is commented below.

2D-Approach: Genomic spatial integration

We scaled up from the single-SNP (1D) to the genomic region (2D) association analysis by considering both genomic distance (LD) between variants and the magnitude of the

association (OR) with the variants. We calculated a LMI score and selected those SNPs with positive LMI or within the top 50% of OR values resulting in a final set of 102,146 SNPs. The LMI scores and *p*-values for these variants showed a direct correlation (Spearman r=0.62; *p*-value=2.2×10^{-16}, **Additional file 1: Fig. S4**). To assess the versatility of LMI, we ran two benchmarks based on the MAFs and the ORs. Out of the 30 PC independent signals (r^2<0.2) derived from the GWAS Catalog, 22 were present in our 102,146 selected set. The observed median rank position for the 22 PC signals in this list was 22,640, an average position significantly higher than that of 10,000 randomly selected sets of the same size (one tail *p*-value=0.0013) (**Additional file 1: Fig. S5**). Moreover, out of the PC genomic loci, LMI was able to capture those reported by at least two studies (21 out of 30 PC independent genomic loci).

An LMI-enriched variant set was generated by selecting the top 0.5% of SNPs according to their LMI scores (LMI≥5.1071) resulting in 510 SNPs, which included 29 out of the 143 SNPs prioritised by the 1D-approach (**Additional file 2: Table S4**). We compared the MAF of the independent SNPs (r2<0.2) (see **Additional file 1: Supplementary methods**) prioritised by the 1D-approach (N=97/143) against the top 97 independent variants, out of 196 independent signals for the 510 SNPs selected by LMI. Notably, the LMI-identified SNPs had a lower MAF than GWAS-identified variants: 0.07 (SD=0.03) vs. 0.24 (SD=0.13) (Wilcoxon statistic *p*-value<2.2×10^{-16}) (**Additional file 1: Fig. S6**). In line with this observation, the average OR for the LMI-based SNPs was significantly higher than that for the GWAS-based SNPs (1.46 *vs.* 1.32, respectively, Wilcoxon statistic *p*-value=1.63×10^{-10}).

The Manhattan plot of the LMI score across the genome displays the hits identified through this approach (**Additional file 1: Fig. S7**). Among the 0.5% top LMI prioritized variants (N=510), there were 8 SNPs in *NR5A2*, including the novel PanGenEU GWAS identified variant (rs3790840). All of them showed a high LMI-score (>6.859) what further

endorses this approach. Other variants of interest identified by the 2D-approach are in *SETDB1, FAM63A/MINDY1, GPRC6A, RFX6, CASC8, CDKN2A, KDM4C, ROR2, MS4A5, SIAH3, LRRC36,* and *CTRB2/BCAR1* loci (**Table 1**). Their potential functionality is discussed below.

A total of 199 *credible sets* were identified among the 510 LMI-based SNP. Of them, 118 (60%) contained the SNP with the lowest *p*-value in the region. Moreover, we observed an enrichment of SNPs with low *p*-values in the 1Mb region for the LMI-based SNP set that was even higher among the 118 *credible sets* (**Additional file 1: Fig. S8**).

3D-Approach: genomic interaction analysis

To gain further insight into the biological function of the 624 candidate SNPs prioritised using the 1D- and 2D-approaches, and to identify additional PC genetic susceptibility loci, we focused on a set of 6,761 significant chromatin interactions (p-values$\leq$1x10^{-5}) identified using Hi-C interaction pancreatic tissue maps at 40Kb resolution [27]. Throughout the rest of the text, we will refer to the chromatin interaction component containing the prioritized SNP as "bait" and to its interacting region as "target". In total, 54 target loci overlapping with 37 genes interacted with bait regions harboring 76/624 (12.1%) SNPs (**Additional file 3: Table S5**).

As a proof of concept of the utility of the 3D-approach to identify novel PC genetic susceptibility loci, we highlight a target region (22:29,197,371-29,237,371bp, p-value=1.3x10^{-9}) interacting with an intronic region of *TTC28* (bait: 22:28,602,352-28,642,352bp) that including four LMI-selected SNPs (rs9620778, rs9625437, rs17487463 and rs75453968, all in high LD, r^2>0.95, in CEU population) (**Fig. 2**). Other loci of interest identified by the 3D-approach are in *SIAH3, CTRB2,* and *MN1* loci (**Table 1**). Their potential functionality is commented below.

Functional in-silico validation

We performed a systematic and exhaustive *in-silico* functional analysis of SNPs prioritized by GWAS (N=143) and LMI (N=510) at the variant, gene, and pathway levels (**Fig. 1** and **Additional file 1: Fig. S1**).

Assessment of potential functionality of the variants. The evidence for potential functionality of the most relevant SNPs for each of the approaches used is reviewed here and summarized in **Table 1**, **Additional file 1: Supplementary methods**, and **Additional file 2: Table S4** and **Additional file 4: Table S6**.

Among the 143 variants prioritized in the 1D-approach, we highlight those in *CASC8* (8q24.21) (**Fig. 3**): 27 variants with *p*-values $<1x10^{-4}$ organized in four LD-blocks, 9 of which were also captured in the 2D approach. The *CASC8* locus is amplified in 5% of PC and codes for a non-protein coding RNA overexpressed in tumor vs normal pancreatic tissue (Log2FC=1.25, *p*-value=$2.29x10^{-56}$). *CASC8* also overlaps with a PC-associated lncRNA [35], suggesting that genetic variants in *CASC8* may contribute to the transcriptional program of pancreatic tumor cells. All *CASC8* variants were also associated with differential leukocyte methylation (mQTL) of *RP11-382A18.1*-cg25220992 in our PanGenEU population sample. Moreover, 20 of them were associated with differential methylation of cg03314633, also in *RP11-382A18.1*. Twenty-three of the variants overlapped with at least one histone mark in either endocrine or exocrine pancreatic tissue. Alterations in *CASC8* significantly co-occur with alterations in *TG* (adjusted *p*-values<0.001), also associated with PC in our GWAS, which is located downstream.

Three of the variants prioritized for *in-silico* analysis in the 1D-approach (but not in the 2D-approach) are located in genes involved in pancreatic function: rs1220684 is in *SEC63*, coding for a protein involved in endoplasmic reticulum (ER) function and ER stress response [44]; rs7212943, a putative regulatory variant, is in *NOC2/RPH3AL*, a gene involved in

exocytosis in exocrine and endocrine cells [45]; and rs4383344 is in *SCTR,* which encodes for the secretin receptor, selectively expressed in ductal cells, involved in the regulation of bicarbonate, electrolyte, and volume secretion. Interestingly, secretin regulation is affected by *H. pylori* which has been suggested as a PC risk factor [46]. High expression of *SCTR* has also been reported in PC [47].

Two variants in high LD (r^2=0.92) and potentially relevant at the functional level are in 1q21.3 (*SETDB1*-rs17661062 and *FAM63A*-rs59942146). *SETDB1* has recently been reported to be required for formation of PC in mice by inhibiting p53-mediated apoptosis [48] and *FAM63A/MINDY1* has been found to interact significantly with diabetes (duration ≥3 years) in a meta-analysis on PC risk conducted within the PanC4 and PanScan consortia [49]. Interestingly, these two variants were also associated with an increased methylation of the cg17724175 in *MCL1*. High mRNA expression of this gene has been associated with poor survival [50], and Mcl-1 has been explored to selectively radiosensitize PC cells [51]. Importantly, these two variants were also the top two prioritized by the 2D-approach with a LMI score ≥16.87 (**Additional file 2: Table S4**).

Using the 2D-approach, we prioritized several other regions with potential functional relevance (**Table 1**, **Fig. 4**, **Additional file 1: Fig. S7**, **Additional file 2: Table S4**). In chromosome 6, we identified rs6907580 (LMI=8.93), a well-characterized stop-gain - and likely disease-causal variant (CADD-score=35) - in exon 1 of *GPRC6A* (*G protein-coupled receptor family C group 6 member A*). *GPRC6A* is expressed in pancreatic acinar, ductal, and β-cells, it participates in endocrine metabolism [52] and it has been involved in pancreatitis using mouse models [53]. Downstream in the same region, LMI approach also identified rs17078438 (LMI=8.90) in *RFX6*, a pancreas-specific gene involved in pancreatic development [39].

Other potentially functional SNPs relevant to PC and prioritized in the 2D-approach comprised 6 SNPs (LMI≤5.60) in the vicinities of *CDKN2A/p16*, a gene that is almost universally inactivated in PC [54] and that is mutated in some hereditary forms of PC [55, 56], with three variants (LMI≤5.48) in *CDKN2A-AS1* and two (LMI≤5.88) in *CDKN2B/p15*, other important cell-cycle regulators; three variants in *KDM4C* (LMI≤11.27), a Lys demethylase 4C highly expressed in PC [57]; and two SNPs tagging *ROR2* (LMI≤5.57), a member of the *Wnt* pathway that plays a relevant role in PC [58].

Another region, in chromosome 16, comprises *BCAR1*-rs7190458, a variant with a relevant role in PC [59] reported in two previous GWAS [8,11], as well as a novel SNP (rs13337397, LMI=6.03) located in the first exon of *BCAR1*. Both SNPs are in low LD (r^2=0.36). This second SNP is intergenic to *CTRB2* and *BCAR1* (**Table 1, Fig. 4, Additional file 1: Fig. S7, Additional file 2: Table S4**). While *BCAR1* is ubiquitous, *CTRB2* code for chymotrypsinogens B2, a protease expressed exclusively in the exocrine pancreas; genetic variation therein has been previously associated with alcoholic pancreatitis [60] and type-2 diabetes [61,62]. The expression of that gene is reduced in tumors vs. normal tissue [63]. Interestingly, *CTRB2* was also identified by the 3D-approach, too. A still undergoing work done by Dr Amundadottir, a co-author of this paper, confirms that a 585bp structural variant in *CTRB2* inhibits chymotrypsin B2 activity and secretion, and confers risk of pancreatic cancer (oral communication).

According to their deleteriousness CADD score [32], we highlight two variants in coding transcripts: *MS4A5*-rs34169848 in chr11:60,197,299 (CADD-score=24.4) and *LRRC36*-rs8052655 in chr16:67,409,180 (CADD-score=24.4). CADD-scores of such magnitude are likely to correspond to disease-causal variants [64].

The 3D approach highlighted *XBP1* as a target region. As said, this target region including the *XBP1* promoter interacts with four of the LMI-selected SNPs (**Fig. 2**) that are in

moderate LD with rs16986177. The alternative allele (T) of this SNP is associated with a decreased expression of *XBP1* in normal pancreas in GTEx (-0.19, *p*-value=$1.3x10^{-4}$) and with an increased risk of PC in our GWAS (OR=1.28, *p*-value=$8.71x10^{-3}$). Expression of *XBP1* is reduced in PC samples from TCGA, compared to normal pancreas samples from GTEx (Log2FC=-1.561, *p*-value=$1.72x10^{-34}$). Chip-Seq data of all pancreatic samples available in ENCODE, as well as PANC-1 pancreatic cancer cells (see Methods), allowed us to find that, in comparison to normal pancreas, the H3K27Ac mark present in the *XBP1* promoter is completely lost in PANC-1 cells and is reduced in a sample of a Pancreatic Intraepithelial Neoplasia 1B, a PC precursor (**Fig. 2**). To further characterize the bait and promoter regions upstream of *XBP1,* we ran eight chromatin states using ChromHMM [65] (**Additional file 1: Supplementary methods**). We observed a clear loss of enhancers/weak promoters in the corresponding target regions in the precursor lesions and in PANC-1 cells. This loss of activity is in line with the observation that *XBP1* expression is reduced in cancer. Moreover, small enhancers are also lost in the bait region of the aforementioned samples. The 3D maps for this region revealed loss of 3D contact in PANC-1 cells (**Fig. 2**).

Gene set enrichment analyses (GSEA). We performed GSEA of the genes harboring the SNPs prioritized using the 1D- and 2D- approaches. Six chromosomal regions were significantly enriched among the 81 genes harboring the 143 prioritized SNPs in the 1D approach (**Additional file 5: Table S7**). GSEA for the gene-trait associations reported in the GWAS Catalog yielded 29 enriched traits (**Additional file 5: Table S7**). The most relevant GWAS traits significantly enriched were 'Pancreatic cancer', 'Lung cancer', 'Prostate cancer', 'Uric acid levels', 'Obesity-related traits', and 'Major depressive disorder'. We also performed a network analysis to visualize the relationships between the enriched GWAS traits and the prioritized genes using the *igraph* R package [66]. Twelve densely connected subgraphs were identified via random walks (**Fig. 5**). Interestingly, 'pancreatic cancer' and 'uric acid levels'

GWAS traits were connected through *NR5A2*, which is also linked to 'chronic inflammatory diseases' and 'lung carcinoma' traits. *NR5A2* is an important regulator of pancreatic differentiation and inflammation in the pancreas [67].

GSEA of the genes harboring the variants included in the *credible sets* corresponding to the 2D-approach revealed enrichment in 'Pancreatic cancer' as well as other GWAS traits related to PC risk factors, including alcoholic chronic pancreatitis, Type 2 diabetes, body mass index and waist-to-hip ratio adjusted for body mass index, and HDL cholesterol (**Additional file 6: Table S8**). These findings lend support to the validity of the 2D-approach as a tool to identify disease-relevant genetic variants.

Pathway enrichment analysis. The genes prioritized in the 1D approach were significantly enriched in 112 Gene Ontology-Biological function (GO:BP) terms (adjusted p-values<0.05, with minimum of three genes overlapping), 7 Cellular component GO terms (GO:CC), and 11 Molecular functions (GO:MF) terms (**Additional file 5: Table S7**). Importantly, GO terms relevant to exocrine pancreatic function were overrepresented. Three KEGG pathways were significantly enriched with ≥2 genes from our prioritized set, including "Glycosaminoglycan biosynthesis heparan sulfate" (*adj-p*=$3.86\text{x}10^{-3}$), "ERBB signaling pathway" (*adj-p*=$3.73\text{x}10^{-2}$) and "Melanogenesis" (*adj-p*=$3.73\text{x}10^{-2}$) (**Additional file 5: Table S7**). Pathways enriched with the genes prioritized in the 2D-approach included GO terms related to the nervous system and G protein-coupled receptor signaling. Interestingly, one of the hallmarks of PC is perineural invasion. Because the standard databases generally lack pathways related to acinar pancreatic function, we generated several curated gene-sets and assessed their enrichment among the SNPs/genes prioritized in the 1D- and 2-D approaches. We found an overrepresentation of LMI-genes in a signature including transcription factors differentially expressed in normal pancreas (GTEx). This signature was also enriched with genes prioritized in the other two approaches

used in our study (including 11 overlapping genes: *SETDB1*, *LHX4*, *NR5A2*, *ZBED6*, *ELK4*, *SIM1*, *RFX6*, *KLF14*, *ZNF32*, *ZNF133,* and *XBP1*).

In summary, the *in-silico* functional analysis revealed a remarkable enrichment of pathways related to the function of acinar and ductal cells, including SNPs associated with novel genes in these pathways.

DISCUSSION

To overcome some of the limitations of standard GWAS analyses, we have expanded the scope of genomic studies of PC susceptibility to include novel approaches that build on spatial genome autocorrelations of LMI and 3D chromatin contacts. An in-depth *in-silico* functional analysis leveraging available genomic information from public databases allowed us to prioritize new candidate variants with strong biological plausibility in well-established (i.e. *NR5A2*) as well as in novel (i.e. *XBP1*) genes playing a key role in acinar function (**Table 1**). We have thus reached a novel landscape on the inherited basis of PC and have paved the way to the application of a similar strategy to any other human disease or interest.

This is the first PC GWAS involving an exclusively Europe-based population sample. Of the previously reported European ancestry population GWAS hits, 42.5% were replicated, supporting the methodological soundness of the study. The lack of replication of other PC GWAS hits may be explained by variation in the MAFs of the SNPs among Europeans, population heterogeneity, differences in the genotyping platform used, and differences in calling methods applied, among others. This result emphasizes that statistical significance for GWAS-SNPs is largely dependent on MAF and the statistical power of the study, highlighting this as a major limitation of classical GWAS analyses.

We applied the LMI (2D-approach) for the first time in the genomics field. LMI captured a new dimension of signals independent from MAF and the statistical power of the

study (**Additional file 1: Fig. S6**). The benchmarking tests evidenced that LMI prioritize SNPs on the basis of OR that were largely present in *credible sets* (**Additional file 1: Fig. S8**). We replicated 6.4% of the previous reported GWAS Catalog signals for PC in European populations by considering the top 0.5% LMI variants, a LMI threshold that is overly conservative, given that many of the GWAS Catalog-replicated signals have lower LMI than the cut-off value we selected. The ability of LMI to prioritize low MAF SNPs, unlike the GWAS approach, may also explain the low replicability rate. LMI helps to identify signals within genomic regions by scoring lower those regions that do not maintain LD structure.

The 3D genomic approach identified a highly potential important chromatin interacting region in *XBP1*. This is a potential candidate detected through a previously uncharacterized "bait" SNP. These findings are particularly important considering the overwhelming evidence of a major role of ER stress and unfolded protein responses in acinar function- two highly relevant processes to acinar homeostasis due to their high protein-producing capacity of these cells - and it plays an important role in pancreatic regeneration [68]. In addition, genetic mouse models have unequivocally shown that *Xbp1* is required for acinar homeostasis and pancreatic ductal adenocarcinoma, the most common form of PC, can be initiated from acinar cells [69]. Overall, these analyses indicate that the SNPs interacting in 3D space with the *XBP1* promoter could contribute to the differential expression of the gene associated with malignant transformation. These findings provide proof of concept that 3D genomics can contribute to identify further susceptibility loci and to decipher the biological relevance of orphan SNPs. Similar results were found with other LMI selected SNPs associated with their target genes only by detecting significant spatial interactions between them (**Additional File 3: Table S5**).

To shed light into the functionality of the newly identified variants, we applied novel post-GWAS approaches to interrogate several databases at the SNP, gene, and pathway levels. We found sound evidence pointing to the functional relevance of several variants prioritized by

the 1D and 2D approaches (**Additional files 2 and 4: Tables S4 and S6**, respectively, and **Additional file 1: Supplemental methods**). The importance of the multi-hit *CASC8* region (8q24.21) is further supported by *in-silico* functional analyses as well as by its previous associations with PC at the gene level [35]. In particular, 12/27 SNPs identified in *CASC8* were annotated as regulatory variants. None of the *CASC8* hits were in LD with *CASC11*-rs180204, a GWAS hit previously associated with PC risk, which is ~205 Kb downstream [10]. *CASC8*-rs283705 and *CASC8*-rs2837237 (r^2=0.68) are likely to be functional with a score of 2b in RegulomeDB (TF binding + any motif + DNase Footprint + DNase peak). *CASC8*-rs1562430, in high LD (r^2>0.85) with 18 *CASC8* prioritized variants, has been previously associated with other cancers (breast, colorectal, and stomach) [70]. None of the prostate cancer-associated SNPs in *CASC8* overlapped with the 27 identified variants in our study. The fact that this gene has not been reported previously in other PC GWAS could be due to the different genetic background of the study populations or to an overrepresentation of the variants tagging *CASC8* in the Oncoarray platform used here.

In addition to confirming SNPs in *TERT*, we found strong evidence for the participation of novel susceptibility genes in telomere biology (*PARN*) and in the post-transcriptional regulation of gene expression (*PRKCA* and *EIF2B5*) (**Additional File 1: Supplemental methods**). Our study also expands the landscape of variants and genes involved in exocrine biology, including *SEC63*, *NOC2*/*RPH3AL* and *SCRT* whose products participate in acinar function and possibly in acinar-ductal metaplasia, a PC pre-neoplastic lesion [71].

KEGG pathway enrichment analysis further validated our results being involved in important pathways for PC, including “Glycosaminoglycan biosynthesis heparan sulfate” and “ERBB signaling pathway”. Heparan sulfate (HS) is formed by unbranched chains of disaccharide repeats which play roles in cancer initiation and progression [72]. Interestingly, the expression of HS proteoglycans increases in PC [73] and related molecules, such as

hyaluronic acid, are important therapeutic targets in PC [74,75]. ERBB signaling is important both in PC initiation and as a therapeutic target [76].

The enrichment analysis indicates that urate levels, depression, and body mass index - three GWAS traits previously reported to be associated with PC risk - were enriched in our prioritized gene set. Urate levels have been associated with both PC risk and prognosis [77, 78]. In addition, patients with lower relative levels of kynurenic acid have more depression symptoms [79]. PC is one of the cancers with the highest occurrence of depression preceding its diagnosis [80]. Furthermore, body mass index has been previously associated with PC risk in diverse populations [81-83] and it has been suggested that increasing PC incidence may be partially attributed to the obesity epidemic. Insulin resistance is one of the mechanisms possibly underlying the obesity and PC association, through hyperinsulinemia and inflammation [84].

The post-GWAS approach used has limitations that should be addressed in future studies. For example, our study has a relatively small sample size, some imbalances regarding gender and geographical areas, and the Hi-C maps that we used have limited resolution (40 kb). To account for population imbalances, regression models were adjusted for gender and for country of origin, as well as for first five principal components. The study of a European-only population allows reducing the population heterogeneity not only at the genetic but also at the non-genetic level. This is particularly advantageous given the novel nature of the analysis performed here and the relatively small sample size of our study. More so, LMI is based on both the summary statistics and LD structure; therefore, it was important to test its validity for the first time in a more homogeneous population, with individuals sharing a more consistent LD pattern. It is now warranted to extend this approach to more generalizable multi-ethnic populations.

Our study has many other strengths: a standardized methodology was applied in all participating centers to recruit cases and controls, to collect information, and to obtain and

process biosamples; state-of-the-art methodology was used to extend the identification of variants, genes, and pathways involved in PC genetic susceptibility. Most importantly, the combination of GWAS, LMI and 3D genomics to identify new variants is completely novel and has proven crucial to refine results, reduce the number of false positives, and establish whether borderline GWAS *p*-value signals could be true positives. These three strategies, together with an in-depth *in*-silico functional analysis, offer a comprehensive approach to advance the study of PC genetic susceptibility.

CONCLUSIONS

We present a novel multilayered post-GWAS assessment on genetic susceptibility to PC. We showed that the combined use of conventional GWAS (1D) analysis with LMI (2D) and 3D genomic approaches allows enhancing the discovery of novel candidate variants involved in PC. Importantly, several of the new variants are located in genes relevant to the biology and function of acinar and ductal cells.

This multi-step strategy, combined with an in-depth *in*-silico functional analysis, offers a comprehensive approach to advance the study of PC genetic susceptibility and could be applied to other diseases.

LIST OF ABBREVIATIONS

PC: Pancreatic Cancer.

GWAS: Genome Wide Association Study.

LMI: Local Moran's Index.

OR: Odds Ratio.

LD: Linkage Disequilibrium.

MAF: Minor Allele Frequency.

SBC: Spanish Bladder Cancer

NCI: National Cancer Institute

PCA: Principal Components

mQTL: methylation Quantitative Trait Locus

eQTL: expression Quantitative Trait Locus

lncRNA: long non-coding RNA

pQTLs: protein Quantitative Trait Locus analysis

DORs: Differentially Open chromatin Regions

CADD: Combined Annotation Dependent Depletion

ER: Endoplasmic Reticulum

GO:BP: Gene Ontology Biological Function.

GO:CC: Gene Ontology Cellular Component.

GO:MF: Gene Ontology Molecular Function.

HS: Heparan sulfate.

DECLARATIONS

Ethics approval and consent to participate: IRB ethical approval and written informed consent was obtained by all participating centres contributing to PanGenEU, SBC/EPICURO, PanC4, and PanScan I-III International consortia, and study participants, respectively. The study was conducted in accordance with the Helsinki Declaration.

Consent for publication: Not applicable.

Availability of data and materials: GWAS summary statistics generated in this study are available in the GWAS catalog repository with the accession numbers GCST90011857 (ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011857) [85] and GCST90011858 (ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011858) [86]. Code for performing the association analysis and metaanalysis is available at https://github.com/EvangelinaLdM/Multilayered_postGWAS_PanGenEU [24]. Code for calculating LMI is freely available at https://github.com/pollicipes/Local-Moran-Index-1D [26].

Competing interests: The authors declare that they have no competing interests.

Funding: The work was partially supported by Fondo de Investigaciones Sanitarias (FIS), Instituto de Salud Carlos III, Spain (#PI061614, #PI11/01542, #PI0902102, #PI12/01635, #PI12/00815, #PI15/01573, #PI18/01347); Red Temática de Investigación Cooperativa en Cáncer, Spain (#RD12/0036/0034, #RD12/0036/0050, #RD12/0036/0073); Fundación Científica de la AECC, Spain; European Cooperation in Science and Technology - COST Action #BM1204: EUPancreas. EU-6FP Integrated Project (#018771-MOLDIAG-PACA), EU-FP7-HEALTH (#259737-CANCERALIA, #256974-EPC-TM-Net); Associazione Italiana Ricerca sul Cancro (#12182); Cancer Focus Northern Ireland and Department for Employment and Learning; and ALF (#SLL20130022), Sweden; Pancreatic Cancer Collective (PCC):

Lustgarten Foundation & Stand-Up to Cancer (SU2C #6179); Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, USA; PANC4 GWAS RO1CA154823; NCI, US-NIH (#HHSN261200800001E).

Authors' contributions: Study conception: **NM, ELM.** Design of the work**: ELM, JAR, DE, MAMR, FXR.** Data acquisition**: EMM, PGR, RTL, AC, MH, MI, XM, ML, CWM, JP, MOR, BMB, AT, AF, LMB, TCJ, LDM, TG, WG, LS, LA, LC, JB, EC, LI, JK, NK, MM, JM, DOD, AS, WY, JY, PanGenEU Investigators, MGC, MK, NR, DS, SBC/EPICURO Investigators, DA, AAA, LBF, PMB, PB, BBM, JB, FC, MD, SG, JMG, PJG, MG, LLM, LD, NRN, UP, GMP, HAR, MJS, XOS, LDT, KV, WZ, SC, BMW, RZSS, APK, LA, FXR, NM.** Data analysis**: ELM, JAR, LA, OL.** Interpretation of data**: ELM, JAR, LA, OL, EMM, MAMR, FXR, NM.** Creation of new software used in the work**: ELM, LA, JAR, OL, MAMR.** Drafting the work or substantively revised it**: ELM, JAR, LA, OL, TCJ, UP, HAR, APK, LA, MAMR, FXR, NM.** Approval of the submitted version (and any substantially modified version that involves the author's contribution to the study): **ALL AUTHORS.** Agreement on both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature: **ALL AUTHORS**. All authors read and approved the final manuscript.

Acknowledgements: The authors are thankful to the patients, coordinators, field and administrative workers, and technicians of the European Study into Digestive Illnesses and Genetics (PanGenEU) and the Spanish Bladder Cancer (SBC/EPICURO) studies. We also thank Marta Rava, former member of the GMEG-CNIO, Guillermo Pita and Anna González-Neira from CGEN-CNIO, and Joe Dennis and Laura Fachal from the University of Cambridge,

for genotyping PanGenEU samples, performing variant calling and SNP imputation, and editing data.

Additional files:

- Additional file 1: Supplementary methods, Tables S1-S3, Figures S1-S9.docx.

Table S1. Characteristics of the study populations

Table S2. Replication of the SNPs reported as associated with pancreatic cancer risk in European population and published in GWAS Catalog.

Table S3. Validated variants (at the nominal p-value), in PanScan and PanC4 populations, among the top 20 SNPs identified in the PanGenEU GWAS study.

Figure S1. Functional *in-silico* analysis strategy followed to identify novel genomic regions previously prioritized using the 1D, 2D and 3D approaches.

Figure S2. GWAS Manhattan plot for the PanGenEU study. The x-axis is the genomic position of each variant and the y-axis is the −log10 *p*-value obtained in the 1D analysis.

Figure S3. Q-Q plots for pancreatic cancer risk of the association results using the PanGenEU case-control study (S2a) and PanGenEU&EPICURO study populations (S2b).

Figure S4. Scatterplot of the local Moran's index (LMI) obtained in the 2D approach and the –log10 *p*-value obtained in the GWAS analysis (1D approach).

Figure S5. Results of the benchmarking test showing that the median rank position of the LMI values for the 22 pancreatic cancer signals from the GWAS Catalog is significantly higher than 10,000 randomly selected sets of the same size.

Figure S6. Minor allele frequency (MAF) distributions for the top 97 SNPs identified by LMI (in pink) and by GWAS (in blue).

Figure S7. LMI Manhattan plot for the PanGenEU study. The x-axis is the genomic position of each variant and the y-axis is the LMI value obtained in the 2D analysis.

Figure S8. Q-Q plots show significant enrichment of SNPs with low *p*-values in the variants prioritized in the 2D-approach (S8a), and in the credible sets derived from them (S8b).

Figure S9. Complementary NETWORK (in blue our input KEGG pathways; in green, the complementary pathways interconnected with them) obtained with Pathway-connector webtool.

- Additional file 2: Table S4.xlsx. List of the 510 SNPs prioritized according to their Local Moran Index (LMI) (.xlsx file).
- Additional file 3: Table S5.xlsx. List of the 76 SNPs overlapping with a chromatin interaction region (bait regions) and their 54 targets (.xlsx file).
- Additional file 4: Table S6.xlsx. Annotation and functional in silico analysis of the 143 prioritized SNPs (.xlsx file).
- Additional file 5: Table S7.xlsx. Results from the gene enrichment analysis performed with FUMA in the 1D prioritized genes (.xlsx file).
- Additional file 6: Table S8.xlsx. Results from the gene enrichment analysis performed with FUMA in the 2D prioritized genes (.xlsx file).

REFERENCES

1. Carrato A, Falcone A, Ducreux M, Valle JW, Parnaby A, Djazouli K, et al. A Systematic Review of the Burden of Pancreatic Cancer in Europe: Real-World Impact on Survival, Quality of Life and Costs. J Gastrointest Cancer. 2015;46:201–11.

2. Malvezzi M, Bertuccio P, Rosso T, Rota M, Levi F, La Vecchia C, et al. European cancer mortality predictions for the year 2015: Does lung cancer have the highest death rate in EU women? Ann Oncol. 2015;26:779–86.

3. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends - An update. Cancer Epidemiol Biomarkers Prev. 2016;25:16–27.

4. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014;74:2913–21.

5. Buniello A, Macarthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47:D1005–12.

6. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009;41:986–90.

7. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010;42:224–8.

8. Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat Genet. 2014;46:994–1000.

9. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet. 2015;47:911–6.

10. Zhang M, Wang Z, Obazee O, Jia J, Childs EJ, Hoskins J, et al. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. Oncotarget. 2016;7:66328–43.

11. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun. 2018;9:556.

12. Chen F, Childs EJ, Mocci E, Bracci P, Gallinger S, Li D, et al. Analysis of heritability and genetic architecture of pancreatic cancer: A PANC4 study. Cancer Epidemiol Biomarkers Prev. 2019;28:1238–45.

13. Anselin L. Local Indicators of Spatial Association—LISA. Geogr Anal. 1995;27:93–115.

14. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295:1306–11.

15. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med. 2015;373:895–907.

16. Montefiori LE, Sobreira DR, Sakabe NJ, Aneas I, Joslin AC, Hansen GT, et al. A promoter interaction map for cardiovascular disease genetics. Elife. 2018;7:e35788.

17. Gomez-Rubio P, Zock J-P, Rava M, Marquez M, Sharp L, Hidalgo M, et al. Reduced risk of pancreatic cancer associated with asthma and nasal allergies. Gut. 2017; doi: 10.1136/gutjnl-2015-310442.

18. Molina-Montes E, Gomez-Rubio P, Márquez M, Rava M, Löhr M, Michalski CW, et al. Risk of pancreatic cancer associated with family history of cancer and other medical conditions by accounting for smoking among relatives. Int J Epidemiol. 2018;47:473–83.

19. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The oncoarray consortium: A network for understanding the genetic architecture of common cancers. Cancer Epidemiol Biomarkers Prev. 2017;26:126–35.

20. Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat Genet. 2010;42:978–84.

21. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5:e1000529.

22. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9:179–81.

23. Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467:1061–73.

24. López de Maturana, E. Multilayered_postGWAS_PanGenEU. Github. https://github.com/EvangelinaLdM/Multilayered_postGWAS_PanGenEU (2020).

25. Viechtbauer W. Conducting meta-analyses in R with the metafor. J Stat Softw. 2010;36:1–48.

26. Rodriguez, J.A. Local Moran Index 1D. Github. https://github.com/pollicipes/Local-Moran-Index-1D (2020).

27. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 2016;17:2042–59.

28. Serra F, Baù D, Goodstadt M, Castillo D, Filion G, Marti-Renom MA. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput Biol. 2017;13:e1005665.

29. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010;38:576–89.

30. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.

31. Martín-Antoniano I, Alonso L, Madrid M, López De Maturana E, Malats N. DoriTool: A Bioinformatics Integrative Tool for Post-Association Functional Annotation. Public Health Genomics. 2017;20:126–35.

32. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47:D886–94.

33. Ardlie KG, DeLuca DS, Segrè A V., Sullivan TJ, Young TR, Gelfand ET, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015;348:648–60.

34. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, et al. PancanQTL: Systematic identification of cis -eQTLs and trans -eQTLs in 33 cancer types. Nucleic Acids Res. 2018;46:D971–6.

35. Arnes L, Liu Z, Wang J, Maurer HC, Sagalovskiy I, Sanchez-Martin M, et al. Comprehensive characterisation of compartment-specific long non-coding RNAs associated with pancreatic ductal adenocarcinoma. Gut. 2019;68:499–511.

36. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558:73–9.

37. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, et al. ENCODE data at the ENCODE portal. Nucleic Acids Res. 2016;44:D726–32.

38. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.

39. Arda HE, Tsai J, Rosli YR, Giresi P, Bottino R, Greenleaf WJ, et al. A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas. Cell Syst. 2018;7:310-322.e4.

40. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.

41. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:11.

42. Li D, Duell EJ, Yu K, Risch HA, Olson SH, Kooperberg C, et al. Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. Carcinogenesis. 2012;33:1384–90.

43. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics. 2012;13:762–75.

44. Linxweiler M, Schick B, Zimmermann R. Let's talk about secs: Sec61, sec62 and sec63 in signal transduction, oncology and personalized medicine. Signal Transduct Target Ther. 2017;2:17002.

45. Matsumoto M, Miki T, Shibasaki T, Kawaguchi M, Shinozaki H, Nio J, et al. Noc2 is essential in normal regulation of exocytosis in endocrine and exocrine cells. Proc Natl Acad Sci U S A. 2004;101:8313–8.

46. Risch HA. Etiology of pancreatic cancer, with a hypothesis concerning the role of N-nitroso compounds and excess gastric acidity. J Natl Cancer Inst. 2003;95:948–60.

47. Körner M, Hayes GM, Rehmann R, Zimmermann A, Friess H, Miller LJ, et al. Secretin receptors in normal and diseased human pancreas: Marked reduction of receptor binding in ductal neoplasia. Am J Pathol. 2005;167:959–68.

48. Ogawa S, Fukuda A, Matsumoto Y, Hanyu Y, Sono M, Fukunaga Y, et al. SETDB1 Inhibits p53-Mediated Apoptosis and is Required for Formation of Pancreatic Ductal Adenocarcinomas in Mice. Gastroenterology. 2020;S0016-5085.

49. Tang H, Jiang L, Stolzenberg-Solomon R, Arslan AA, Beane Freeman LE, Bracci P, et al. Genome-wide gene-diabetes and gene-obesity interaction scan in 8,255 cases and 11,900 controls from the PanScan and PanC4 Consortia. Cancer Epidemiol Biomarkers Prev. 2020; doi.org/10.1158/1055-9965.EPI-20-0275.

50. Castillo L, Young AIJ, Mawson A, Schafranek P, Steinmann AM, Nessem D, et al. MCL-1 antagonism enhances the anti-invasive effects of dasatinib in pancreatic adenocarcinoma. Oncogene. 2020;39:1821–9.

51. Wei D, Zhang Q, Schreiber JS, Parsels LA, Abulwerdi FA, Kausar T, et al. Targeting Mcl-1 for radiosensitization of pancreatic cancers. Transl Oncol. 2015;8:47–54.

52. Pi M, Quarles LD. Multiligand specificity and wide tissue expression of GPRC6A reveals new endocrine networks. Endocrinology. 2012;153:2062–9.

53. Zhang X, Jin T, Shi N, Yao L, Yang X, Han C, et al. Mechanisms of pancreatic injury induced by basic amino acids differ between L-arginine, L-ornithine, and L-histidine. Front Physiol. 2019;10.

54. Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. Nature. 2016;538:378–82.

55. Bartsch DK, Sina-Frey M, Lang S, Wild A, Gerdes B, Barth P, et al. CDKN2A germline mutations in familial pancreatic cancer. Ann Surg. 2002;236:730–7.

56. Lynch HT, Brand RE, Hogg D, Deters CA, Fusaro RM, Lynch JF, et al. Phenotypic variation in eight extended CDKN2A germline mutation familial atypical multiple mole melanoma-pancreatic carcinoma-prone families: The familial atypical multiple mole melanoma-pancreatic carcinoma syndrome. Cancer. 2002;94:84–96.

57. Gregory BL, Cheung VG. Natural variation in the histone demethylase, KDM4C, influences expression levels of specific genes including those that affect cell growth. Genome Res. 2014;24:52–63.

58. Morris JP, Wang SC, Hebrok M. KRAS, Hedgehog, Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma. Nat Rev Cancer. 2010;10:683–95.

59. Duan B, Hu J, Liu H, Wang Y, Li H, Liu S, et al. Genetic variants in the platelet-derived growth factor subunit B gene associated with pancreatic cancer risk. Int J Cancer. 2018;142:1322–31.

60. Rosendahl J, Kirsten H, Hegyi E, Kovacs P, Weiss FU, Laumen H, et al. Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. Gut. 2018;67:1855–63.

61. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè A V., Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44:981–90.

62. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun. 2018;9:2941.

63. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45:W98–102.

64. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.

65. Ernst J, Kellis M. ChromHMM: Automating chromatin-state discovery and characterization. Nat Methods. 2012;9:215–6.

66. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Syst. 2006; doi:10.1186/1471-2105-12-455.

67. Cobo I, Martinelli P, Flández M, Bakiri L, Zhang M, Carrillo-De-Santa-Pau E, et al. Transcriptional regulation by NR5A2 links differentiation and inflammation in the pancreas. Nature. 2018;554:533–7.

68. Hess DA, Humphrey SE, Ishibashi J, Damsz B, Lee A, Glimcher LH, et al. Extensive pancreas regeneration following acinar-specific disruption of Xbp1 in mice. Gastroenterology. 2011;141:1463–72.

69. Guerra C, Schuhmacher AJ, Cañamero M, Grippo PJ, Verdaguer L, Pérez-Gallego L, et al. Chronic Pancreatitis Is Essential for Induction of Pancreatic Ductal Adenocarcinoma by K-Ras Oncogenes in Adult Mice. Cancer Cell. 2007;11:291–302.

70. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010;42:504–7.

71. Notta F, Hahn SA, Real FX. A genetic roadmap of pancreatic cancer: Still evolving. Gut. 2017;66:2170–8.

72. Nagarajan A, Malvi P, Wajapeyee N. Heparan sulfate and Heparan Sulfate Proteoglycans in cancer initiation and progression. Front Endocrinol (Lausanne). 2018;9:483.

73. Theocharis AD, Skandalis SS, Tzanakakis GN, Karamanos NK. Proteoglycans in health and disease: Novel roles for proteoglycans in malignancy and their pharmacological targeting. FEBS J. 2010;277:3904–23.

74. Hingorani SR, Zheng L, Bullock AJ, Seery TE, Harris WP, Sigal DS, et al. HALO 202: Randomized phase II Study of PEGPH20 Plus Nab-Paclitaxel/Gemcitabine Versus Nab-Paclitaxel/Gemcitabine in Patients with Untreated, Metastatic Pancreatic Ductal Adenocarcinoma. J Clin Oncol. 2018;36:359-366

75. Provenzano PP, Cuevas C, Chang AE, Goel VK, Von Hoff DD, Hingorani SR. Enzymatic Targeting of the Stroma Ablates Physical Barriers to Treatment of Pancreatic Ductal Adenocarcinoma. Cancer Cell. 2012;21:418–29.

76. Mccleary-Wheeler AL, Mcwilliams R, Fernandez-Zapico ME. Aberrant signaling pathways in pancreatic cancer: A two compartment view. Mol Carcinog. 2012;51:25–39.

77. Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. Nat Med. 2014;20:1193–8.

78. Stotz M, Szkandera J, Seidel J, Stojakovic T, Samonigg H, Reitz D, et al. Evaluation of uric acid as a prognostic blood-based marker in a large cohort of pancreatic cancer patients. PLoS One. 2014;9:e104730.

79. Botwinick IC, Pursell L, Yu G, Cooper T, Mann JJ, Chabot JA. A biological basis for depression in pancreatic cancer. Hpb (Oxford). 2014;16:740-743.

80. Eguia V, Gonda TA, Saif MW. Early detection of pancreatic cancer. JOP. 2012; 13:131–4.

81. Carreras-Torres R, Johansson M, Gaborieau V, Haycock PC, Wade KH, Relton CL, et al. The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study. J Natl Cancer Inst. 2017, doi:10.1093/jnci/djx012.

82. Koyanagi YN, Matsuo K, Ito H, Tamakoshi A, Sugawara Y, Hidaka A, et al. Body-mass index and pancreatic cancer incidence: A pooled analysis of nine population-based cohort studies with more than 340,000 Japanese subjects. J Epidemiol. 2018; 28:245–52.

83. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer - Viewpoint of the IARC working group. N Engl J Med. 2016; 375:794–8.

84. Park J, Morley TS, Kim M, Clegg DJ, Scherer PE. Obesity and cancer - Mechanisms underlying tumour progression and recurrence. Nat Rev Endocrinol. 2014; 10:455–65.

85. López de Maturana E, Rodríguez JA, Alonso L, Lao O, Molina-Montes E, Martín-Antoniano I, et al. A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. PanGenEU GWAS summary statistics. GWAS Catalog. 2020. ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011857.

86. López de Maturana E, Rodríguez JA, Alonso L, Lao O, Molina-Montes E, Martín-Antoniano I, et al. A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. PanGenEU and EPICURO controls GWAS summary statistics. GWAS Catalog. 2020. ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011858.

Table 1. Novel pancreatic cancer genetic susceptibility hits prioritized by approaches 1D, 2D, and 3D, as well as by *in silico* functional analyses.

Genomic location	# Prioritized novel SNPs	Nearest gene	Selection approach	Relevance in pancreas physiology or carcinogenesis
1:150902203, 150974311	2	*SETDB1, FAM63A/MINDY1*	1D, 2D	Histone methyltransferase that cooperates in the development of PC; Lysine 48 deubiquitinatinase
1:200016460	1	*NR5A2*	1D, 2D	Transcription factor required for acinar differentiation and PC susceptibility gene
2:120278171	1	*SCTR*	1D	Secretin receptor expressed in ductal cells
6:108238917	1	*SEC63*	1D	ER protein involved in ER stress response
6:117150008	1	*GPRC6A*	2D	Disease-causal variant (CADD score=35)
6:117196211	1	*RFX6*	2D	Transcription factor involved in pancreatic development and adult endocrine cell function
8:128302062-128494384	27	*CASC8*	1D, 2D	LncRNA cancer associated susceptibility gene
9:21967751-21995300	6	*CDKN2A*	2D	Tumor suppressor mutated in >95% of PC; also involved in Familial PC
9:6772101, 6785243, 6831637	3	*KDM4C*	2D	Highly expressed in PC
9:94601093, 94603970	2	*ROR2*	2D	Wnt pathway (activated in PC)
11:60197299	1	*MS4A5*	2D	Disease-causal variant (CADD score=24.4)
13:46446427, 46405544, 46471859-46511859	2/CIR	*SIAH3*	2D, 3D	E3 ubiquitin ligase
16:67387817, 67397580	2	*LRRC36*	2D	Disease-causal variant (CADD score=24.4)
16:75263661, 75295639, 75201482-75241482	2/CIR	*BCAR1/CTRB2*	2D, 3D	CTRB2 is chymotrypsinogen 2, a major pancreatic protease; associated with chronic pancreatitis and PC
17:143542	1	*RPH3AL*	1D	Regulatory variant; Rabphilin 3a is involved in exocytosis in acinar cells
22:28476910-29165195	CIR	*MN1*	3D	No pancreas-related function has been discovered so far
22:28602352-28642352	CIR	*XBP1*	3D	Major regulator of the ER stress and unfolded protein responses in acinar cells

PC: pancreatic cancer; CIR: Chromatine interacting region

MAIN FIGURE AND LEGENDS

Figure 1. Study Flowchart: Overview of the complementary approaches adopted in this study to identify new pancreatic cancer susceptibility regions.

Figure 2. Three-dimensional genome organization in healthy and PANC-1 cells and association results corresponding to the genomic region around *XBP1* using the standard GWAS and 2D approaches. A) Coverage-normalized Hi-C maps of healthy samples and PANC-1 cells at 40Kb resolution. Green ellipses highlight the interaction between the region harboring four Local Moran's Index (LMI)-selected SNPs and the *XBP1* promoter. B) Tracks of the ChromHMM Chromatin for 8 states in healthy pancreas, PANC-1 cells, and a Pancreatic Intraepithelial Neoplasia 1B. Promoters are colored in light purple, strong enhancers in dark green and weak enhancers in yellow. Note that the strong enhancer in the target region is lost in the PANC-1 and PanIN-1B samples, compared to the healthy samples. C) UCSC tracks of H3K27ac, an enhancer-associated mark, and arcs linking significant interactions called by Homer. Interactions in healthy pancreas samples are in green and those in PANC-1 and in the PanIN-1B sample are in purple. Red arc represents the interaction between LMI-prioritized SNPs and the *XBP1* promoter (highlighted region in Hi-C map in A). D) Scatterplots of SNPs in region chr22:28,400,000-29,600,000 (hg19) and their –log10 (p-value), LMI and odds ratio. Bait and target chromatin interaction regions are highlighted in yellow and blue, respectively.

Figure 3. Zoom plot of the 8q24.21 CASC8 (cancer Susceptibility 8) region and linkage disequilibrium pattern of the PanGenEU GWAS prioritized variants. Red and green points indicate OR<1 and OR>1, respectively.

Figure 4. Scatterplots of the –log10 *p*-values, local Moran's index (LMI) values and odds ratios (OR) for three genomic regions prioritized based on their LMI value. Highlighted regions show the hits identified in the 2D, but not in the 1D approach.

Figure 5. Network of traits in the GWAS catalog enriched with the genes prioritized in the 1D approach of PanGenEU GWAS. Twelve densely connected subgraphs identified via random walks are displayed in different colors.

Figure 1

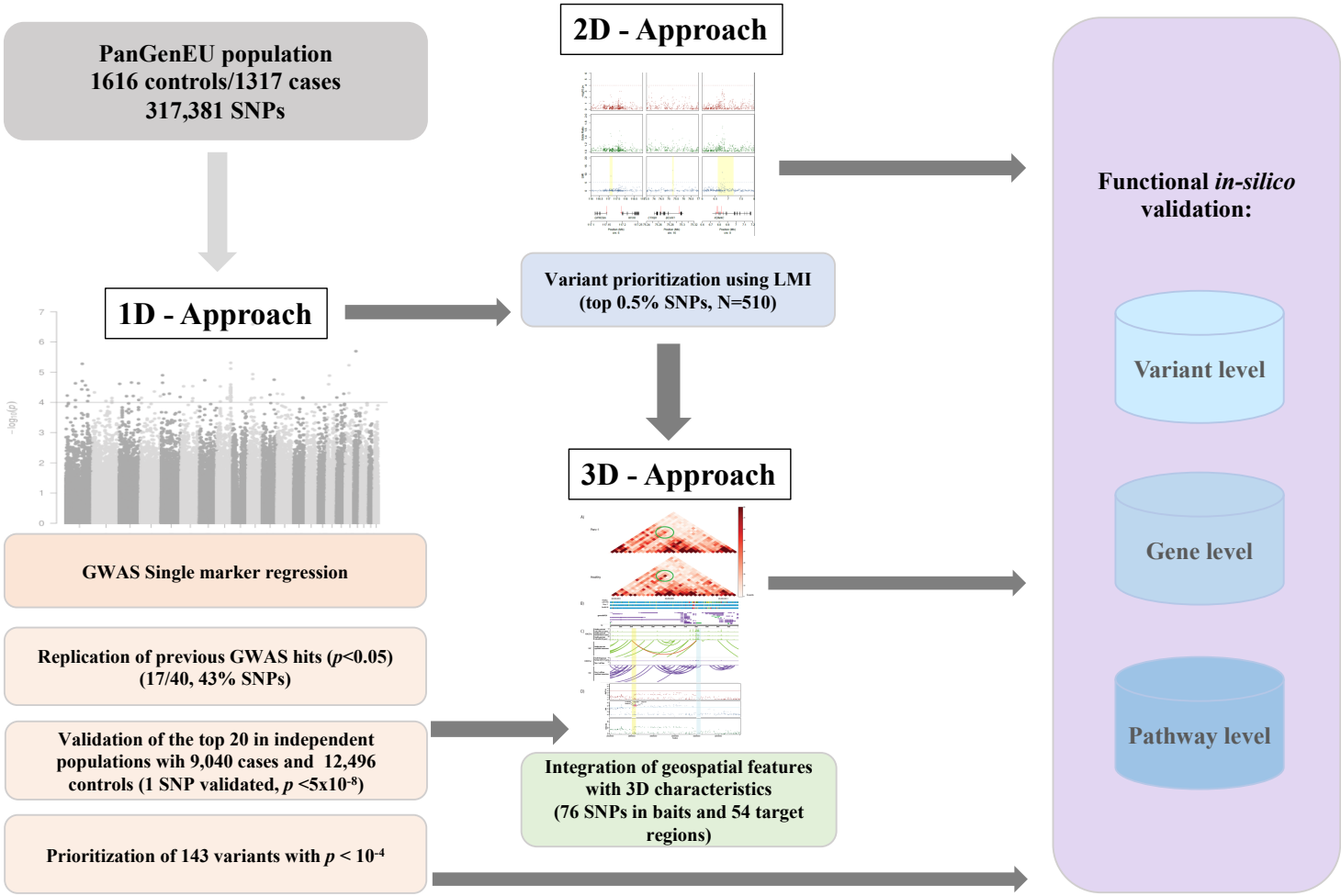

PanGenEU population
1616 controls/1317 cases
317,381 SNPs
1D - Approach
GWAS Single marker regression
Replication of previous GWAS hits (p<0.05)
(17/40, 43% SNPs)
Validation of the top 20 in independent populations wih 9,040 cases and 12,496 controls (1 SNP validated, p <5x10-8)
Prioritization of 143 variants with p < 10-4
2D - Approach
Variant prioritization using LMI
(top 0.5% SNPs, N=510)
3D - Approach
Integration of geospatial features with 3D characteristics
(76 SNPs in baits and 54 target regions)
Functional in-silico validation:
Variant level
Gene level
Pathway level

Figure 2

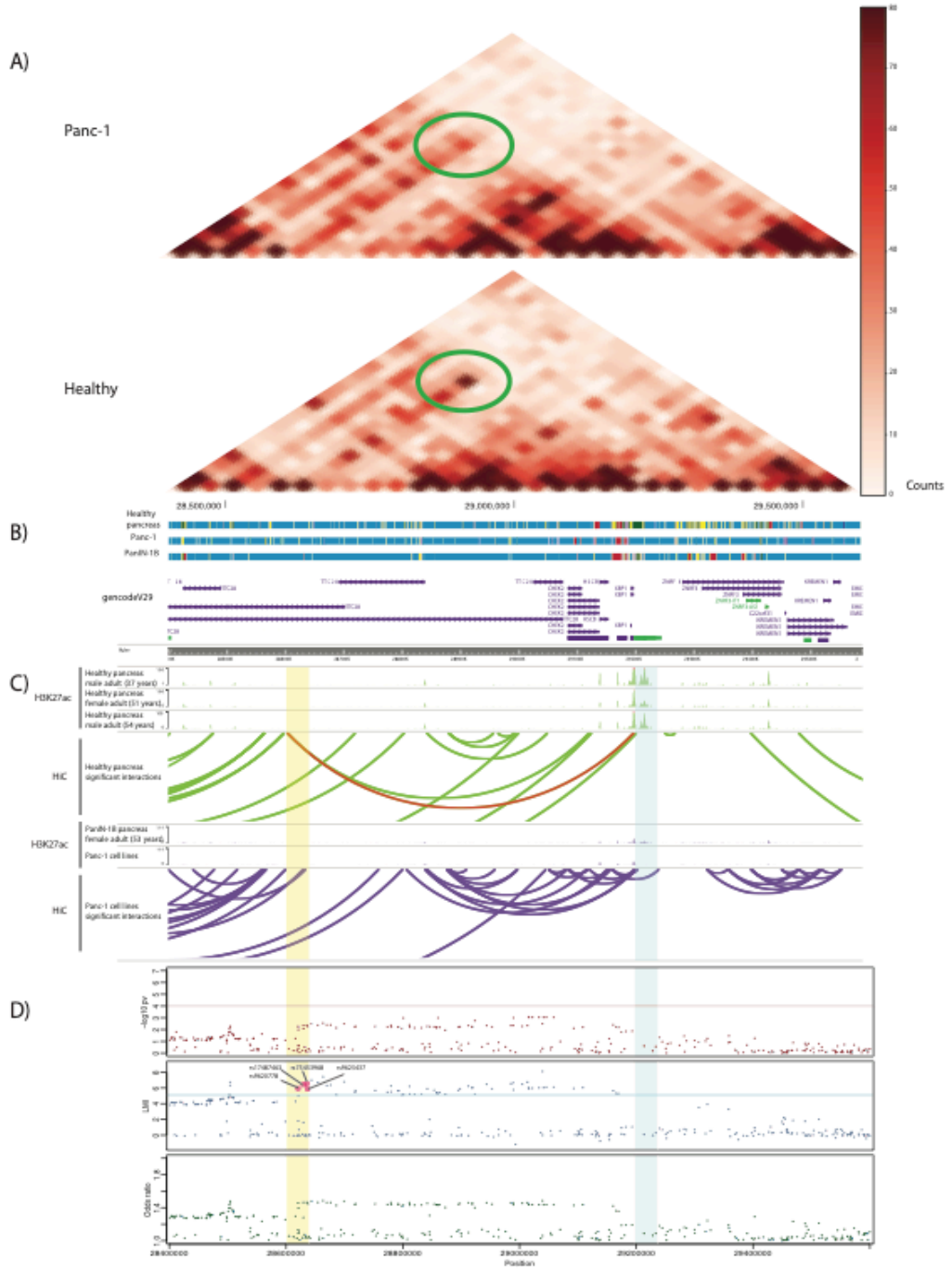


Figure 3

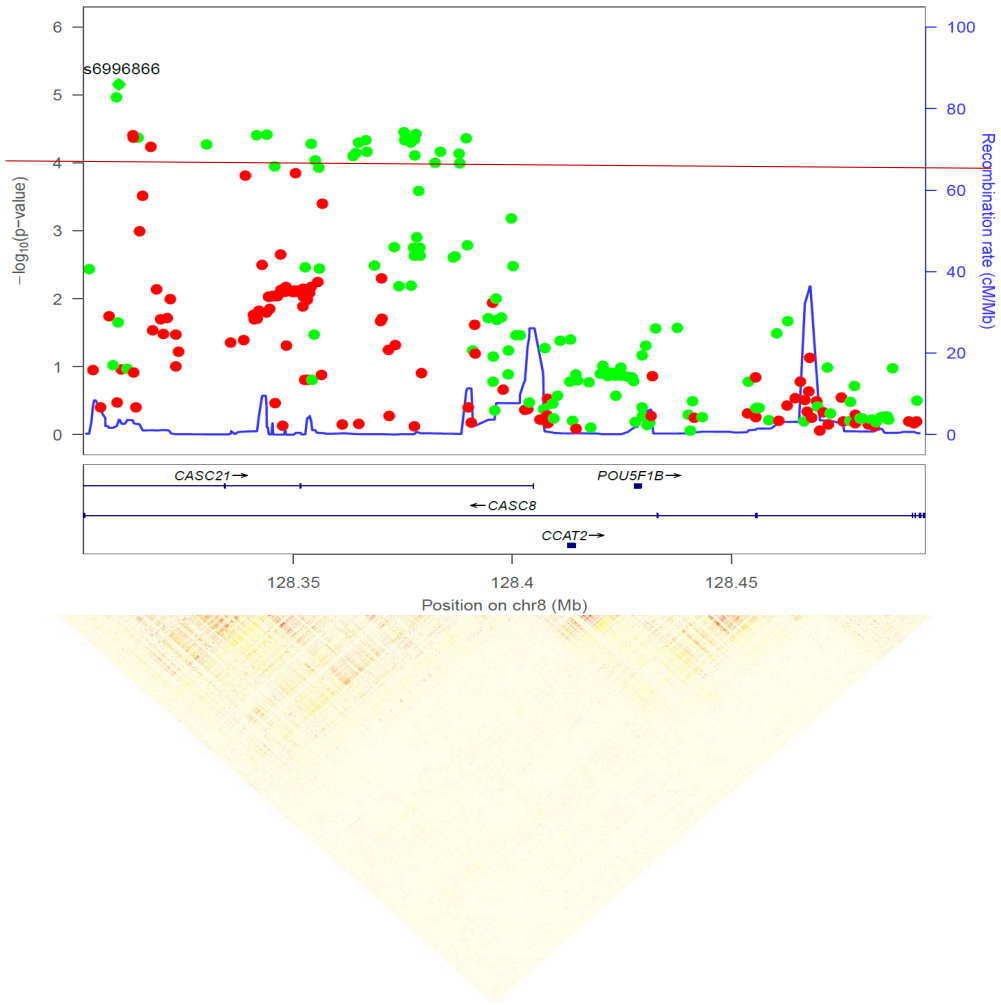

s6996866
6
5
4
3
2
1
0
100
80
60
40
20
0
Recombination rate (cM/Mb)
−log10(p−value)
CASC21
POU5F1B
CASC8
CCAT2
128.35
128.4
128.45
Position on chr8 (Mb)

Figure 4

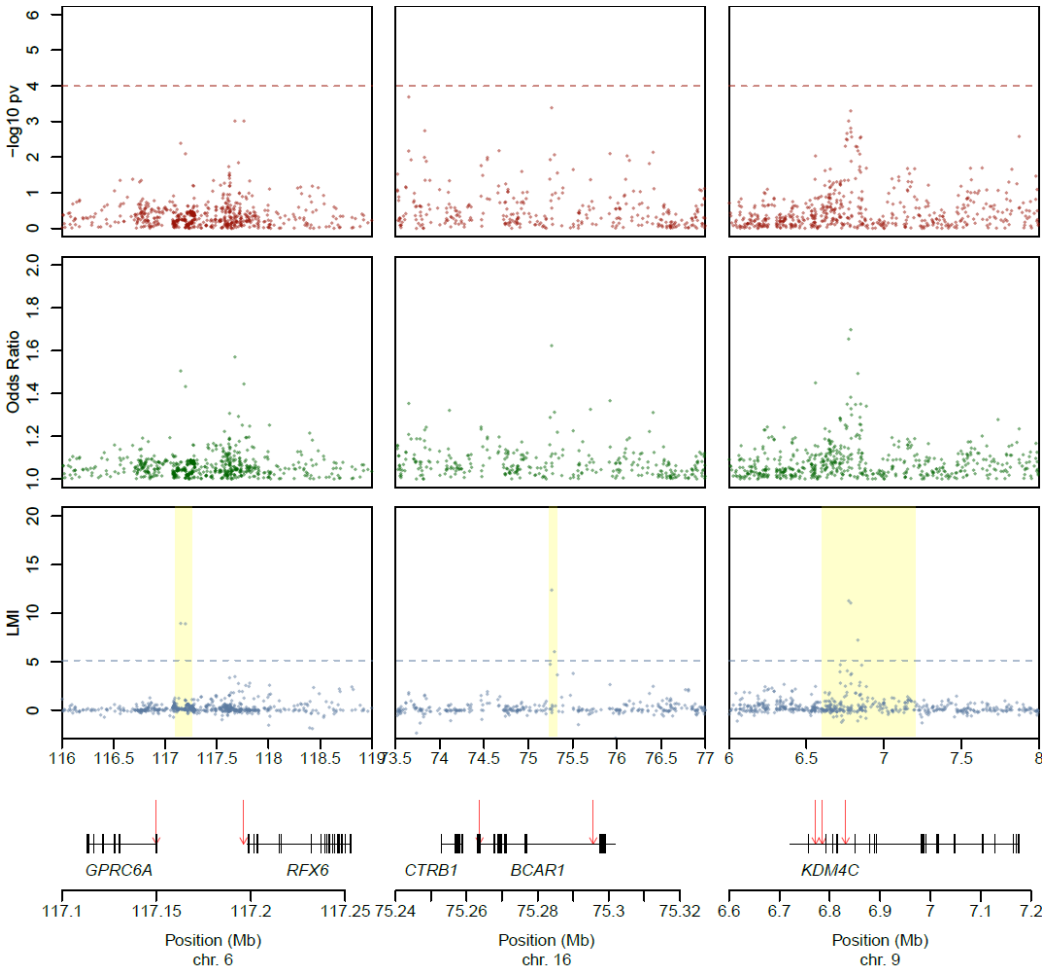
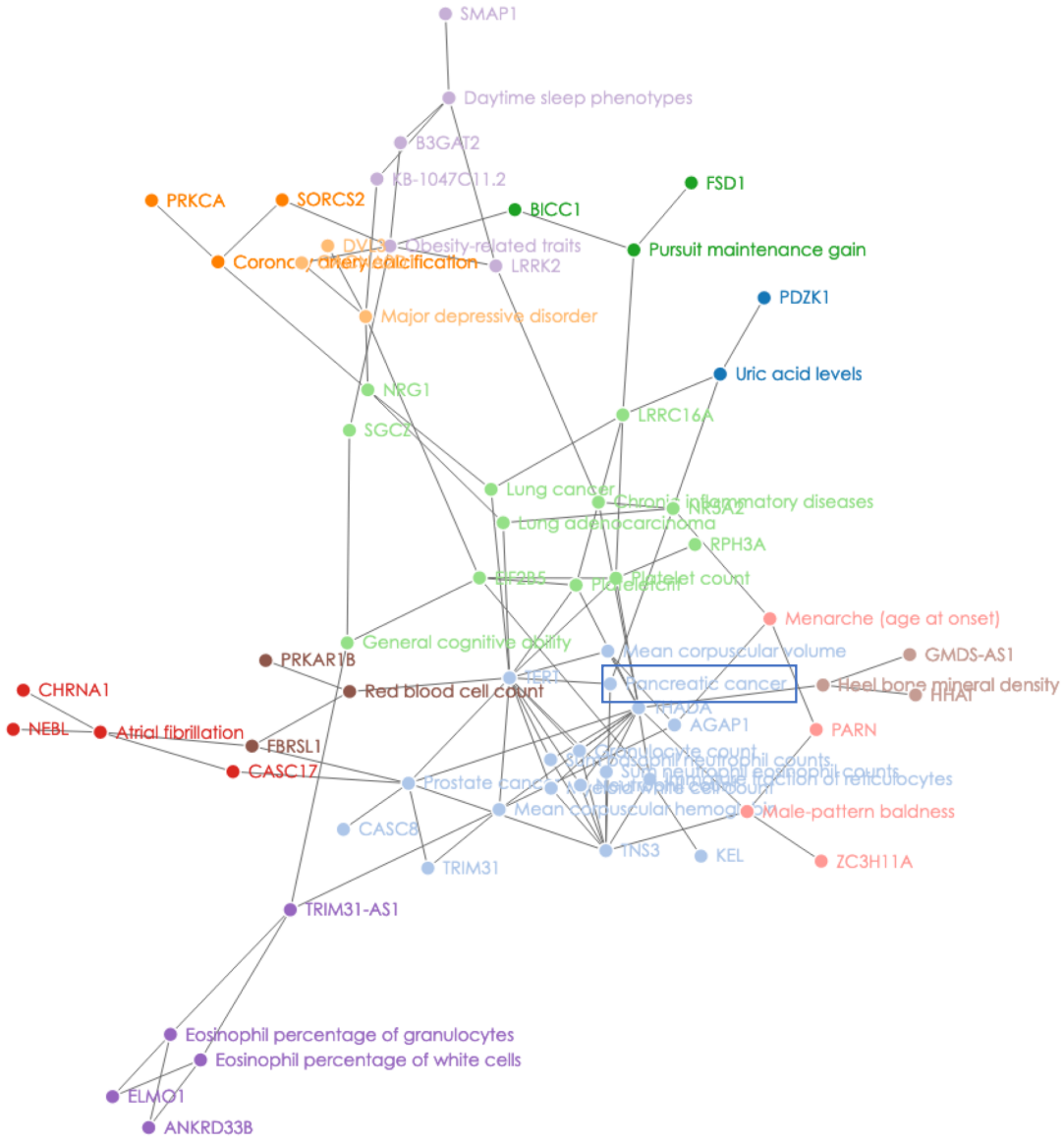

−log10 pv
Odds Ratio
LMI
GPRC6A
RFX6
CTRB1
BCAR1
KDM4C
Position (Mb)
chr. 6
Position (Mb)
chr. 16
Position (Mb)
chr. 9

Figure 5


SMAP1
Daytime sleep phenotypes
B3GAT2
KB-1047C11.2
PRKCA
SORCS2
BICC1
FSD1
DVL3
Obesity-related traits
Pursuit maintenance gain
Coronary artery calcification
LRRK2
PDZK1
Major depressive disorder
Uric acid levels
NRG1
LRRC16A
SGCZ
Lung cancer
Chronic inflammatory diseases
INKA2
Lung adenocarcinoma
RPH3A
EIF2B5
Platelet count
Menarche (age at onset)
General cognitive ability
Mean corpuscular volume
GMDS-AS1
PRKAR1B
Pancreatic cancer
Heel bone mineral density
CHRNA1
Red blood cell count
HHAT
RNADA
NEBL
Atrial fibrillation
AGAP1
PARN
FBRSL1
CASC17
Prostate cancer
Male-pattern baldness
CASC8
TRIM31
TNS3
KEL
ZC3H11A
TRIM31-AS1
Eosinophil percentage of granulocytes
Eosinophil percentage of white cells
ELMO1
ANKRD33B

Additional File 1

López de Maturana E, et al. A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer

Annex 1. PanGenEU Centres and Investigators

Spanish National Cancer Research Centre (CNIO), Madrid, Spain: Núria Malats[1], Francisco X Real[1], Evangelina López de Maturana, Paulina Gómez-Rubio, Esther Molina-Montes, Lola Alonso, Mirari Márquez, Roger Milne, Ana Alfaro, Tania Lobato, Lidia Estudillo.

Verona University, Italy: Rita Lawlor[1], Aldo Scarpa, Stefania Beghelli.

National Cancer Registry Ireland, Cork, Ireland: Linda Sharp[1], Damian O'Driscoll.

Hospital Madrid-Norte-Sanchinarro, Madrid, Spain: Manuel Hidalgo[1], Jesús Rodríguez Pascual.

Hospital Ramon y Cajal, Madrid, Spain: Alfredo Carrato[1], Carmen Guillén-Ponce, Mercedes Rodríguez-Garrote, Federico Longo-Muñoz, Reyes Ferreiro, Vanessa Pachón, M Ángeles Vaz.

Hospital del Mar, Barcelona, Spain: Lucas Ilzarbe[1], Cristina Álvarez-Urturi, Xavier Bessa, Felipe Bory, Lucía Márquez, Ignasi Poves✝, Fernando Burdío, Luis Grande, Mar Iglesias, Javier Gimeno.

Hospital Vall d´Hebron, Barcelona, Spain: Xavier Molero[1], Luisa Guarner✝, Joaquin Balcells.

Technical University of Munich, Germany: Christoph Michalski[1], Jörg Kleeff, Bo Kong.

Karolinska Institute, Stockholm, Sweden: Matthias Löhr[1], Jiaqui Huang, Weimin Ye, Jingru Yu.

Hospital 12 de Octubre, Madrid, Spain: José Perea[1], Pablo Peláez.

Hospital de la Santa Creu i Sant Pau, Barcelona, Spain: Antoni Farré[1], Josefina Mora, Marta Martín, Vicenç Artigas, Carlos Guarner, Francesc J Sancho, Mar Concepción, Teresa Ramón y Cajal.

The Royal Liverpool University Hospital, UK: William Greenhalf[1], Eithne Costello.

Queen's University Belfast, UK: Michael O'Rorke[1], Liam Murray✝, Marie Cantwell.

Laboratorio de Genética Molecular, Hospital General Universitario de Elche, Spain: Víctor M Barberá[1], Javier Gallego.

Instituto Universitario de Oncología del Principado de Asturias, Oviedo, Spain: Adonina Tardón[1], Luis Barneo.

Hospital Clínico Universitario de Santiago de Compostela, Spain: Enrique Domínguez Muñoz[1], Antonio Lozano, Maria Luaces.

Hospital Clínico Universitario de Salamanca, Spain: Luís Muñoz-Bellvís[1], J.M.

Sayagués Manzano, M.L. Gutíerrrez Troncoso, A. Orfao de Matos.

University of Marburg, Department of Gastroenterology, Phillips University of Marburg, Germany: Thomas Gress[1], Malte Buchholz, Albrecht Neesse.

Queen Mary University of London, UK: Tatjana Crnogorac-Jurcevic[1], Hemant M Kocher, Satyajit Bhattacharya, Ajit T Abraham, Darren Ennis, Thomas Dowe, Tomasz Radon

Scientific advisors of the PanGenEU Study: Debra T Silverman (NCI, USA) and Douglas Easton (U. of Cambridge, UK)

[1] Principal Investigator in each centre

Annex 2. Spanish Bladder Cancer/EPICURO Centres and Investigators

Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, Barcelona –Coordinating Center1: M. Kogevinas, N. Malats, F.X. Real, M. Sala, G. Castaño, M. Torà, D. Puente, C. Villanueva, C. Murta-Nascimento, J. Fortuny, E. López, S. Hernández, R. Jaramillo, G. Vellalta, L. Palencia, F. Fermández, A. Amorós, A. Alfaro, G. Carretero.

National Cancer Institute, NIH, USA -–Coordinating Center2: Debra Silverman, Mustafa Dosemeci✝, Nathaniel Rothman, Montserrat García-Closas, Stephen Chanock.

Hospital del Mar, Universitat Autònoma de Barcelona, Barcelona: J. Lloreta, S. Serrano, L. Ferrer, A. Gelabert, J. Carles, O. Bielsa, K. Villadiego.

Hospital Germans Trias i Pujol, Badalona, Barcelona: L. Cecchini, J.M. Saladié, L. Ibarz.

Hospital de Sant Boi, Sant Boi de Llobregat, Barcelona: M. Céspedes.

Consorci Hospitalari Parc Taulí, Sabadell: C. Serra, D. García, J. Pujadas, R. Hernando, A. Cabezuelo, C. Abad, A. Perera, J. Prat.

Centre Hospitalari i Cardiològic, Manresa, Barcelona: M. Domènech, J. Badal, J. Malet.

Hospital Universitario de Canarias, La Laguna, Tenerife: R. García-Closas, J. Rodríguez de Vera, A.I. Martín.

Hospital Universitario Nuestra Señora de la Candelaria, Tenerife: J. Taño✝, F. Cáceres.

Hospital General Universitario de Elche, Universidad Miguel Hernández, Elche, Alicante: A. Carrato, F. García-López, M. Ull, A. Teruel, E. Andrada, A. Bustos, A. Castillejo, J.L. Soto.

Universidad de Oviedo, Oviedo, Asturias: A. Tardón.

Hospital San Agustín, Avilés, Asturias: J.L. Guate, J.M. Lanzas, J. Velasco.

Hospital Central Covadonga, Oviedo, Asturias: J.M. Fernández, J.J. Rodríguez, A. Herrero.

Hospital Central General, Oviedo, Asturias: R. Abascal, C. Manzano, T. Miralles.

Hospital de Cabueñes, Gijón, Asturias: M. Rivas, M. Arguelles.

Hospital de Jove, Gijón, Asturias: M. Díaz, J. Sánchez, O. González.

Hospital de Cruz Roja, Gijón, Asturias: A. Mateos, V. Frade.

Hospital Alvarez-Buylla, Mieres, Asturias: P. Muntañola, C. Pravia.

Hospital Jarrio, Coaña, Asturias: A.M. Huescar, F. Huergo.

Hospital Carmen y Severo Ochoa, Cangas, Asturias: J. Mosquera.

Supplementary Methods and Results

Benchmarking of LMI

To assess the usefulness of the LMI score for SNP prioritization, we ran two tests using SNPs known to be associated with PC in European populations [GWAS Catalog, n=40]. Before performing this benchmarking tests, we corrected the 40 signals by LD using a custom made "greedy" algorithm. First, we calculated all pairwise LD values (r^2) for all the SNPs on the same chromosome. Then, we reviewed the list of SNPs ordered by ascending position chromosome-wise and considered as a cluster all the SNPs that had $r^2 > 0.2$ with the SNP under consideration. We considered this set of SNPs as a unique genomic signal, filtered out the SNPs assigned to the cluster from the ranked list, and then proceeded to the next SNP. This resulted in a total of 30 independent clusters of >1 SNPs. When more than one SNP was included within the same cluster, the SNP with the highest LMI was selected. Then, the two benchmarking tests consisted of 1) evaluating whether the GWAS Catalog PC-associated SNPs known to be associated with PC in European populations (GWAS Catalog, n=40) had a LMI value higher than expected, and 2) assessing how many of the previously reported loci were also identified according to the LMI out of the 30 independent signals of >1 SNPs.

Pathway-enrichment analysis

To ensure that key pathways functionally connected to the selected ones in the pathway enrichment analysis are not lost, we used the *Pathway-connector* webtool (Minadakis et al., 2019) (**Figure S9**), which also found six complementary KEGG pathways: 'Tyrosine metabolism', 'Metabolic pathways', 'Glycolysis/Gluconeogenesis', 'Glycerolipid metabolism', 'PI3K-Akt signaling pathway', 'mTOR signaling pathway'.

Functional *in-silico* analyses

We performed an in-depth, systematic, *in silico* functional analysis at SNP, gene, and pathway levels (**Figure S1**) for variants with a GWAS *p*-value$<1\times10^{-4}$ (N=143). The functional impact of variants and their annotation was performed using DoriTool, an integrative pipeline developed at CNIO, built to combine different bioinformatics algorithms and public databases (Martín-Antoniano et al., 2017). We also investigated if the prioritized variants were identified as PC-associated lncRNAs in a catalogue computed using a systems and experimental biology approaches (Arnes et al., 2019).

eQTL analysis was done using three independent pancreatic datasets: (1) GTEx version 7 dataset for 220 normal pancreatic tissue samples was obtained from the GTEx Portal on

06/11/2018 (Ardlie et al., 2015); 2) Laboratory of Translational Genomics (LTG) dataset with 95 histologically normal pancreatic samples (Zhang et al., 2018); and 3) The Cancer Genome Atlas (TCGA) dataset with 178 pancreatic tumors (PAAD samples), using the PancanQTL webtool (Gong et al., 2018). RefSeq genes located within +/−500 kb of the marker SNP for each GWAS significant locus were assessed for *cis*-eQTL effects. The eQTL analysis was performed separately for each dataset.

The **mQTL** analysis was based on the methylation and genotyping data from leukocytes DNA generated in 265 controls from the PanGenEU study. Leukocyte DNA methylation data were obtained using the Illumina Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA, USA, 850k) according to the manufacturer's protocol (Infinium HD Methylation Assay). Preprocessing of methylation data included background correction, normalization procedure using the *preprocessQuantile* function of *minfi* (Aryee et al., 2014), and exclusion of 'failed' probes and cross-reactive probes (Pidsley et al., 2016). A linear model was assumed to test the association between each prioritized SNP and the M-value of each individual CpG site 5mC levels. The models were adjusted for age and sex. All analyses were adjusted for multiple testing correction using Benjamini-Hochberg's method (Benjamini & Hochberg, 1995).

For the **pQTL** analysis, we interrogated the genetic atlas of the human plasma proteome including 1,927 genotype-protein associations, recently published (Sun et al., 2018).

Histone marks annotation was performed using the DNase hypersensitive sites, histone modifications, transcription start sites (TSS), active promoters, and transcription factor-binding sites in human cell lines and tissues from assays conducted in pancreas tissue catalogued in the ENCODE project (Sloan et al., 2016).

SNPs were mapped to significant **3D chromatin interaction regions** revealed by HOMER (Heinz et al., 2010) using the Hi-C map of pancreas tissue (Schmitt et al., 2016). Then, we annotated those SNPs in the CI region 1 overlapping with an enhancer region, as well as those interacting with a CI region overlapping with a promoter region. Both enhancer and promoter regions were obtained from Roadmap Epigenomics Projects for pancreas (E098) (https://egg2.wustl.edu/roadmap/web_portal/DNase_reg.html#delieation) and were predicted using DNase peaks and core 15-state chromatin state model. Enhancers were linked to genes using the predictions of enhancer-gene links for the Roadmap Epigenomics reference epigenome for pancreas (E098) for states 6, 7 and 12 (http://www.biolchem.ucla.edu/labs/ernst/roadmaplinking/).

We annotated prioritized SNPs in the **differentially opening regions (DORs)** in human pancreatic endocrine and exocrine lineages reported by Arda et al (Arda et al., 2018).

We interrogated one of the largest publicly available collections of genes and variants associated to human diseases using DisGeNET (Piñero et al., 2017), to investigate whether the prioritized variants are also associated with **PC comorbidities or other types of cancer**.

We performed **enrichment analysis at the gene level** to interpret the potential relevance of our prioritized variants in a broader context of genes and molecular pathways. To this end, we used FUMA GWAS platform (Watanabe et al., 2017),. Moreover, we used DisGeNet R package to analyse the properties of disease genes.

cBioPortal for Cancer Genomics was used to investigate **molecular profiling** at gene level in PC samples and to identify the molecular profiles including mutations and copy number alterations (amplifications/deletions) that our SNP-associated gene set present in PC samples from the four available studies.

Finally, the differential gene **expression profiling** in cancer and normal tissue was evaluated by comparing tumor to normal tissue to find tumor-specific genes using GEPIA, a web-based tool using TCGA, and GTEx data (Ardlie et al., 2015).

Functional *in-silico* results

We present here the *in-silico* functional evidences for the 1D-approach prioritized variants (**Table S6**):

- Two variants at **1q21.3** (rs17661062 and rs59942146) in high LD (r^2=0.92) and intronic to ***SETDB1*** and ***FAM63A*** were associated with an increased methylation of the cg17724175 in ***MCL1***. *SETDB1* appears altered in 4% of the PC tumors analyzed in cBioPortal.
- rs12756006 at **1p22.3** is a regulatory variant in a promoter flanking region which overlaps with a DOR in ductal cells. Moreover, it overlaps with H3K4me1 and H3K9me3 in pancreas.
- *NR5A2*-rs4465241 and *NR5A2*-rs3790840 (**1q32.1**) overlap with one and 10 histone marks, respectively (H3K27ac in pancreas; and DNase, H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3 and H3K9me3 in pancreas and H3K27ac, H3K27me3 and H3K36me3 in endocrine pancreas). As discussed before, variants tagging ***NR5A2*** have previously been associated with PC risk in GWAS. This orphan nuclear receptor participates in a wide variety of processes such as cholesterol and glucose metabolism in the liver, resolution of endoplasmic reticulum stress, intestinal glucocorticoid production, pancreatic development and acinar differentiation, and inflammatory response. In addition, *NR5A2* gene has been previously associated with uric acid levels, also associated with PC and lung adenocarcinoma risk (Cobos et al

2018). Moreover, the expression of *NR5A2* in tumor is significantly decreased compared with its counterpart in normal tissue (Log2=-1.84, adjusted-*p*-value= $1.85x10^{-43}$).

- rs6697813 (**1p36.21**) is upstream to ***DDI2*** (*DNA damage inducible 1 homolog 2*) and was significantly associated with differential methylation in leukocytes of four CpG sites annotated in ***CHCHD2P6-RP4-680D5.2***, ***CASP9***, and ***AGMAT-DNAJC16***. The *G* allele is associated with an increased expression of ***C1orf144*** in normal pancreas. Moreover, it overlaps with 7 chromatin marks: two in endocrine pancreatic tissue and five in pancreas tissue. Another *DDI2* variant in linkage equilibrium with our hit has been previously associated with alcoholic chronic pancreatitis (Rosendahl et al., 2018).
- *DHRS3*-rs12136952 (**1p36.22**) is an intronic variant of ***DHRS3*** located in an enhancer region of the gene, overlapping a H3K4me1 chromatin mark in pancreas tissue, enriched at active enhancers. This gene is a highly conserved member of the short-chain dehydrogenases/reductases family, which encodes the DHRS3 protein, an endoplasmic reticulum protein (Deisenroth et al., 2011). The variant also overlaps a CTCF binding site and is in a chromatin region contacting the promoter of the *Vacuolar Protein Sorting 13 Homolog D* (***VPS13D***), involved in vesicle transport. This variant is a potential regulator of the expression of *DHRS3*. *DHRS3* appears to be overexpressed in PC tumors versus normal tissue.
- The rs11118832 variant at **1q41** is located in an intron of ***DUSP10***. It overlaps with 5 and 6 chromatin marks, respectively. Variants in *DUSP10* were previously reported as associated with colorectal cancer. *DUSP10* is mutated in 1.9% of pancreatic adenocarcinoma/cancer samples in cBioPortal and is differentially expressed in tumor vs normal tissue (Log2= 1.135; *p*-value=7.23E-37).
- *SCTR*-rs4383344 (**2q14.2**) is associated with decreased methylation of cg24309134, located in an openSea region. ***SCTR*** codes for the secretin receptor, crucially involved in the function of healthy pancreatic ductal epithelial cells where it stimulates pancreatic bicarbonate, electrolyte, and fluid secretion; its silencing may contribute to tumor growth and progression of PC (Ding et al., 2002). Secretin receptors are overexpressed in non-neoplastic pancreas ducts and its isoforms may be correlated with decreased secretin binding in pancreatic ductal tumors.
- *KIAA1257*-rs1683813 (**3q21.3**) is associated with a decreased methylation of cg27310733 and with the overexpression of ***LOC653712*** and ***AC112484.3*** in both pancreatic tissue and whole blood (GTEx). Moreover, it overlaps with H3K27me3 in pancreatic tissue. *KIAA1257*-rs9810890 *(r^2*=0.11 with our hit) is associated with dental

caries. ***KIAA1257*** is mutated in 0.8% of pancreatic adenocarcinoma/cancer samples in cBioPortal.

- rs2584051 (**3p26.2**) is a non-coding transcript exon variant at the *Leucine Rich Repeat Neuronal 1* (***LRRN1***) and intronic of sulfatase modifying factor 1 (***SUMF1***). This regulatory variant overlaps with a DOR in ductal cells and with two chromatin marks associated with active transcription (H3K4me1 and H3K4me3).
- *SEC63*-rs12206846 (**6q21**) is associated with an over-methylation of cg00234027. ***SEC63*** encodes a protein involved in protein translocation in the endoplasmic reticulum. The SNP overlaps with H3K36me3 and H3K9me3 in both endocrine and pancreas tissue. An independent variant in this gene (rs11153123) also overlaps with five chromatin marks in pancreas: H3K27ac, H3K36me3, H3K4me1, H3K4me3 and H3K9me3.
- *GMDS-AS1*-rs761098 (**6p25.2**) is a regulatory region variant in *GDP-Manose 4, 6-dehydratase Antisense* RNA 1 (Head To Head), a non-coding RNA gene that has been associated with pancreatic ductal adenocarcinoma (Arnes et al., 2019), suggesting that genetic variants therein may contribute to transcriptional regulation in pancreatic cancer. ***GMDS-AS1*** catalyzes the conversion of GDP-mannose to GDP-4-keto-6-deoxymannose, the first step in the synthesis of GDP-fucose from GDP-mannose (using NADP+ as a cofactor). GDP-fucose is the fucose donor used to synthesize all mammalian fucosylated glycans, including ABO blood group antigens, present on the surface of all cell membranes. *GMDS-AS1* appears with a deep deletion in 0.8% of the pancreatic adenocarcinoma/cancer samples in cBioPortal. *GMDS* gene is expressed in PANC-1 cells and in well-differentiated tumors. *GMDS* is enriched in the epithelium in primary tumors (p-adj<0.05). A loss-of-function variant in this gene associated with tumor progression. *GMDS* is mutated in two PC samples in cBioPortal and is differentially expressed in PC (TCGA) vs normal tissue (GTEx) (Log2FC=1.625, p-value=$1.65x10^{-45}$).
- *TNS3*-rs2271311 (**7p12.3**) is associated with an increased methylation of *TNS3*-cg06114556 and overlaps with 5 chromatin marks in pancreas (DNase; H3K27ac; H3K36me3; H3K4me1; and H3K4me3) and 2 in endocrine pancreatic tissue (H3K9ac H3K27ac). The SNP is in linkage equilibrium with rs78417682, a pancreatic cancer susceptibility variant reported in (Klein et al., 2018). ***TNS3*** is altered in 1% of the pancreatic adenocarcinoma cases in cBioPortal.
- Two variants at **7q22.1** (rs62484781, and rs62482372; $r^2_{rs62484781\text{-}rs62482372}$=0.7) are located upstream to ***DPY19L2P2***. rs6955512 is associated with increased methylation

of SMURF1-cg27297376 and it overlaps with H3K27ac and H3K4me1 marks in both endocrine and pancreatic tissue. Variants in *DPY19L2P2* were previously associated with breast cancer (Michailidou et al., 2017). *DPY19L2P2*-rs62484781 overlaps with H3K27ac and H3K4me1 histone marks in both endocrine and pancreatic tissue. Approximately 2% of PC in cBioPortal showed alterations in ***SMURF1***.

- *FBRSL1*-rs6560884 (**12q24.33**) is associated with decreased expression of ***FBRSL1*** in normal pancreatic tissue (GTEx and LTG datasets) and it overlaps with three chromatin marks in pancreas tissue and with one in endocrine pancreatic tissue. This variant is in moderate LD with rs10870474 (r^2=0.43), also in *FBRSL1*, which was associated with increased methylation of 5 cpgs (cg25130710; *FBRSL1*-cg03621470, promoter associated; cg15310701, cg15785681 and cg24135151), with an increased expression of the same gene in whole blood (GTEx), and overlaps with H3K27ac, H3K36me3 and H3K9me3 marks in pancreatic tissue. *FBRSL1* is altered in 1.3% of PC tumors.
- *PRKCA*-rs11654719 at **17q24.2** is associated with an increased methylation of cg08055746 (OpenSea) and it overlaps with H3K4me1 in both pancreatic and endocrine tissues, and with H3K27ac, H3K4me1 and H3K9me3 marks in pancreatic tissue. ***PRKCA*** participates in the SCTR pathway. Moreover, it is altered in 1.3% of the pancreatic tumors.
- rs8111858 (**19p13.3**) is an intronic variant of ***STAP2*** *(Signal-transducing adaptor family member-2*) and *FSD1 (Fibronectin type III and SPRY domain containing 1*). *STAP2* encodes an adaptor protein that regulates various intracellular signaling pathways and promotes tumorigenesis in melanoma and breast cancer (Kitai et al., 2017). This variant seems to be functional because the T allele is associated with lower methylation in leukocytes and with an increased expression of *STAP2* in normal pancreatic tissue. Moreover, it overlaps with three histone chromatin marks which are associated with activation of gene transcription (H3K36me3, H3K4me1 and H3K4me3). *STAP2* is overexpressed in pancreatic tumors in comparison with normal pancreatic tissue.
- rs62209634 (**20q11.22**) is an intergenic variant interacting with a chromatin region harboring the promoters of three genes: ***PXMP4***, ***ZNF341****,* and ***RP4-553F4.2***. *PXMP4* is overexpressed in tumors, compared to normal tissue. The SNP is associated with differential expression of ***BPIFB2****,* ***CHMP4B*** and ***DYNLRB1*** in normal pancreatic tissue.

No overlapping was found between our 143 variants and pQTLs obtained in the plasma proteome.

No significant enrichment of our gene set was observed for gene expression in normal pancreatic tissue (GTEx v7).

Shared variants between mental diseases and PC

PC is one of the tumor entities with one of the highest incidences of depression preceding cancer diagnosis (Eguia et al., 2012). To explore the relationship between mental disorders and PC risk, we analyzed the association between variants in genes associated with PC risk in our study and both neuroticism (***RBFOX1*** and ***PRKCA***) and major depression disorder (***CACNA2D1***, ***NRG1***, ***NRG1-IT2*** and ***KB-1047C11.2***) in GWAS catalog and the risk of mental disorders in PC cases from PanGenEU study. None of the associations remained significant after multiple testing correction, the minimum *p*-value being for ***RBFOX1***-rs10852686 ($4.95x10^{-03}$).

Drug repurposing on newly identified genes

We analyzed whether some of the genes we retrieved either through GWAS-LMI (n=338) or 3D interactions (n=37) were drug-targets. We searched for all these genes detected through LMI and 3D in PharmaGKB database (*downloaded full data on May 28th, 2019* from pharmgkb.org) and searched for their occurrences in that database. We did not find direct evidence of these genes being target for PC current treatments. A total of 23/338 (6.8%) genes in the selected GWAS-LMI distribution were annotated in the list of clinically actionable gene-drug associations for other cancer types or conditions associated with PC. Four genes were already pharmacological targets for several types of cancer (***BCL2***; ovarian neoplasms, ***CDKN2A,*** and ***FOLH1***; lymphoblastic leukemia, ***ZNF423***; breast neoplasms) and two genes (***AQP2*** and ***NBEA***) were clinically actionable to treat type II diabetes mellitus. Intriguingly, ~30% of the actionable genes had associations with drugs to treat mental conditions (*e.g.*, anxiety, depression or psychotic disorders), which have been associated with PC (Eguia et al., 2012). Furthermore, ***ALK, BDNF, DDHD1*** and ***XBP1***, which is 10.8% of the 37 target genes annotated through Hi-C interactions, had at least an entry in the database for any of the categories. Remarkably, ***ALK*** is already a target for many cancer treatments (neuroblastoma, non-small cell lung cancer and anaplastic lymphoma), and is currently undergoing clinical trial as a target for Ceritinib for treating pancreatic cancer metastasis: clinicaltrials.gov/ct2/show/NCT02227940. ***BDNF*** is the target for many psychiatric disorders (depression, schizophrenia, substance abuse disorder, among others). ***DDHD1*** has been associated with warfarin dose-response, and, finally, SNP variants in ***XBP1*** affect response to platinum compounds in lung cancer treatment.

REFERENCES

Arda, H. E., Tsai, J., Rosli, Y. R., Giresi, P., Bottino, R., Greenleaf, W. J., Chang, H. Y., & Kim, S. K. (2018). A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas. *Cell Systems*, *7*(3), 310-322.e4. https://doi.org/10.1016/j.cels.2018.07.007

Ardlie, K. G., DeLuca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., Trowbridge, C. A., Maller, J. B., Tukiainen, T., Lek, M., Ward, L. D., Kheradpour, P., Iriarte, B., Meng, Y., Palmer, C. D., Esko, T., Winckler, W., Hirschhorn, J. N., Kellis, M., … Lockhart. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, *348*(6235), 648–660. https://doi.org/10.1126/science.1262110

Arnes, L., Liu, Z., Wang, J., Maurer, H. C., Sagalovskiy, I., Sanchez-Martin, M., Bommakanti, N., Garofalo, D. C., Balderes, D. A., Sussel, L., Olive, K. P., & Rabadan, R. (2019). Comprehensive characterisation of compartment-specific long non-coding RNAs associated with pancreatic ductal adenocarcinoma. *Gut*, *68*(3), 499–511. https://doi.org/10.1136/gutjnl-2017-314353

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Deisenroth, C., Itahana, Y., Tollini, L., Jin, A., & Zhang, Y. (2011). p53-inducible DHRS3 is an endoplasmic reticulum protein associated with lipid droplet accumulation. *Journal of Biological Chemistry*, *286*(32), 28343–28356. https://doi.org/10.1074/jbc.M111.254227

Ding, W. Q., Cheng, Z. J., McElhiney, J., Kuntz, S. M., & Miller, L. J. (2002). Silencing of secretin receptor function by dimerization with a misspliced variant secretin receptor in ductal pancreatic adenocarcinoma. *Cancer Research*, *62*(18), 5223–5229.

Eguia, V., Gonda, T. A., & Saif, M. W. (2012). Early detection of pancreatic cancer. *JOP : Journal of the Pancreas*, *13*(2), 131–134. https://doi.org/10.1097/mpa.0000000000001024

Gong, J., Mei, S., Liu, C., Xiang, Y., Ye, Y., Zhang, Z., Feng, J., Liu, R., Diao, L., Guo, A. Y., Miao, X., & Han, L. (2018). PancanQTL: Systematic identification of cis -eQTLs and trans -eQTLs in 33 cancer types. *Nucleic Acids Research*, *46*(D1), D971–D976. https://doi.org/10.1093/nar/gkx861

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining

Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Kitai, Y., Iwakami, M., Saitoh, K., Togi, S., Isayama, S., Sekine, Y., Muromoto, R., Kashiwakura, J. I., Yoshimura, A., Oritani, K., & Matsuda, T. (2017). STAP-2 protein promotes prostate cancer growth by enhancing epidermal growth factor receptor stabilization. *Journal of Biological Chemistry*, *292*(47), 19392–19399. https://doi.org/10.1074/jbc.M117.802884

Klein, A. P., Wolpin, B. M., Risch, H. A., Stolzenberg-Solomon, R. Z., Mocci, E., Zhang, M., Canzian, F., Childs, E. J., Hoskins, J. W., Jermusyk, A., Zhong, J., Chen, F., Albanes, D., Andreotti, G., Arslan, A. A., Babic, A., Bamlet, W. R., Beane-Freeman, L., Berndt, S. I., … Amundadottir, L. T. (2018). Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nature Communications*, *9*(1), 556. https://doi.org/10.1038/s41467-018-02942-5

Martín-Antoniano, I., Alonso, L., Madrid, M., López De Maturana, E., & Malats, N. (2017). DoriTool: A Bioinformatics Integrative Tool for Post-Association Functional Annotation. *Public Health Genomics*, *20*(2), 126–135. https://doi.org/10.1159/000477561

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., … Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, *551*(7678), 92–94. https://doi.org/10.1038/nature24284

Minadakis, G., Zachariou, M., Oulas, A., & Spyrou, G. M. (2019). PathwayConnector: Finding complementary pathways to enhance functional analysis. *Bioinformatics*, *35*(5), 889–891. https://doi.org/10.1093/bioinformatics/bty693

Piñero, J., Bravo, Á., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., & Furlong, L. I. (2017). DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, *45*(D1), D833–D839. https://doi.org/10.1093/nar/gkw943

Rosendahl, J., Kirsten, H., Hegyi, E., Kovacs, P., Weiss, F. U., Laumen, H., Lichtner, P., Ruffert, C., Chen, J. M., Masson, E., Beer, S., Zimmer, C., Seltsam, K., Algül, H., Bühler, F., Bruno, M. J., Bugert, P., Burkhardt, R., Cavestro, G. M., … Sahin-Tóth, M. (2018). Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut*, *67*(10), 1855–

1863. https://doi.org/10.1136/gutjnl-2017-314454

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., Li, Y., Lin, S., Lin, Y., Barr, C. L., & Ren, B. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*, *17*(8), 2042–2059. https://doi.org/10.1016/j.celrep.2016.10.061

Sloan, C. A., Chan, E. T., Davidson, J. M., Malladi, V. S., Strattan, J. S., Hitz, B. C., Gabdank, I., Narayanan, A. K., Ho, M., Lee, B. T., Rowe, L. D., Dreszer, T. R., Roe, G., Podduturi, N. R., Tanaka, F., Hong, E. L., & Cherry, J. M. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Research*, *44*(D1), D726–D732. https://doi.org/10.1093/nar/gkv1160

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., Oliver-Williams, C., Kamat, M. A., Prins, B. P., Wilcox, S. K., Zimmerman, E. S., Chi, A., Bansal, N., Spain, S. L., Wood, A. M., … Butterworth, A. S. (2018). Genomic atlas of the human plasma proteome. *Nature*, *558*(7708), 73–79. https://doi.org/10.1038/s41586-018-0175-2

Watanabe, K., Taskesen, E., Van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Zhang, M., Lykke-Andersen, S., Zhu, B., Xiao, W., Hoskins, J. W., Zhang, X., Rost, L. M., Collins, I., Bunt, M. Van De, Jia, J., Parikh, H., Zhang, T., Song, L., Jermusyk, A., Chung, C. C., Zhou, W., Matters, G. L., Kurtz, R. C., Yeager, M., … Amundadottir, L. (2018). Characterising cis-regulatory variation in the transcriptome of histologically normal and tumour-derived pancreatic tissues. *Gut*. https://doi.org/10.1136/gutjnl-2016-313146

Table S1. Characteristics of the study populations.

	PanGenEU+EPICURO		PanScan I+II		PanScan III		PanC4	
	Cases (N=1317)	Controls (N=1616)	Cases (N=3525)	Controls (N=3642)	Cases (N=1582)	Controls (N=5203)	Cases (N=3933)	Controls (N=3651)
Age (mean, SD)	64.2 (12.4)	64.5 (11.5)	66.2 (10.1)	66.6 (10.2)	*	*		
Sex								
Male	575 (43.66%)	391 (24.20%)	1841 (52.2%)	1885 (51.8%)	781 (49.4%)	3866 (74.3%)	2280 (58.0%)	2038 (55.8%)
Female	742 (56.34%)	1225 (75.80%)	1684 (47.8%)	1757 (48.2%)	801 (50.6%)	1337 (25.7%)	1653 (42.0%)	1613 (44.2%)
Region[a]								
1	32 (2.43%)	47 (2.91%)						
2	1045 (79.35%)	1513 (94.63%)						
3	240 (18.22%)	56 (3.47%)						

a: EPICURO control population
*Age for PanScan III cases and controls was provided as age groups (10 year). See Wolpin *et.al* Nature Genetics 2014

Table S2. Replication of the SNPs reported as associated with pancreatic cancer risk in European population and published in GWAS Catalog.

Gene	SNP	Chromosome region	OR	*p*-value
WNT2B	rs351365	1p13.2	1.18	3.92E-02
LOC105371682-RNU6-716P	rs10919791	1q32.1	0.80	9.67E-03
NR5A2	rs2816938	1q32.1	1.28	6.97E-04
NR5A2	rs3790844	1q32.1	0.81	9.35E-03
ETAA1-LOC107985891	rs1486134	2p14	0.91	1.74E-01
EDNRA	rs6537481	4q31.22	0.86	4.21E-02
TERT	rs2736098	5p15.33	0.75	7.52E-05
CLPTM1L	rs401681	5p15.33	1.18	9.47E-03
CLPTM1L	rs31490	5p15.33	1.18	8.85E-03
TSN3	rs73328514	7p12.3	0.77	1.05E-02
LINC-PINT	rs6971499	7q32.3	0.79	2.04E-02
HNF4G	rs294147	8q21.11	1.14	4.60E-02
SMC2	rs10991043	9q31.1	1.14	4.50E-02
RNY1P8-MARK2P12	rs9543325	13q22.1	0.85	9.96E-03
BCAR1	rs7190458	16q23.1	1.62	4.14E-04
LINC00673	rs11655237	17q24.3	1.46	1.43E-04
LINC00673	rs7214041	17q24.3	1.47	9.12E-05

Table S3. Validated variants (at the nominal p-value), in PanScan and PanC4 populations, among the top 20 SNPs identified in the PanGenEU GWAS study.

Cytoband	SNP	I^2	OR	95%CI	*p-value (metaanalisis)*
1q32.1	*NR5A2*-rs3790840	58.07	1.23	(1.12, 1.34)	5.91E-06
19p13.11	*NWD1*-rs773914	80.37	1.27	(1.02, 1.58)	3.39E-02

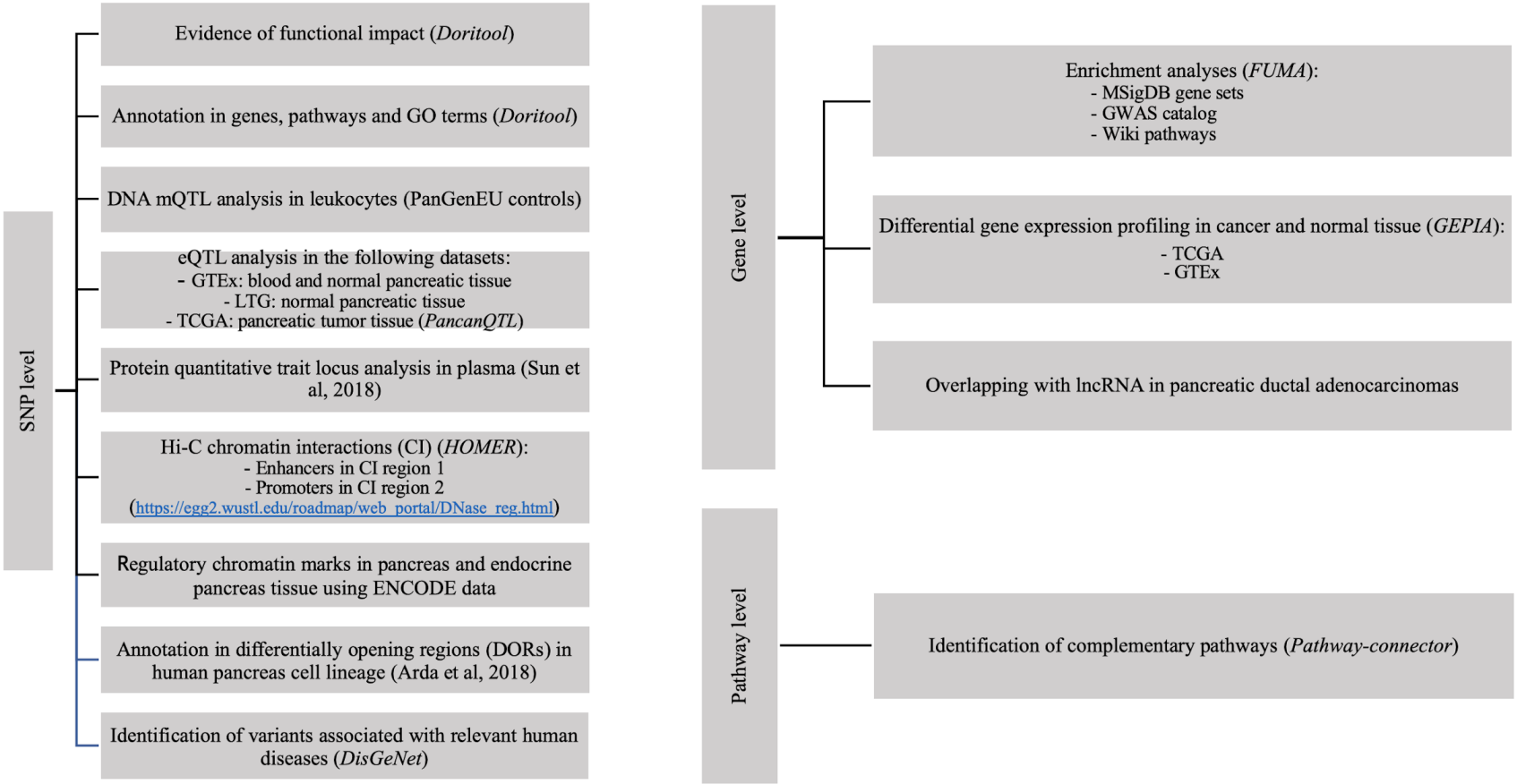


Figure S1. Functional *in-silico* analysis strategy followed to identify novel genomic regions previously prioritized using the 1D, 2D and 3D approaches.

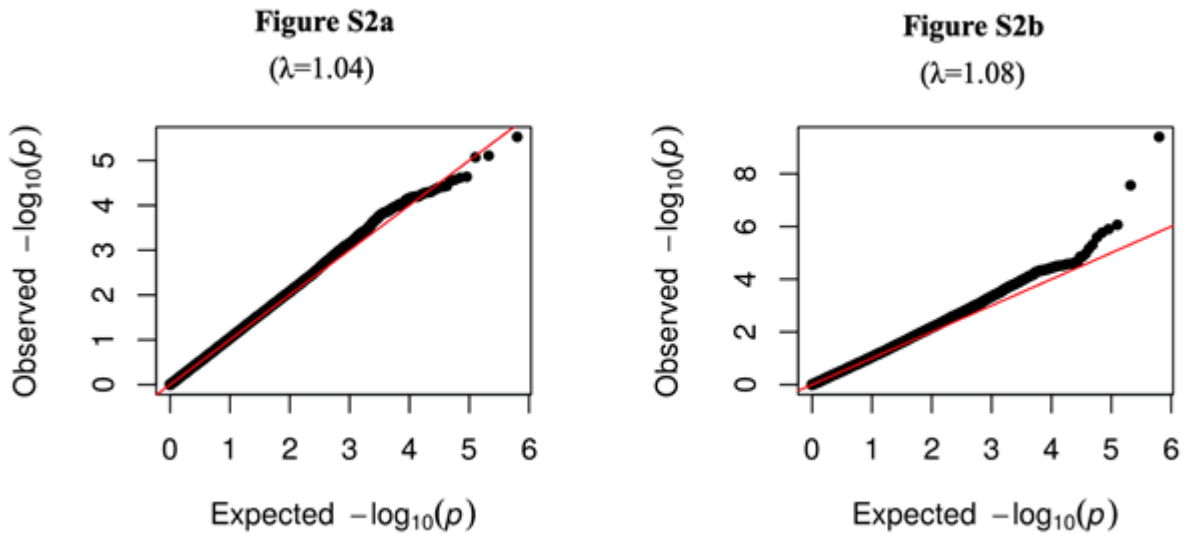


Figure S2. GWAS Manhattan plot for the PanGenEU study. The x-axis is the genomic position of each variant and the y-axis is the −log10 *p*-value obtained in the 1D analysis.

Figure S3a

(λ=1.04)

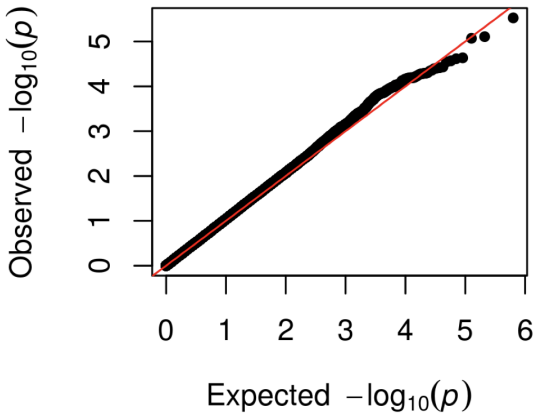


Figure S3b

(λ=1.08)

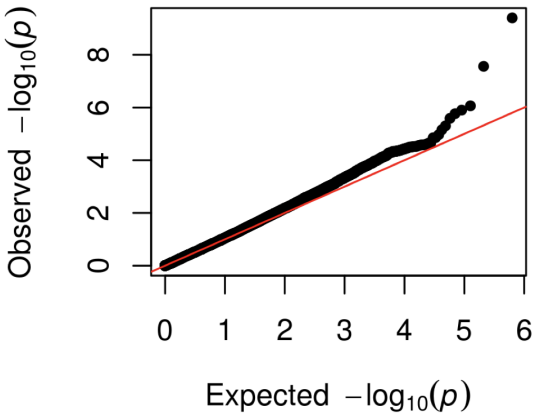


Figure S3. Q-Q plots for pancreatic cancer risk of the association results using the PanGenEU case-control study (S2a) and PanGenEU&EPICURO study populations (S2b).

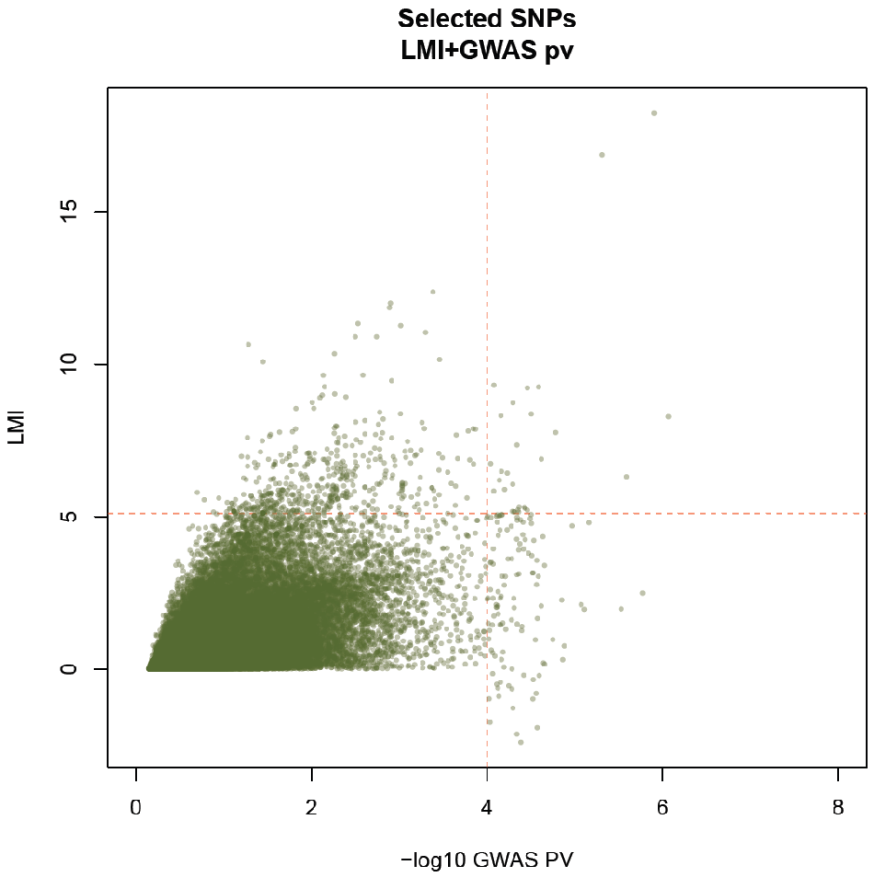


Figure S4. Scatterplot of the local Moran's index (LMI) obtained in the 2D approach and the –log10 *p*-value obtained in the GWAS analysis (1D approach).

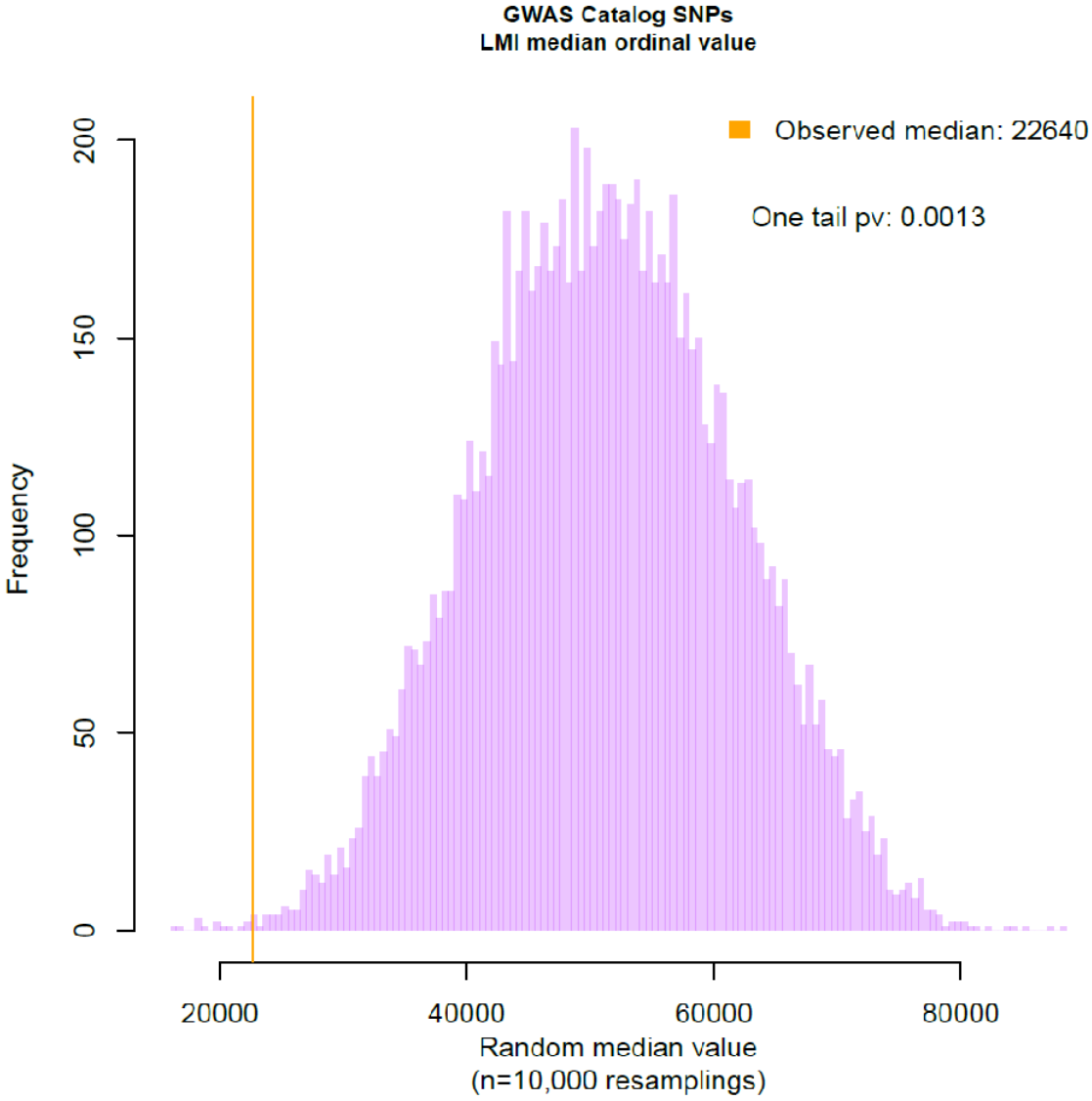


Figure S5. Results of the benchmarking test showing that the median rank position of the LMI values for the 22 pancreatic cancer signals from the GWAS Catalog is significantly higher than 10,000 randomly selected sets of the same size.

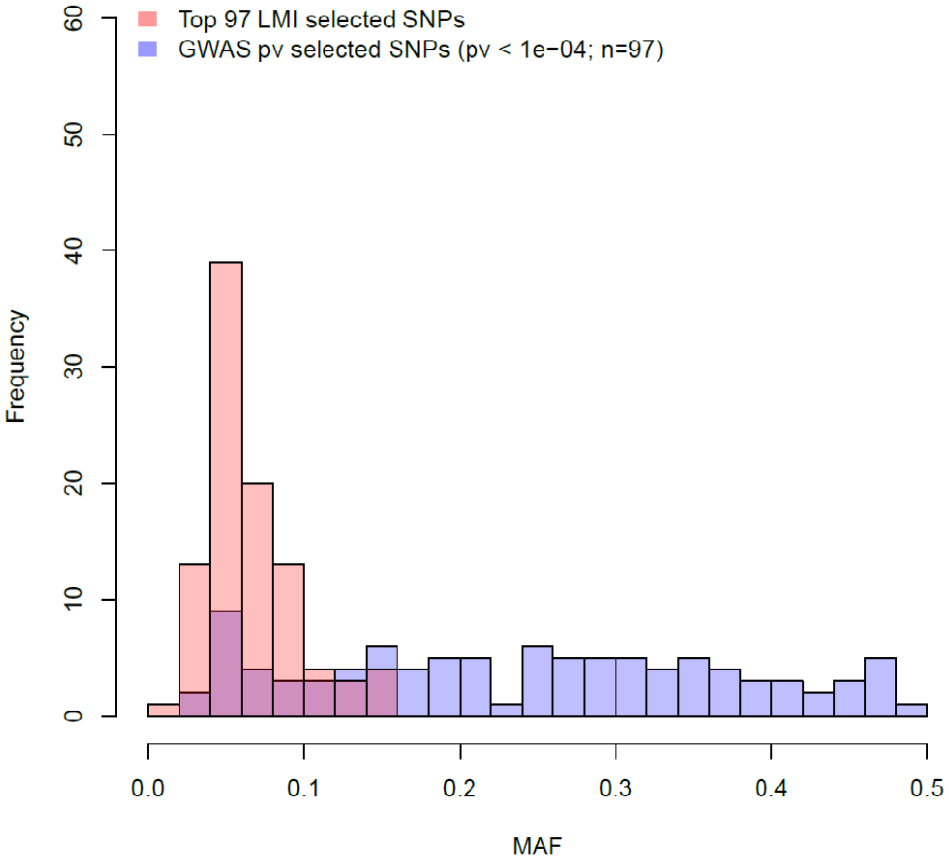


Figure S6. Minor allele frequency (MAF) distributions for the top 97 SNPs identified by LMI (in pink) and by GWAS (in blue).

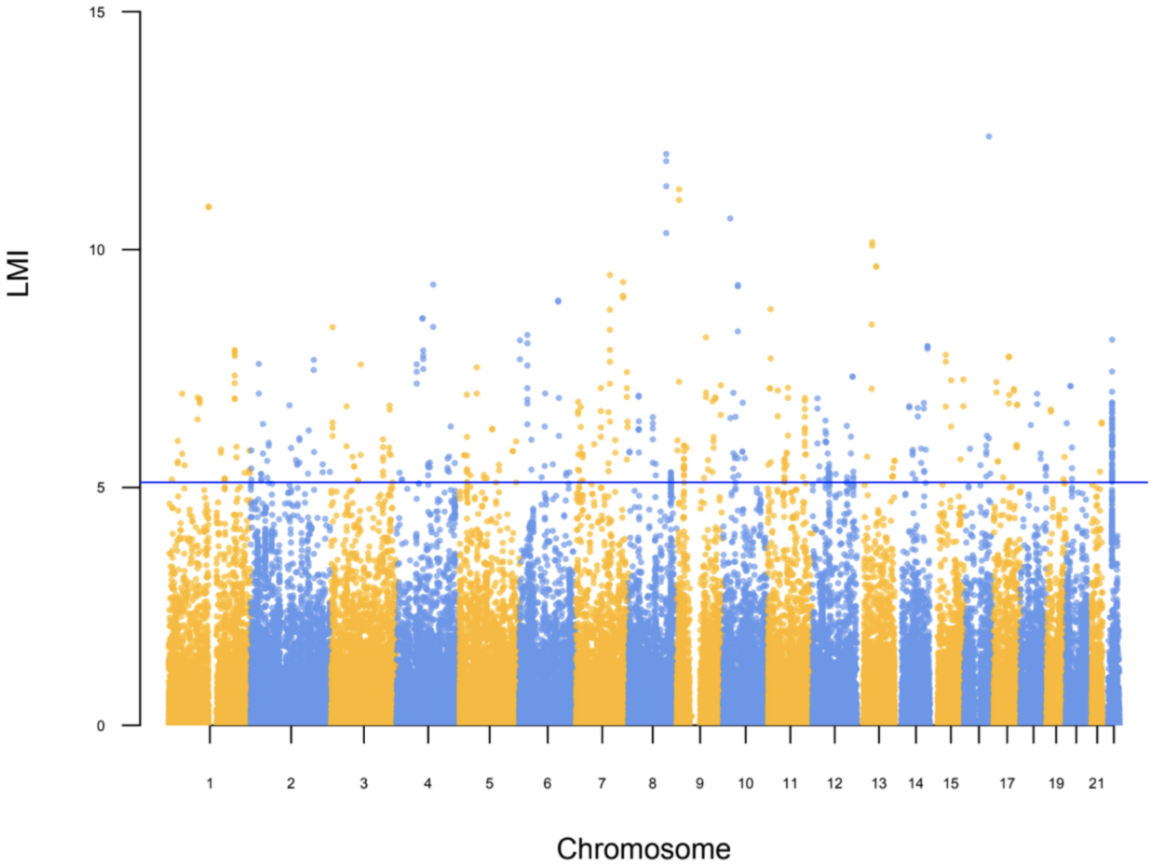


Figure S7. LMI Manhattan plot for the PanGenEU study. The x-axis is the genomic position of each variant and the y-axis is the LMI value obtained in the 2D analysis

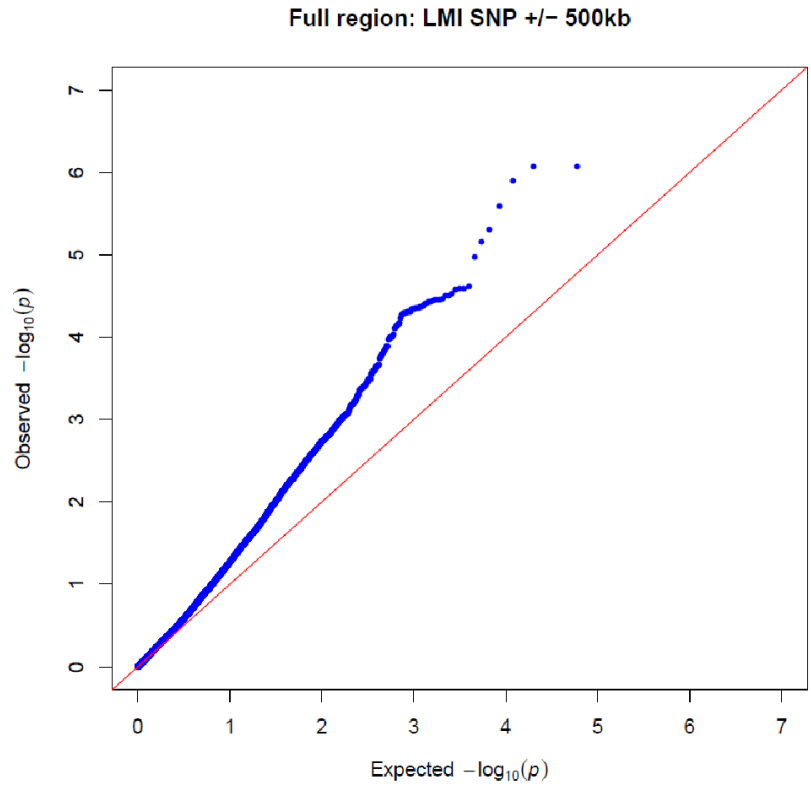


Figure S8a

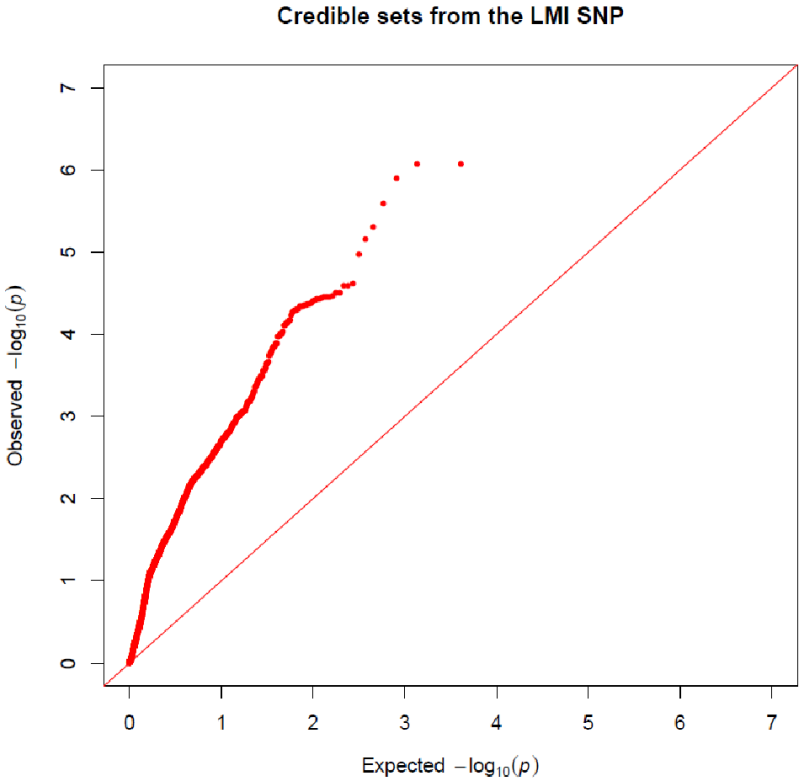


Figure S8b

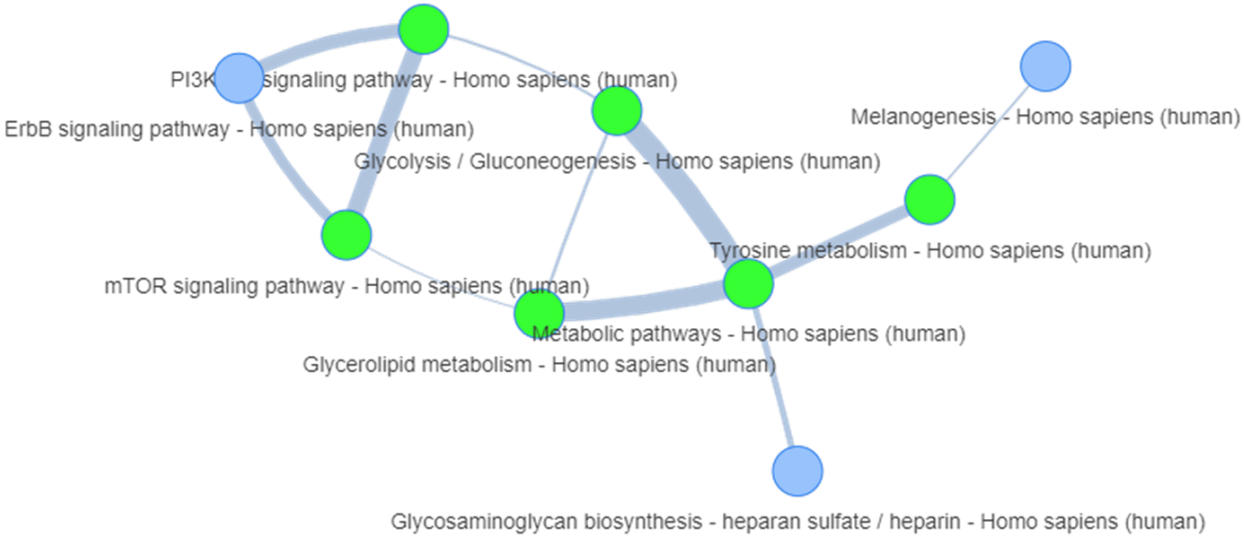


Figure S9. Complementary NETWORK (in blue our input pathways; in green, the complementary nodes included) obtained with Pathway-connector webtool.